The Spontaneous Emergence of Language Variation from a Homogeneous Population

Robin Clark[1] and Steven O. Kimbrough[2]

[1] University of Pennsylvania, Philadelphia PA 19104, USA,
Department of Linguistics
rclark@babel.ling.upenn.edu
[2] University of Pennsylvania, Philadelphia PA 19104, USA,
Department of Operations, Information, and Decisions
kimbrough@wharton.upenn.edu

Abstract. We present a large scale model of socio-phonetic micro-variation; the simulation involves a system of multiple vowels, each consisting of two acoustic formants, a continuous variable over which vowels may range. The population consists of approximately four thousand agents divided into two groups, each with a related underlying vowel system. In addition, the agents are equipped with a simple method of concept formation in response to their linguistic experience; this allows them to class acoustic input into phonemes. We show, first, that a heterogeneous population will, over time, blend their vowels into a coherent system of vowels. Second, when there are influential agents—*Language Leaders*—these leaders will introduce distortions into the blending process, although the vowels ultimately blend into a coherent system. Third, when the Language Leaders form a network over which they preferentially talk to each other, the resulting system will be distorted and the segregated groups will fail to blend. Finally, from this result we will show that a homogeneous population, segregated into two groups, will spontaneously exhibit language variation under the influence of language leaders who preferentially signal to each other over a network.

1 Introduction

Language change is a central example of social change, and yet "the mechanism by which cultural patterns drive language change remains mysterious" [4, page 121]. Language change is of particular interest to social scientists because (1) it is pervasive so that there is a great deal of data available and (2) it is known that change in the language system tends to reduce, rather than enhance, ease of communication. On the latter point, it is accepted that blendings and mergers are dominant forces in language change, and yet they necessarily reduce the ability of the language to make contrasts, which of course interferes with the ability to make expressive contrasts.

Further, it is generally held that language variation is a necessary prerequisite for language change, so the question of the emergence of language variation would seem to be an intimate part of the study of language change. Current thinking on language variation suggests that it arises when groups with distinct linguistic properties come in contact with each other; these groups can represent distinct languages or distinct sociolinguistic groups (for example, different classes within a society).[3] This line of reasoning involves a regress, for these distinct sociolinguistic forms themselves had to develop somehow.

Broadly speaking, social science offers three modes of explanation for cultural change. The first is Darwinian selection (taken in a wide sense). The second is self-organization and the third is history or path dependence [1]. The standard view of language change (above) is a Darwinian one—variants are selected—but the ontogeny of the variants is left unaddressed. In this paper, we will provide and analyze a hypothesis for how language variation can arise spontaneously even in a population that is linguistically homogeneous. Our explanation, embodied in an agent-based model using a real vowel system [6], shows how self-organization and path dependence alone can spontaneously generate linguistic variation that is the necessary input to any process of language change.

2 Summary of the Model

We present, in particular, an Agent-Based Model consisting of a population of agents, split into two groups (known in the model as "Lefties" and "Righties"). The model name is Exemplars and Simple Segregation. It is implemented in NetLogo and available on the Modeling Commons (`http://modelingcommons.org`) as *Exemplars and Simple Segregation.nlogo.* The agents are equipped with an underlying linguistic system of eight vowels, each vowel specified by two acoustic parameters: the first and second formants. Formants can be thought of as resonances in speech sounds that correlate with the shape of the vocal tract [7]. These resonances give clues to the position of the articulators, shape of the vocal tract, and so on. The first formant (F1) correlates with tongue height and the second formant (F2) correlates with the frontness of the tongue. Figure 1 shows the formants for the front vowels of English, as in the words *beat*, *bit*, *bait*, *bet* and *bat* (the formants are highlighted in red in the figure).

The underlying systems may be systematically varied according to the group or the systems of the two groups can be identical. In addition, each agent's representation of the vowel is unique, determined by its previous phonetic experience; each agent has queues consisting of a sequence of exemplars, a separate queue for each vowel. Since each agent's individual experience is different, each agent's vowel queues will be distinct; there is, thus, no overarching shared representation for the vowels. Although each vowel has a queue of exemplars associated with it, these queues receive summary representations in the form of averages for each of the vowel's acoustic formants; we will refer to these summary representations as *prototypes*. The formants of the basic system, drawn from Javanese, is shown

[3] For a discussion of the importance of language contact to language change see the classic work of Weinreich, [8], as well as the discussion in the first volume of Labov's *Principles of Linguistic Change*, [2]

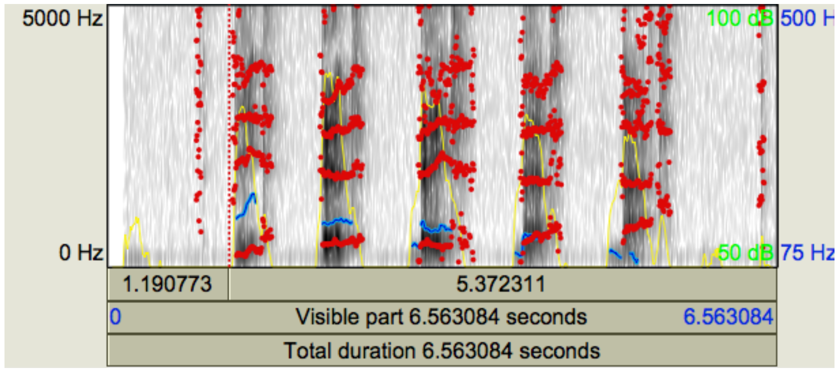


Fig. 1: Formants for “beat”, “bit”, “bait”, “bet” and “bat”.

in Table 1. Example systems for the two groups, Lefties and Righties, are shown in Table 2, which shows the identifying number of the vowel, the corresponding IPA symbol, the frequencies of the first and second formants, and the ratio of F2 to F1.

Number	IPA Symbol	F1: First Formant	F2: Second Formant	F2/F1
1	i	299	2190	7.324
2	e	364	2260	6.209
3	ɛ	481	1980	3.929
4	ɑ	750	1250	1.667
5	ə	377	1400	3.714
6	u	290	630	2.172
7	o	392	690	1.760
8	ɔ	502	790	1.574

Table 1: Formants in the basic system. Formant values in Hz.

Each agent receives a signal, consisting of the two formants of a single vowel, from another agent. This vowel is generated by randomly selecting one vowel prototype from the sender’s system and adding noise to the formants. Once the receiver gets the vowel formants, it tries to classify the signal relative to its own system of vowels; the goal is to have the sender and receiver coordinate on the vowel that was sent, though no feedback is given as to whether the receiver correctly classified the input vowel or not.

The method of classification used in the model tries to find an invariant property of the signal, although the sender and receiver might be from different social groups and, thus, have different subsystems. Our method is as follows, each prototype vowel can be summarized by a real number given by the ratio of F2 to F1; the ratios are shown in the last column of Table 1. Briefly, F2/F1 of

No.	IPA	F1	F2	F2/F1
1	i	269.1	1971.0	7.324
2	e	327.6	2034.0	6.209
3	ɛ	432.9	1782.0	3.929
4	ɑ	675.0	1125.0	1.667
5	ə	339.3	1260.0	3.714
6	u	261.0	567.0	2.172
7	o	352.8	621.0	1.760
8	ɔ	451.8	711.0	1.574

(a) Standard Lefty system, shifted down 10% from Table 1.

No.	IPA	F1	F2	F2/F1
1	i	328.9	2409.0	7.324
2	e	400.4	2486.0	6.209
3	ɛ	529.1	2178.0	3.929
4	ɑ	825.0	1375.0	1.667
5	ə	414.7	1540.0	3.714
6	u	319.0	693.0	2.172
7	o	431.2	759.0	1.760
8	ɔ	552.2	869.0	1.574

(b) Standard Righty system, shifted up 10% from Table 1.

Table 2: Lefty and Righty vowel systems.

the input vowel is compared with the ratios of the agent's prototype vowel. The number that is the closest to the input vowel is then guessed and the new vowel is placed on the guessed vowel's exemplar queue. Although the method is quite simple, the results are extremely accurate. Table 3 and Table 4 show confusion matrices for both the Righties and the Lefties after an experimental run of 40,000 ticks. The confusion matrices can be read as follows: Rows represent the signal sent by the signaler while columns represent the guess made by the receiver; the numbers in each cell show the number of times the receiver guessed the vowel in the column when the signal sent was the vowel in the row. As can be seen from the tables, our method is extremely accurate, with less than 1% errors.[4]

	i	e	ɛ	ɑ	ə	u	o	ɔ
i	9145733	1	0	0	0	0	0	0
e	0	9149090	0	0	0	0	0	0
ɛ	0	0	9150376	0	4	0	0	0
ɑ	0	0	0	9152556	0	0	1356	1505
ə	0	0	331	0	9153398	0	0	0
u	0	0	0	0	0	9151695	1	0
o	0	0	0	87117	0	0	9059696	0
ɔ	0	0	0	13813	0	0	0	9133328

Table 3: Lefty confusion matrix at end of run: Bigotry=0, leader-effect=20%, and no leaders.

[4] Let us take care to note that our system is a drastic simplification of actual speech recognition. In our system, the vowels are unaffected by the coarticulatory effects of surrounding consonants, for example, and we don't attempt to treat diphthongs; both of these would require vectors sampled from the speech stream as opposed to points. In addition, we have followed sociolinguistic practice in focusing on F1 and F2; in reality, more information about the vowel's acoustics would be necessary. Nevertheless, our system shows a high degree of tolerance for speaker variation as we desired.

	i	e	ɛ	ɑ	ə	u	o	ɔ
i	9454202	0	0	0	0	0	0	0
e	0	9459837	0	0	0	0	0	0
ɛ	0	0	9458157	0	0	0	0	0
ɑ	0	0	0	9452060	0	0	557	646
ə	0	0	113	0	9450892	0	0	0
u	0	0	0	0	0	9456369	0	0
o	0	0	0	67079	0	0	9385023	0
ɔ	0	0	0	7601	0	0	0	9447464

Table 4: Righty confusion matrix at end of run: Bigotry=0, leader-effect=20%, and no leaders.

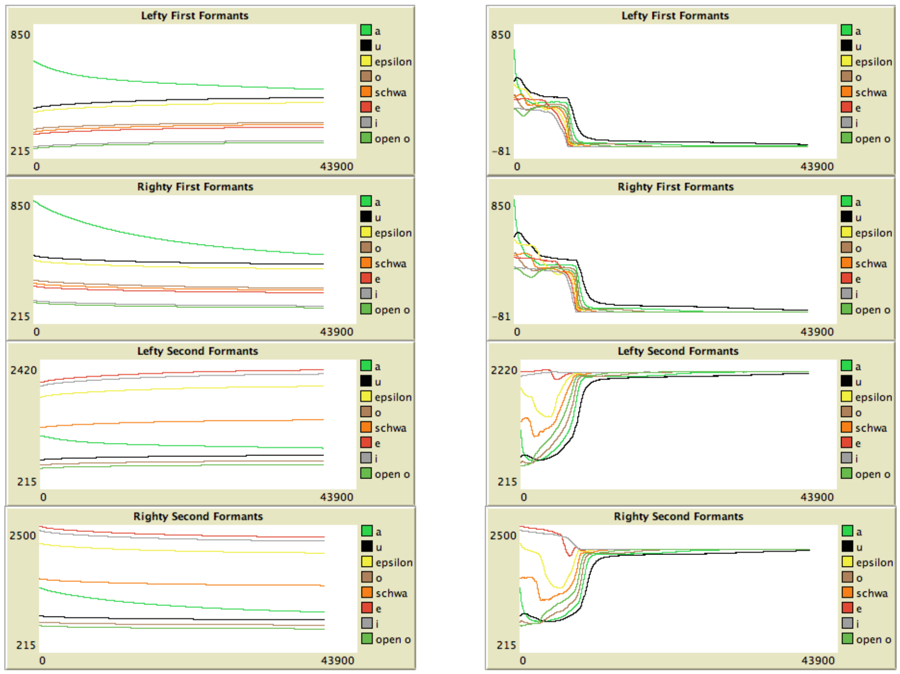
Finally, a subset of the agents of each group can be designated as leaders; these leaders may be connected in a network over which they may communicate. The distinctive property of leaders is that agents from the leader's group give their signals greater weight than is accorded to the signals of non-leaders. In other words, language leaders are highly influential among members of their own group, although they are treated as regular agents by members of the opposite group. Two crucial property of the network that connects leaders of the same group are (1) that members of the network may talk "long distance", transcending the normal spatial effects other agents are subject to; and (2) leaders may preferentially talk to each other more frequently than they talk to ordinary agents. This last features mean that leaders can form a clique within which they may coordinate their signaling.

3 Results and Analysis

For present purposes we can see the gist of our results by comparing two extreme cases. In the first, Figure 2a, we see that the vowel systems of the Lefties and the Righties, initially separated by about 20% for each vowel, blend and asymptotically merge. The two societies contain no leaders and are not hierarchical.

Compare the situation in Figure 2a with Figure 2b, which is blending only in the sense that the systems collapse. In the setup whose results are shown in Figure 2b, there are 320 leaders within each group (Lefties or Righties); these leaders are networked with one another and preferentially signal over the network; and when they speak to non-leaders they have strong influence on them, their utterances are weighted 80 times that of a normal agent. Of course, this is an extreme condition, one we do not expect occurs in nature, but it serves to illustrate the behavior of the model.

Without the distinctions in formant frequencies among the vowels, as in Figure 2b, communication would be impossible. That it is so is clearly reflected in the confusion matrices. We recorded by vowel for the length of the runs, at 100 tick intervals, the cumulative confusions for all of the utterances. Figure 3 compares the cumulative confusions—the number of times the agents guessed the



(a) Blending without leaders.

(b) Blending with leaders.

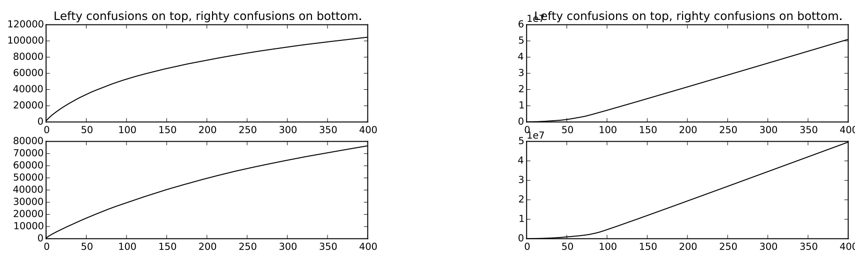
Fig. 2: Blending without (a) and with (b) leaders.

target vowel incorrectly—for the setup with no leaders, Figure 3a, and the setup with leaders, Figure 3b. What Figure 3a indicates is that at the beginning of the run confusions were occurring at a higher rate, especially for the lefties, with the slope of the line fairly high. Later, as blending occurs, both slopes become more gentle, presumably because the populations—Lefties and Righties—have increased their mutual intelligibility.

The story for the run with highly influential networked leaders, Figure 3b, is quite different. Here we see that initially the slope of the confusion line is quite gentle compared to what happens after about 75 ticks on the x-axis. Note that the scales of the two graphs are very different. On the left, the leaderless groups are increasing linearly with total confusion counts after 40,000 model ticks in the 70–100,000 range. On the right the rate of increase is steeper and the total counts after 40,000 ticks are about 5×10^7, about 2 orders of magnitude greater. In other words, the leaderless group is achieving greater mutual intelligibility over time (they are co-ordinating on which signal is sent for each vowel), and in the group with highly influential leaders the confusions are increasing to a point where receivers are guessing more or less at random.

A different look at the data confirms this account. Figure 4 plots the lag-1 values for the cumulative confusion counts, and thus plots the absolute confusion values, as they change over the course of the run. On the left, Figure 4a, we see that the leaderless populations move from having a modest level of confusion at the start to rather low levels of confusion at the end. (Note: the two graphs have slightly different scales. This account for the generally more jagged appearance of the bottom graph, for the Righties.)

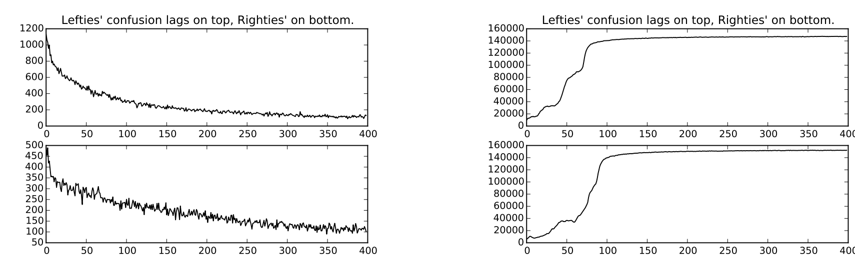
On the right, Figure 4b, we see exactly the opposite story: starting from a low level of confusion, the populations rapidly increase their confusion rates (and counts) so that after 100 graph ticks they reach a ceiling of mutual incomprehension. (Notice the scale differences between Figures 4a and 4b. The level of confusion in the populations with leaders is hugely higher than the populations without leaders.)



(a) Cumulative confusions without leaders.

(b) Cumulative confusions with leaders.

Fig. 3: Cumulative confusions (a) without leaders and (b) with leaders.



(a) Confusion lags without leaders.

(b) Confusion lags with leaders.

Fig. 4: Confusion lags (a) without leaders and (b) with leaders.

As noted above, the configuration underlying the right hand sides of the Figures 2–4 is extreme. Let us now look at a considerably less extreme case. Figure 5 shows the screen at the end of a run in which:

1. The Lefties and Righties do not speak to each other (`bigotry`=0). The effect is to have a single run that conducts two experiments, one on the Lefties and one on the Righties.
2. The Lefties have 160 leaders (half the number of leaders in the experiment of Figure 2b), the Righties have none.
3. The `leader-effect` is 60, somewhat weaker than the 80 of the previous example; see Figure 2b.
4. The probability that a leader, when it gets a chance to speak, will automatically pick another leader to speak with is `pleaderleader` = 0.2. This compares with 0.6 for the experiment of Figure 2b.

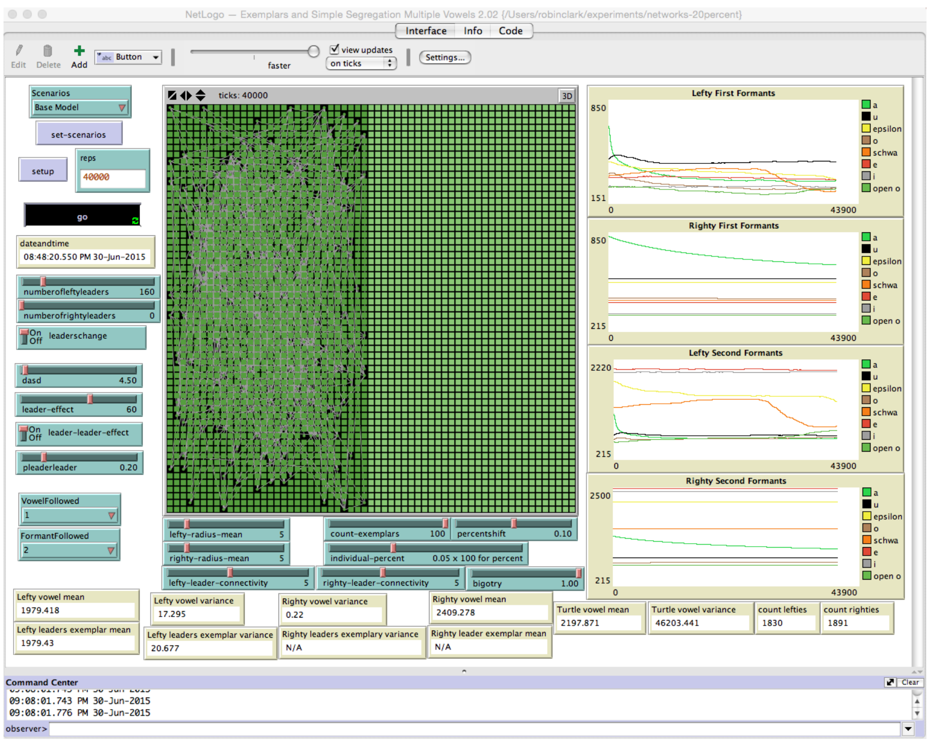


Fig. 5: Intermediate example: Network of 160 Lefty leaders with `pleaderleader` = 0.2 and `leader-effect`=60. Lefties and Righties do not talk with each other.

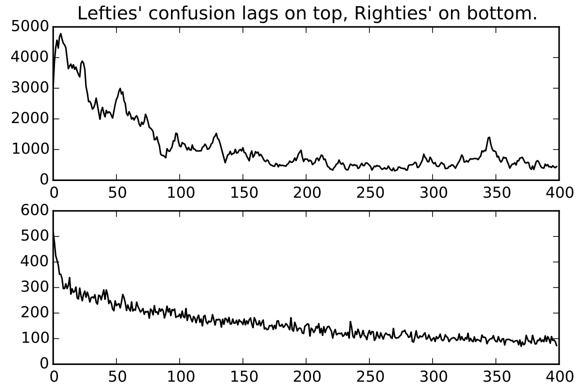


Fig. 6: Confusion lags: (Top) Network of 160 Lefty leaders with `pleaderleader` = 0.2 and `leader-effect`=60; (Bottom) Righties, no leaders. Lefties and Righties do not talk with each other.

Figure 6 presents plots of lag-1 confusion values for the Lefties and the Righties. As such it corresponds to Figures 4a and 4b, with the top of Figure 6 corresponding to Figure 4b and the bottom to Figure 4a. Notice that there is a drastic qualitative ("phase") change: In Figure 4b confusions increase to essentially a maximum and in Figure 6 (Top) confusions decline quite substantially. In the one, the population collapses into mutual incomprehensibility; in the other the population evolves its vowel system to achieve a good degree of coordination in their signaling, so that at the end of the run communication is much more effective than at the start.

Further, if we now compare the top and bottom of Figure 6 we see that the Righties with no leaders also have a high degree of mutual comprehension when compared to the Lefties with their network of leaders. The Righty vowel system does evolve, due largely to the vowel /a/. This can be seen readily in the plots of Figures 2a and 5, where the green line signifies the vowel /a/. The explanation for this change is easily seen when we consider our method of using formant ratios (see the last column Table 1, which gives the formant ratios for the various vowels in the system). The F2/F1 ratio for [a], is very close to the ratios for [o] and [ɔ] and so there is likely to be a high degree of confusion among these vowels. While the change from /a/ to /ɔ/ is a common one, and we think the behavior of our model is broadly accurate, we would not want to push our method as a full explanation of this fact. That is, although we believe the behavior of the model in this aspect is roughly correct, the actual mechanics of speech perception are quite complex and are not mimicked in our model.

Above all, what Figure 6 teaches, in the context of the other experiments above, is that linguistic variation can arise spontaneously due to the mere presence of a network of language leaders. Thus, language change does not necessarily require contact between disparate groups. Language variation is an inevitable product of a hierarchical social structure. We now turn to a discussion of the consequences of this finding in the next section.

4 Discussion

In addition to the three experiments discussed above, which are quite representative, we conducted a few score other experiments, which space limitations prevent us discussing in detail here. What emerges reliably is that replications of a given experiment vary very little, so that the single results we present here are indeed representative. More importantly, the general pattern that emerges is that with networked leaders, leader influence and the degree that the leaders talk among themselves, in distinction to the non-leaders, are the main factors driving variation. To see this, we can consider a 2×2 design, with two levels of leader influence, A-hi and A-lo, and two levels of tendency for leaders to talk among themselves, B-hi and B-lo. The four cases when listed are:

– A-hi/B-hi. Figure 2b (vowel system collapse).
– A-lo/B-hi. Figures 5 and 6 (leader-induced variation with long-term coordination).
– A-hi/B-lo. Figure 8 (enhanced leader-induced variation; see discussion below).
– A-lo/B-lo. Figure 2a (no leaders, pure blending). Figure 7 (reduced leader-induced variation; see discussion below).

Figure 7 is a non-extreme case of the A-lo/B-lo condition. What we see is that as always the Righties (lacking leaders) are pure blending, have low variation, and increasingly coordinate (reduce confusions) over the life of the run. The Lefties are also quite well behaved, but with more apparent variation. Visually, it is apparent that the two populations are each coordinating on their vowel systems and that at the end of the run they are approximately equal in their levels of confusion, although overall there is more variation among the Lefties. As a rough measure of this we looked at the last 1000 ticks, sampled every 100 ticks. The confusion data is 114, 143, 140, 117, 103, 127, 133, 88, 95, 75 for the Lefties and 94, 87, 85, 93, 88, 90, 99, 67, 84, 94 for the Righties. The means and standard deviations are 113.5 (21.75) and 88.1 (8.30) for the Lefties and Righties respectively. In sum, although the presence of leaders induces some variation beyond that present without leaders, with low influence and low concentration of leader-leader interactions, the level of variation induced is low and hardly disruptive.

Figure 8 is a non-extreme case of the A-hi/B-lo condition. What we see, again, is that the Righties (lacking leaders) are pure blending, have low variation, and increasingly coordinate (reduce confusions) over the life of the run. The Lefties also tend towards successful coordination in the long run, but now they are not quite as well behaved, having more apparent variation than their counterparts in Figure 7. Looking at the last 1000 ticks, sampled every 100 ticks, the confusion

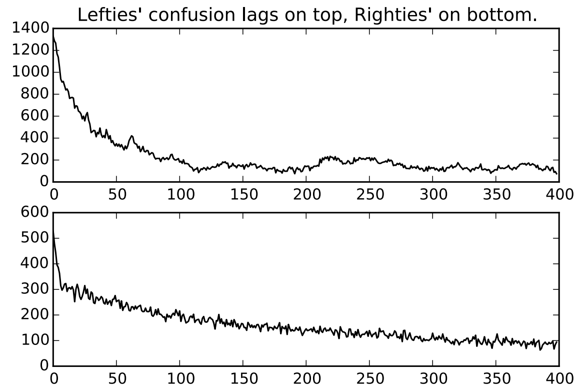


Fig. 7: Confusion lags: (Top) Network of 160 Lefty leaders with `pleaderleader` = 0.2 and `leader-effect`=20; (Bottom) Righties, no leaders. Lefties and Righties do not talk with each other.

data is 388, 401, 519, 546, 543, 539, 607, 607, 606, 491 for the Lefties and 89, 90, 95, 86, 82, 83, 76, 83, 81, 97 for the Righties. The means and standard deviations are 524.7 (75.0) and 86.2 (6.2) for the Lefties and Righties respectively. In sum, the presence of highly influential leaders induces variation well beyond that present without leaders, and indeed well beyond that present with high leader-leader interaction but low influence leaders.

The short summary is that endogenous language variation inevitably occurs in the presence of highly influential language leaders. The leaders form a subsample with their own variance, which is imposed on the larger population by virtue of the high influence of the leaders. Variation can also spontaneously arise when less influential leaders are linked by a social network, since the network amplifies the natural variance in the population of leaders. Sociolinguistics has, in recent years, begun to appreciate the contribution of social networks (as well as language leaders) to shaping language variation [3, 5]. Our results suggest that networks of language leaders will cause variation, even when variation is otherwise absent in the population. We anticipate that an exploration of leader networks and the interaction of these networks with social identity will aid in our understanding of social change and cultural evolution.

The ramifications of these findings may have a deep impact on our understanding of social evolution. Non-darwinian factors, specifically, self-organization in hierarchical populations may well be an operant cause of the variation on which social evolution depends.

Of course, this is only an initial, exploratory study. As a next step, we need to come to a more systematic understanding of how the structure and organization of leader networks influence the population. In addition, and most crucially, a more systematic exploration, incorporating field data much of it of a type not

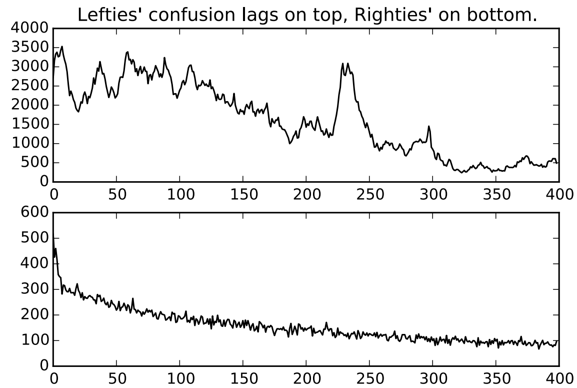


Fig. 8: Confusion lags: (Top) Network of 160 Lefty leaders with `pleaderleader` = 0.05 and `leader-effect`=60; (Bottom) Righties, no leaders. Lefties and Righties do not talk with each other.

heretofore widely collected, will be required to discern the true causes of social evolution.

References

1. Arthur, W.B.: Complexity and the Economy. Oxford University Press, Oxford, UK (2015)
2. Labov, W.: Principles of Linguistic Change: Internal Factors. Blackwell Publishers, Oxford, UK (1994)
3. Labov, W.: Principles of Linguistic Change, Vol. 2: Social Factors (Language in Society). Blackwell Publishers (2001)
4. Labov, W.: Dialect Diversity in America: The Politics of Language Change. University of Virginia Press, Charlottesville, VA (2012)
5. Milroy, L.: Social networks. In: Chambers, J.K., Trudgill, P., Schilling-Estes, N. (eds.) The Handbook of Language Variation and Change, chap. 22, pp. 549–572. Blackwell Publishing, Inc., Malden, MA (2002)
6. Nettle, D.: Linguistic Diversity. Oxford University Press, New York, NY (1999)
7. Stevens, K.N.: Acoustic Phonetics. Current Studies in Linguistics, The MIT Press, Cambridge, MA (2000)
8. Weinreich, U.: Languages in Contact: Findings and Problems. Mouton & Co, The Hague (1968)