Evolution of language diversity: the survival of the fitness

Zach Solan, Eytan Ruppin, David Horn

Faculty of Exact Sciences

Tel Aviv University

Tel Aviv, Israel 69978

{*rsolan,ruppin,horn*}*.post.tau.ac.il*

Shimon Edelman

Department of Psychology

Cornell University

Ithaca, NY 14853, USA

se37@cornell.edu

October 29, 2002

Abstract

We examined the role of fitness, commonly assumed without proof to be conferred by the mastery of language, in shaping the dynamics of language evolution. To that end, we introduced island migration (a concept borrowed from population genetics) into the shared lexicon model of communication (Nowak et al., 1999). The effect of fitness linear in language coherence was compared to a control condition of neutral drift. We found that in the neutral condition (no coherence-dependent fitness) even a small migration rate – less than 1% – suffices for one language to become dominant, albeit after a long time. In comparison, when fitness-based selection is introduced, the subpopulations stabilize quite rapidly to form several distinct languages. Our findings support the notion that language confers increased fitness. The possibility that a shared language evolved as a result of neutral drift appears less likely, unless migration rates over evolutionary times were extremely small.

1   Introduction

Many of the recent studies that explored the evolution of language in a population of communicating agents assume that the possession of a common language increases the fitness of its users, e.g., (Cangelosi, 2001; Reggia et al., 2001; Nowak et al., 1999; Smith, 2001). Language-conferred fitness is defined as the increase in individual survival probability that stems from successful communication: a group whose members communicate well is supposed to leave, on the average, more offspring. Although the assumption that language increases human evolutionary fitness is intuitively appealing, its validity needs to be examined in the light of the indications that a coherent language can emerge in the absence of language-related fitness (Kirby, 2000; Oliphant, 1999; Oliphant and Batali, 1997), e.g., through a combination of genetic drift and random population diffusion.

1

The present paper examines and compares the effects of fitness and of neutral drift on the emergence of coherent languages. Our approach is motivated both by communication models employed in studies of language evolution (Nowak et al., 2000; Nowak et al., 1999; Nowak and Krakauer, 1999; Komarova and Nowak, 2001; Kirby, 2000; Kirby, 2002; Batali, 1998), and by migration models that are widely used in population genetics, such as the island model (Wright, 1931) and migration matrix (Bodmer and Cavalli-Sforza, 1968). Specifically, we integrate communication and migration effects by constructing a system of several distinct populations that exchange migrants at a controlled rate; within each population, evolutionary dynamics of language is governed by the communication model.

The spread of languages within and between populations in this model can be compared to empirical data on linguistic diversity, a field of study that has been drawing increasing attention over the recent years (Nichols, 1992; Nettle, 1999). Although there are as many as 6500 known languages, about 95% of the world population speaks only 100 languages. The distribution of languages around the world is far from uniform (Grimes, 2001). Hundreds of languages have evolved in Africa and Oceania, while very few evolved in North America and Europe. Examining the distribution of languages, most of which are nearly extinct, can shed light on the evolutionary forces that were involved in their emergence.

The markedly uneven geographic distribution both of the emergence of languages and of their disappearance suggests that powerful dynamic mechanisms are at work. The evolutionary dynamics behind linguistic diversity can be explored using computational simulations, as in (Nettle, 1999). That work studied vowel learning in the presence of migration, in a population distributed over a 7×7 spatial grid, with 20 individuals at each position. Each new offspring learned two continuous parameters (the first two formants), by sampling their values among all the adults in its group. Individuals thus learned the approximate average values of the parameters in their group. Each simulation typically lasted for several hundreds of generations. To simplify the calculations and to keep the group sizes constant, migration always involved an exchange of individuals between two groups.

Nettle examined the dynamics of this system under three modes of fitness influence. The first of those was the neutral mode, corresponding to an absence of selection pressure. The second was the functional selection mode, based on the notion that some variables may confer fitness because they are easier to identify, learn, or use than the alternatives. The third mode corresponded to raising the "social status" of some individuals in each group, and having the new offspring learn selectively from the high-ranking adults in their group. Nettle found that in the absence of social selection even a very low level of migration destroys local diversity. Social selection was found to contribute very strongly both to diversity and to its stability in the face of higher migration rates. Finally, functional selection was seen to amplify and reinforce diversity, but could not bring it about on its own.

2   The model

In the present work, we extend Nettle's migration matrix model (controlling the population exchange between spatially separated "islands") by introducing a measure of the individual fitness based on communication success (Nowak et al., 1999). Thus, in our model both the individual and the social fitness are determined by the communication performance, rather than by external parameters. In addition, we allow true matrix-controlled migration (compared to symmetric swapping of individuals used by Nettle).

The population we study consists of n individuals ($I_1 \ldots I_n$) that can exchange m possible signals ($S_1 \ldots S_m$). For each individual, the signal exchange is described by two matrices, P and Q, the first one defining production and the second comprehension. The elements p_{ij} of the production matrix P are the probabilities that a certain individual will refer to object i with signal j; each row in the matrix P sums to 1. The elements q_{ji} of the comprehension matrix Q are the probabilities that an individual will associate signal j with object i; here too each row sums to 1. As in Nowak's work, the ability of two individuals to communicate depends on their comprehension and production matrices. The payoff value corresponds to the mutual understanding between these two individuals; an individual's average payoff is calculated by summing its payoff with respect to each of its group members, divided by the group size (see equation 1).

$$
\begin{aligned}
F(I_k, I_l) &= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{ij}^{(k)} q_{ji}^{(l)} + p_{ij}^{(l)} q_{ji}^{(k)} \right) \\
F(I_u) &= \frac{1}{n-1} \sum_{v \neq u} F(I_u, I_v)
\end{aligned}
\tag{1}
$$

The evolution of language in this model is based on lexicon transmission from one generation to the next. Individual I_k learns from I_l by sampling the responses of I_l to every object, inducing its own lexicon from the responses of I_l. Three learning modes have been implemented and compared:

1. **Parental learning with fitness** (SELECTION). In this mode, the offspring learn from their parents, with individuals that communicate well producing more offspring. Of the many possible fitness functions linking the number of offspring with the communication payoff we chose the linear fitness.

2. **Parental learning with no fitness** (DRIFT). In this mode, there is no advantage to individuals who communicate well.

3. **Role model** (SOCIAL SELECTION). In this mode, the members of a group acquire their language from a few selected "role model" individuals, singled out by their ability to communicate with their group members.

The migration component of our model is very similar to the original n-island approach of (Wright, 1931), and to the migration matrix approach of (Bodmer and Cavalli-Sforza, 1968). The former assumes that individuals migrate from one subpopulation to another with the same rate, which means that the distances between subpopulations are not taken into account (see matrix M_1 matrix in equation 2 and Figure 1). In the latter approach, the migration rates between subpopulations can be different, e.g., depending on distance (cf. matrix M_2 in equation 2 and Figure 1). In our implementation, each sub-population consisted of N individuals. In each cycle, the migration rate determined the number of individuals relocating from one island to another; a migration matrix controlled the migration rate between the sub-populations.

$$M_1 = \begin{pmatrix} 0 & m & m & m \\ m & 0 & m & m \\ m & m & 0 & m \\ m & m & m & 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 0 & m & \frac{m}{2} & \frac{m}{4} \\ m & 0 & m & \frac{m}{2} \\ \frac{m}{2} & m & 0 & m \\ \frac{m}{4} & \frac{m}{2} & m & 0 \end{pmatrix} \qquad (2)$$
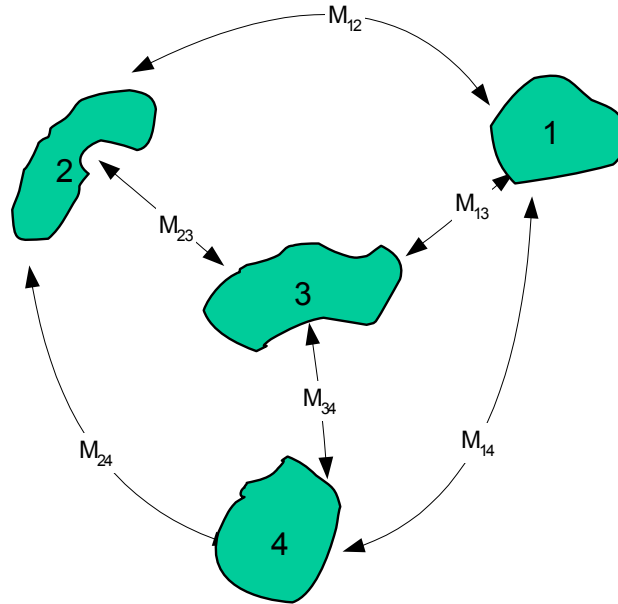

Figure 1: The island migration model (general case).

3   Results

We first analyzed the diversity of languages that evolved under each of the three fitness modes, by considering two kinds of indicators: (1) the *internal payoff*, which is the average payoff within a sub-population, and

(2) the *external payoff*, which is the average payoff between sub-populations. The migration matrix entries were constant and equal for all groups (in the more realistic case the migration matrix is merely symmetric). Figure 2(a) illustrates the time course of the internal and external payoffs across generations in the DRIFT mode. Notice that the after the first 500 generations the internal payoff is higher than the external payoff. After 2000 generations, only one dominant language is left, at which point the internal payoff is equal to the external payoff. The same phenomenon was observed in the SOCIAL SELECTION, where, however, it took much longer for one dominant language to emerge. In comparison, in the SELECTION mode (linear fitness), a diversity of languages is obtained under high internal and low external payoff.

To visualize the process of language development, one may consider the distance matrix whose elements d_{ij} correspond to the quality of understanding between any two individuals within the entire population:

$$d(ij) = 1 - F(I_i, I_j)/n \tag{3}$$

where $F(I_i, I_j)$ is the communication payoff between individuals I_i and I_j. Note that $d = 0$ means maximal understanding between two individuals, and $d = 1$ means no understanding. A simulation with 10 sub-populations of 50 individuals each yields a 500×500 distance matrix. The diagonal values are all equal to zero, because each individual understands itself perfectly.

By embedding this matrix into a two-dimensional space using multidimensional scaling, or MDS (Shepard, 1980), we obtained a series of snapshots of the 2D *understanding space*, which affords a certain insight into the sub-population pattern in each cycle. Figure 3 presents a sequence of snapshots taken every 200 generations in the DRIFT, SOCIAL SELECTION, and SELECTION modes. In the DRIFT mode, the first several snapshots contain several clusters of languages; eventually, the clusters become unstable, that is, most of the languages become extinct and one language prevails. In comparison, in the SOCIAL SELECTION mode the system passes through a stage with several discrete clusters; after 2000 generations one dominant language remains. Likewise, in the SELECTION mode, 10 sub-populations stabilize after 200 generations; after 800 generations only six of these are left.

In Figure 3, there is a large cluster in the DRIFT mode; in the SOCIAL SELECTION mode each of the much smaller clusters eventually converges to a single dot. These singleton clusters represent sub-populations with the most efficient language (recall that the size of a cluster is determined by its internal payoff value). In this sequence, all the clusters become singletons with the maximal internal payoff, which means that this fitness mode finds an optimally efficient language, but cannot stabilize the system in a multiple-language mode; role model learning has been noted to lead to better payoff results (Nowak et al., 1999).

Figure 4 presents a sensitivity analysis of the three fitness modes for different migration rates. The only mechanism that remained stable against a higher rate of migration was the SELECTION mode with linear
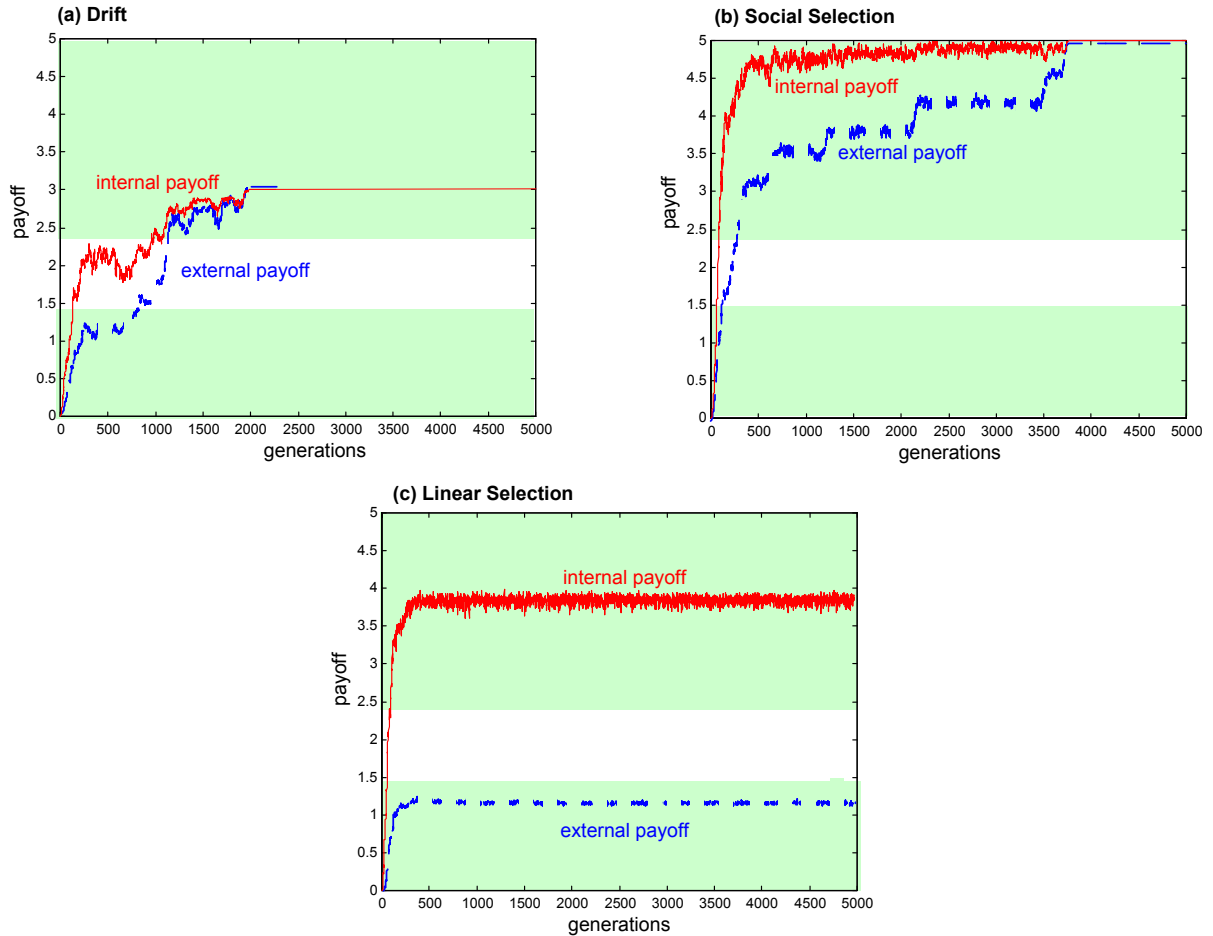
Figure 2: The internal payoff (solid line) and the external payoff (dashed line) in successive generations (the maximum possible payoff according to equation 1 is equal to the number of distinct signals $m = 5$). The upper band of the graph corresponds to a situation with one dominant language (all individuals understand each other regardless of their location). When the red curve is in the upper band and the blue one in the lower band, individuals within the sub-populations understand each other but not the members of other sub-populations (in other words, several languages co-exist). In this run we simulated 10 sub-populations of 50 individuals each. The migration rate between the islands was constant and equal to 0.01 (1% of migration per generation).
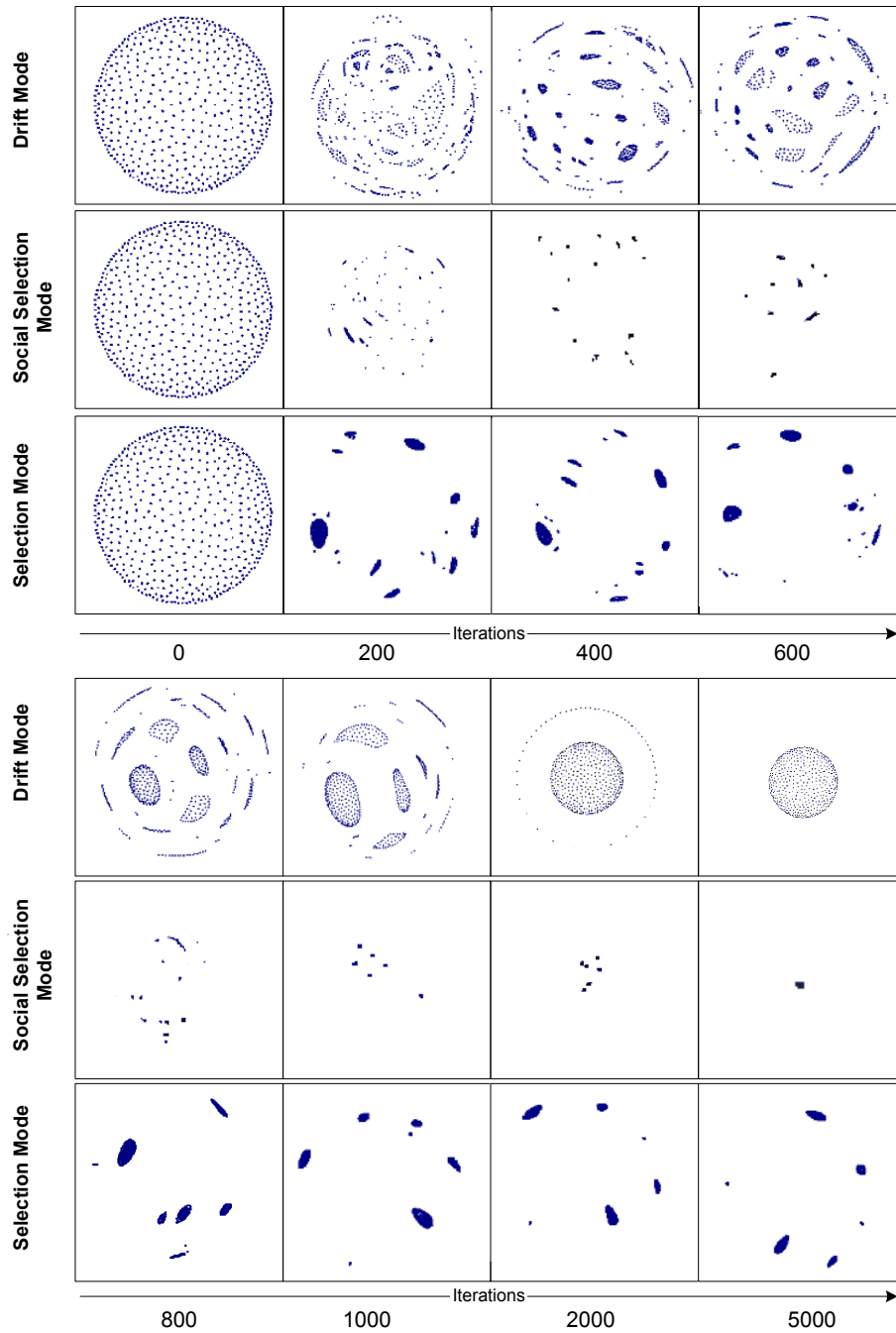
Figure 3: Multidimensional scaling (MDS) snapshots of the communication distance matrix under each of the three fitness modes. The circular arrangement of the clusters is a well-known artifact introduced by MDS, which is of no import in the present case (we are interested in the number of clusters and their shapes and not in their mutual arrangement).

7

fitness. Even in that mode, however, the system converged to a single language when the migration rate exceeded a critical value (about 4%).
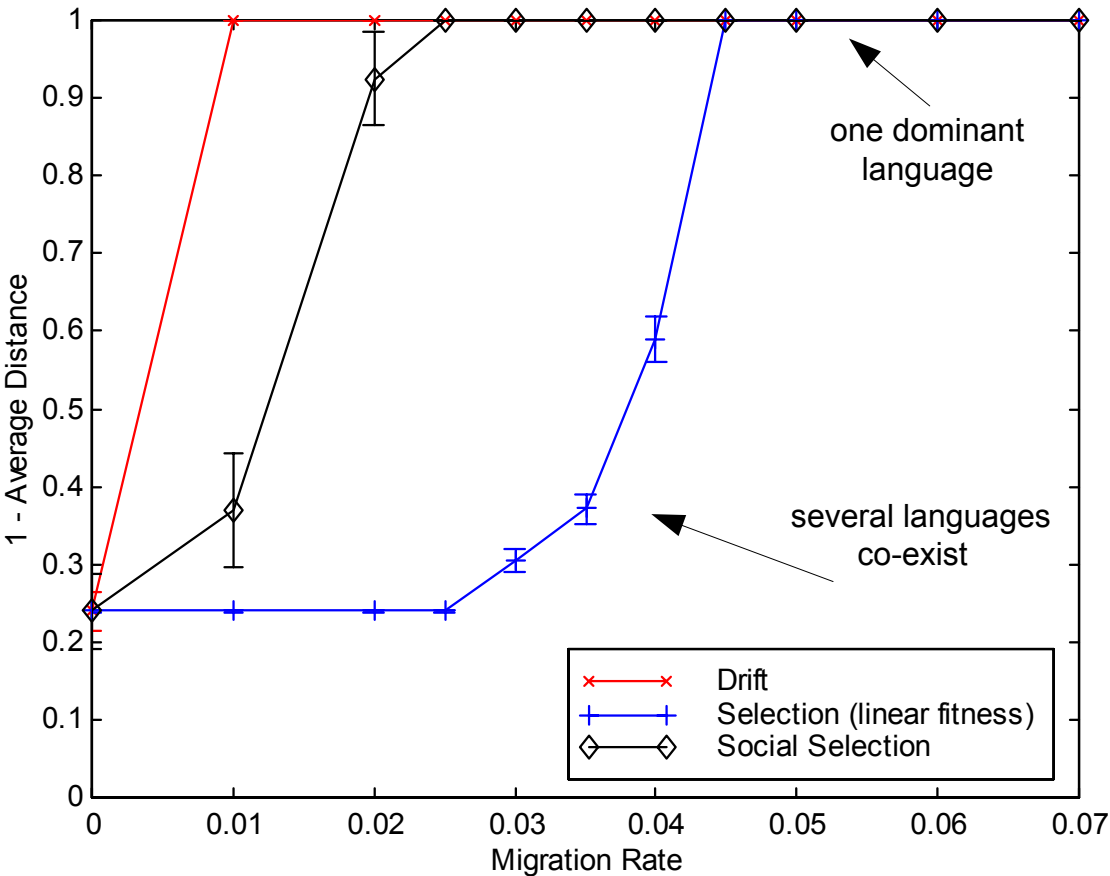

Figure 4: The dependence of linguistic diversity on migration rate under the three fitness modes (DRIFT, pink; SOCIAL SELECTION, blue; and SELECTION, black). The abscissa shows the migration rate and the ordinate – the internal-external payoff ratio, computed by averaging the distance defined by equation 3. A ratio of 1 corresponds to a single dominant language; low ratios indicate high linguistic diversity.

A more realistic scenario in which the sub-populations are randomly located on a 50×50-cell grid is illustrated Figure 5. The migration rate between the groups is determined by a Gaussian function of their distance. The MDS snapshots of the simulation were taken over three different iterations. As pointed out before, only the linear fitness mode can preserve the co-existence of several distinct languages.
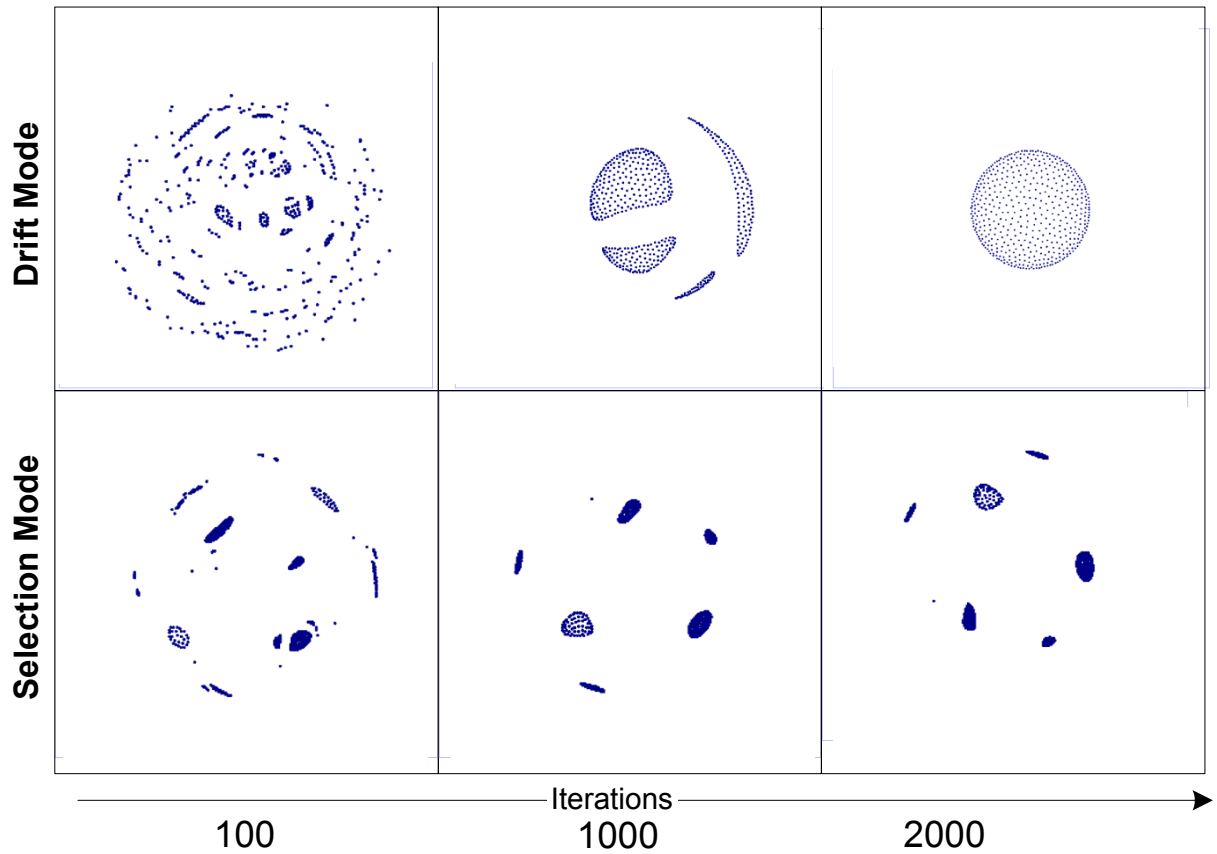
Figure 5: A spatial migration model. The migration rate between the groups is determined by a Gaussian function of their distance on a 50×50 grid. The figure shows the MDS snapshots of the state of the simulation, taken over three different iterations.

4 Discussion

In the language evolution literature, one frequently encounters arguments based on the intuitive assumption that successful communication leads eventually to an increased chance of individual survival. This assumption occupies a prominent place in many models of language evolution, despite the scarcity of evidence supporting it. In this work, we addressed this issue by studying the evolutionary dynamics of language diversity with and without language-conferred fitness.

Our model combines the shared lexicon approach (Nowak et al., 1999) with island migration methods (Nettle, 1999). The results of our simulations suggest that the two main mechanisms responsible for present-day language diversity are: (1) the contribution of language to the fitness of individuals, and (2) the migration of individuals between sub-populations. We found that in the DRIFT and SOCIAL SELECTION conditions it takes much longer for the system to reach a steady state, and that even a very small migration rate suffices for one language to become dominant. In the SELECTION case, when the fitness of individuals is assumed to be proportional to the success of their shared communication, the subpopulations stabilize quite rapidly to form several distinct languages. In this case, only relative high migration rates will force the system into a regime of one dominant language. For each case, a critical value of the migration rate can be identified, beyond which only one language can exist in the long run.

Our findings differ from those of (Nettle, 1999), most likely due to differences in the manner in which selection pressure and social effects have been implemented. In Nettle's model, these factors were not influenced by the ability of the individuals to communicate; in comparison, in the present study mutual understanding between individuals was the key factor shaping the evolutionary dynamics.

The consistent and stable emergence either of one dominant language or of linguistic diversity for different values of a control parameter — the migration rate — has an intriguing analogy in solid-state physics: when certain materials are cooled rapidly, they crystallize heterogeneously (several crystals develop), while a slow gradual annealing of the same material leads to a homogeneous crystallization. Similarly to these phase transition effects, our model predicts, for the SELECTION mode, the emergence of bounded groups with distinct languages rather than continuous dialects. The predominant factor is the critical value of migration rate. In locations where the migration rate is above the critical value, the languages will be fused into one. In locations that share a lower rate of migration, languages will remain divided. In comparison, the DRIFT and SOCIAL SELECTION modes drive languages into extinction very quickly.

Thus, we predict a non-constant rate of language evolution, where the appearance of new languages is very rapid (the first stage of language evolution takes very few iterations), and the extinction of languages is slow. It has been suggested that language evolution undergoes periods of rapid expansion, during which many languages evolve, intermixed with long periods of near-equilibrium, during which languages diffuse and converge, as societies interact, intermarry, fission and fuse (Dixon, 1997). Given that migration rates

10

in human societies consistently increase with time, fewer places in the emerging world remain below the critical rate of migration. Thus, less diverse linguistic environments are to be found in geographic locations where humans arrived late (as in the New World), and, eventually, in the world at large.

References

Batali, J. (1998). *Computational simulations of the emergence of grammar*. Cambridge University Press, Cambridge.

Bodmer, W. F. and Cavalli-Sforza, L. L. (1968). A migration matrix model for the study of random genetic drift. *Genetics*, 59:565–592.

Cangelosi, A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*, 5(2):93–101.

Dixon, R. M. W. (1997). *The rise and falls of languages.* Cambridge University Press.

Grimes, P. (2001). *data from Ethnologue: Languages of the World, 14th Edition*. SIL International.

Kirby, S. (2000). *Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners.*, pages 303–323. Cambridge University Press.

Kirby, S. (2002). *Learning, Bottlenecks and the Evolution of Recursive Syntax*, chapter 6. Cambridge University Press.

Komarova, N. L. and Nowak, M. A. (2001). The evolutionary dynamics of the lexical matrix. *Bull. Math. Biol*, 63(3):451–485.

Nettle, D. (1999). *Linguistic Diversity*. Oxford University press.

Nichols, J. (1992). *Linguitics Diversity in Space and Time*. University of Chicago Press.

Nowak, M. A. and Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14):8028–8033.

Nowak, M. A., Plotkin, J. B., and Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature*, 404:495–498.

Nowak, M. A., Plotkin, J. B., and Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7(3/4).

Oliphant, M. and Batali, J. (1997). Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language*, 11(1).

Reggia, J. A., Schulz, R., Wilkinson, G., and Uriagereka, J. (2001). Conditions enabling the evolution of inter-agent signaling in an artificial world. *Artificial Life*, 7(1):3–32.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397.

Smith, K. (2001). The importance of rapid cultural convergence in the evolution of learned symbolic communication. In *ECAL01*, Lectures Notes in Computer Science, pages 637–640, Prague. Springer.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16:97–159.