The Evolution of Dialect Diversity

Introduction: From Dialect to Dialectology

Observations on dialect diversity have been recorded for thousands of years, including Old Testament stories and early writings and literature from around the globe. Language diversity remains the subject of much study today – largely in the related fields of socio-linguistics, historical linguistics and dialectology.

One key question is *why* is there so much difference in dialects. To some the question was irrelevant as diversity was somehow obviously a natural feature of human language, one not requiring much explanation – it simply was. Currently the question of why diversity should exist to the degree that it does has been taken quite seriously, with strong differences of opinion apparent.

On one side, there are various arguments that linguistic factors alone could not be responsible for language diversity, and hence there must be some specific reason for its emergence and degree. Some additional extra-linguistic factor which causes diversity where otherwise none would exist. The extent of dialect diversity, and the speed of its evolution caused Dunbar to remark that it “is so striking and so universal that it cannot be a simple accident of evolution: it must have a purpose” (Dunbar, 1996, page 158).

Opposing this is the view that the nature of language, how it is transmitted and learned and the role of contact between individuals, provides sufficient means to explain the origins of language diversity.

In this chapter we briefly set out some of the arguments for both positions, Concentrating on arguments based on the possible requirements of social factors to effect change and diversity.

With plausible and convincing arguments presented on both sides of the divide, it is hoped that the use of simulation-based models may help resolve the debate. We review a number of existing models, and present an overview of our own approach, including some recent work applying de Boer’s model of evolving sound systems (Chapter 3) to this problem. Using simulation-based models is not the only possible approach, and we also briefly describe some of the difficulties that may result from the use of more traditional mathematical modeling techniques.

We conclude that simulation models have much potential for the investigation of linguistic diversity and dialect. First, a brief overview of some characteristics of linguistic diversity, and a brief literature review.

Dialect and Linguistic Diversity

As stated, the study of dialects and language diversity is of interest to academics in a number of sub-disciplines of linguistics, as well as to academics in other, related, fields. A good starting point for an overview of some of the different approaches taken is the Cambridge Encyclopedia of Language (Crystal, 1987).
Crystal presents an overview of some key characteristics of human language diversity. One important finding of dialectology is that the boundaries between different dialects or languages are not always easy to define. Geographically close dialects may have sufficiently similar grammars or lexicons to allow speakers from each dialect to understand each other quite well, despite differences in the dialects. Geographically distant dialects may not be at all mutually intelligible, despite being at either end of a chain of dialects where every dialect is intelligible to speakers of the neighboring dialects, Figure 1. Several such dialect continua exist in Europe, blurring the boundaries between the different languages within the Germanic, Scandinavian, Romance and Slavic language groups.

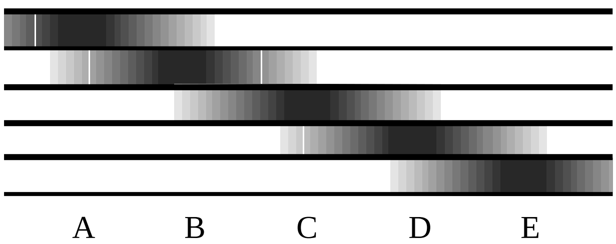

Figure 1. A schematic dialect continuum from dialect A to dialect E, showing some degree of mutual intelligibility between adjacent dialects (After Crystal, 1987, p25)

Further, the geographical boundaries between dialects may not be easy to determine. Dialects may differ in their lexical, semantic, morphological or phonological features. Sampling the language use of individuals in some area results in a map, on which boundaries, termed *isoglosses*, may be drawn to show where language use is distinct on either side according to a particular linguistic feature.
It may be expected that these lines will be largely coincident, forming clear dialect boundaries. Often, however, the boundaries are not even nearly coincident, as individuals near a boundary may use differing mixtures of lexical and grammatical items from the surrounding major dialects, Figure 2. Only when viewed at a more distant scale is it possible to determine distinct dialect areas – with poorly demarcated boundaries between them.

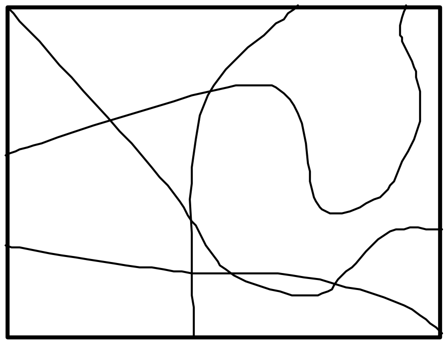

Figure 2. Dialect boundaries (After Crystal, 1987, p28)

As if this picture were not complicated enough, it is also widely recognized no two individuals use language in quite the exact same way – in a sense every individual speaks their own particular dialect, or *idiolect*. What is viewed as a dialect is merely some norm derived by sampling many idiolects. This, as we shall see, is highly significant when examining or attempting to explain linguistic diversity.

Sharp linguistic boundaries that do exist often coincide with significant geographical boundaries or with strong cultural boundaries (c.f. Chambers, 1980, p111). Where such boundaries exist, limiting the interactions of individuals across the divide, there may be many coincident isoglosses splitting the sides and a resultant lower degree of mutual intelligibility.

The aim of modeling language diversity is not to reproduce these characteristic patterns, but to examine what processes may give rise to them, and before we proceed to build models we must examine the explanations which already exist.

In recent years much of the research has centered on social variations in language use, propelled by the impact of the work of Labov, who detailed significant quantitative differences in language use by individuals according to social groupings (Labov 1972). This fuelled the growth of the field of socio-linguistics. Within socio-linguistics the predominant view of why such variation exists is that it serves some social function, the benefits provided leading people to learn and use particular variants over others (for example, Chambers, 1995, Trudgill and Cheshire 1998).

In contrast, in historical linguistics the prevalent view has been that language changes are due to pressures internal to language as a system – most famously with Grimm's 'Sound Laws' (Grimm, 1822). Such a view holds that changes in one part of the system may lead to increased opportunities for further change. The internal structure of a language, rather than the requirements of the speakers, is seen as the major force driving change. In both cases it appears that the favored explanation is heavily influenced by the different forms of evidence and phenomena that each discipline studies.

For example, two opposing arguments are presented by Milroy (1993) and Lass (1997), skeletons of which are repeated here.

Milroy's argument emphasizes that the best linguistic systems, for optimal communication, would be ones where every speaker used the same language. The differences in dialect are not just learned but actively maintained by speakers, where otherwise it would be expected that the dialects would merge and unify. Therefore, the differences must be of some use to the language users.

Lass' reply is unusual in historical linguistics literature, in its extensive use of evolutionary theory. It also gives a very central role to the idiolect. The language used by a single individual is considered to be a member of a hyper-variable species. With no two idiolects the same, language norms are simply averages of the different language features in use. Linguistic 'junk' and redundancy allows variability without serious loss of communicative ability. Another result of this is that learners are not required to learn perfectly the existing language norms.

It is not easy to resolve this argument using the available linguistic evidence. To prove either theory conclusively, language and dialect evolution must be tested in a population where the precise effects of social influence on language change are well understood.

In this chapter, we hope to show that simulation models are useful in bypassing this problem, and for providing reliable supplementary evidence. Simulation models are not the only technique available, however. In the next section we give an overview of other modeling approaches and argue that simulation models provide a more flexible and powerful means for testing theories explaining the emergence of dialects.

Approaches to Modeling Dialect Diversity

Here we very briefly review some of the different modeling approaches that have been used to study the development of linguistic diversity. This is not a thorough critique of the different approaches but, it is hoped, it provides some justification for the use of simulation micro-models in attempts to better understand the evolution of dialects.

Mathematical models may be a better approach than the use of simulation to describe the process of language change and the evolution of dialects. Pagel (2000) presents a mathematical model for examining the rate and pattern of linguistic evolution. The equations presented are useful for describing the degree of diversity that exists or the rate of diversification, but not for explaining what causes diversity in the first place.

Pagel's equations are useful in comparing the rate at which languages evolve into other, distinct, languages, and the rate of the linguistic diversification within different language groups. The model developed does not allow for close investigation of particular changes, or of what happens within languages as they develop, presenting a more global picture instead.

This is quite different to the approach of Cavalli-Sforza and Feldman (1978). Their theory of Cultural Evolution, although intended more generally for modeling the evolution of cultural traits, can be applied to the evolution of language. They present a mathematical model that is, in essence, quite simple – as summation of the influences on as single individual determines which traits are acquired by that individual. With the presence of mutation, innovations may occur and over time the cultural traits in the population may evolve – particular traits succeeding in the population at large or otherwise.

The possibility that this model might also be used to describe language evolution is emphasized by similarities it has with the description of the action of Social Networks on language (Milroy, 1980), appearing to be a mathematical formulation of the same.

To actually use this model requires the use of computer simulation – the very large number of calculations required would make the use of this model infeasible on all but the smallest of populations with the minimum of cultural traits otherwise.

Where models that abstract away the spatial distribution of speakers are unable to capture this important aspect of dialect evolution, mathematical models which seek to represent this are complex and not feasible without the aid of computers. An alternative approach, fully leveraging the power of computers is to build a simulation model, and a number of such models have been built.

For example, Niyogi & Berwick (1995) use a dynamical systems based model to study language change. Even so, there would be some problems in extending or adapting their approach to study linguistic diversity. By considering the distribution of dialects only as a proportion of the population which uses it, no spatial information is used in the model. The effect of this is that the acquisition of grammars is not reliant on any spatial constraints, and there can be no comparison with the emergence of dialects across geographical or social distances in the evolution of human languages. Briscoe (2000) argues that a micro-simulation model has a number of advantages over Niyogi and Berwick's more mathematical approach, and such simulations with more realistic population models may even produce different results.

In the next sections we review some works that have applied the micro-simulation approach to the evolution of linguistic diversity and closely related problems.

Related Simulation-based Models

The emergence of 'dialects' has been observed in a number of Artificial Life and other micro-simulation models of language and signal evolution. In many cases the existence of dialects is noted but not investigated – research being focused elsewhere, such as the evolution of a signaling ability or the emergence of conventionalized signals (e.g. Werner and Dyer, 1991; Hutchins and Hazelhurst, 1995 #98; Livingstone and Fyfe, 2000).

In some cases even work that is not about language at all may be relevant. Axelrod (1997) presents a model to investigate the dissemination of culture through a spatially arranged population. In the model, neighboring sites may interact if they have at least one cultural trait in common, and as sites interact they share traits and slowly converge. Eventually a stable distribution emerges where a limited number of groups survive, within each group all sites having identical sets of traits, and no traits being shared with sites belonging to neighboring groups.

Viewing language as a cultural trait, this is obviously relevant to the evolution of dialects. The results are at odds with observed phenomenon in human language, however (see figure 2, above, and related discussion).

In other cases the relevance of the results to the question of the origins of linguistic diversity is clearer. For example, Kirby (1998) builds a simulation model to show that Universal Grammar constraints may not be innate constraints at all, but merely the outcome of learning over time leading to a reduced set of surviving grammars. In this work Kirby shows results indicating the existence of geographically distributed dialects of grammars.

Maeda and Sasaki (1997) examine the effect of language contact. The results show subsequent language reorganization, but after this dialect diversity is absent from the population – again, not results that contrast with those observed in the real world.

Other related work has looked at the process of language change, apart from the question of dialect. Steels and Kaplan (1998) demonstrate how various linguistic and extra-linguistic errors can lead to continued language change. While natural language errors are somewhat more systematic than the random errors introduced in this model, the model successfully demonstrates the large influence such errors may have on language innovation.

A similar, artificial neural network based, model is presented by Stoness and Dircks (1999). In this model, it is found that noise is not required to maintain competition between forms. This appears to be due to the networks learning either one of two similar signals for particular internal meanings – an ability not present in the Steels and Kaplan model, where similarities in the lexical forms are ignored by the agents.

There yet remain a further number of works that explicitly explore the origins and nature of linguistic diversity, and these we consider next.

Simulation Micro-Models of Dialect Diversity

Arita and Taylor (1996) present what is possibly the first attempt to explain the origin of linguistic diversity using a micro-simulation model. They hold that it is the spatial distribution of individuals that is the key factor in the emergence of dialects. While this is a plausible position, it is not strongly supported by the model, which relies on *genetic* mutation for the emergence of linguistic diversity. Language is inherited, with mutation producing diversity and learning leading to increased convergence. If the spatial distribution of speakers is indeed a factor in the emergence of dialect diversity, then it must be able to work when the only means of language transmission is through learning – as it is for human language

Innate language is again used by Arita and Koyama (1998), in their investigation into the evolutionary dynamics of vocabulary sharing. Mutation rate is again identified as being an important factor in the emergence of diversity in the vocabularies, but without an identified linguistic equivalent. The degree of vocabulary sharing is also related to the availability of resources – rather than vocabularies, it is cooperative strategies that are being evolved here, as evidenced by cases where the evolved communication strategy is not to communicate at all.

A significant contribution to the use of micro-simulation for exploring the emergence of linguistic diversity has been made by Daniel Nettle in a series of papers. Nettle also sees learning as a force for convergence, and argues that additional, socially motivated, factors are required for the emergence of dialect diversity.

First, in Nettle and Dunbar (1997) a model is developed which shows how dialects may be used to indicate group membership, and how such a marker may be used in the evolution of cooperation. Groups of cooperative agents are able to resist invasion from non-cooperative individuals. This is used as the basis for an argument that dialects emerged for this *reason* – something that we argue against later in this chapter.

Nettle presents two further models that support his arguments that social status and social functions of dialect differences are pre-requisites for the emergence of dialect diversity (Nettle, 1999a, Nettle, 1999b).

The model presented in Nettle, 1999a, arranges language learners into a series of small groups, the language used consisting of a model of a vowel sound system. Learners pass through five life-cycle stages, and all language acquisition occurs during the first stage, where the new language agents learn from the other agents in the same village. Each group contains four individuals at each of the life-cycle stages (twenty in total at any time). After the fifth life stage, the elderly are replaced by a new set of infants. The infants each learn a sound system according to the set of sound systems in use by the existing group members, plus a small amount of noise. After this, all the individuals are 'aged' one stage. No learning occurs after the first stage.

It is seen that, unless the groups are completely isolated from one another, diversity does not emerge. Adding in social status changes the findings significantly. Each individual has a 25% chance of gaining high social status after the first life stage. Learners only learn language from those individuals with high status within the village. Otherwise, for any vowel, the sound learnt is the average formant frequency values used by all of the adults in the population for that vowel, plus a slight perturbation due to noise.

With social status included small differences between groups may become magnified over time and it is found that contact between groups no longer eliminates diversity.

The model presented in Nettle, 1999b, has many major differences, but retains the same agent life-cycle, where agents pass through five life-stages, before being replaced by new learners. Again, learning only occurs during the first life stage. Apart from this there are few similarities. Inspired by social impact theory (Latané, 1981), there are no sub-groups within this model, all agents existing on a single spatial array. Instead of learning vowels, agents acquire one of two grammars, p or q. In determining which grammar a learner acquires, the impact of all the surrounding grammars is calculated. This forms a sum of all the surrounding grammars, weighted by distance. Then, if the result is in favor of one grammar, that is the grammar acquired.

Several factors may be varied in this model, but the general finding is that sustained diversity requires that social status exerts a very large influence upon the acquisition of grammar.

These last two models each have design features which lead directly to these results. In the former, vowels are learnt by an explicit averaging of the vowels in use already in the local group. In the latter, the impact measurement and forced selection of a grammar from one of two distinct grammars – without the possibility of acquisition of elements of different grammars – is a form of threshholding.

Nettle argues in his work that the effects of averaging and thresholding would work to stifle diversity, were it not for the effect of social status, and uses these models as demonstrations. He then uses models in which these are enforced by the language acquisition rules he has built in. It is not proven that under more realistic learning conditions, where language is acquired as the result of many interactions or where there is a possibility of learning grammars that are different but compatible with surrounding grammars, that averaging or thresholding will prove to be the barrier to diversity that Nettle argues they are.

In the next section we present our own model, which was developed after an earlier model of the co-evolution of language and physiology (Livingstone and Fyfe 1998, 2000). The acquisition of signals occurs over many stochastic interactions between the signal learners, and as we shall see the results are quite unlike those described by Nettle.

Emergence of Dialects in Spatially Organized Populations of Simple Signal Learning Agents

We observed the existence of what appeared to be dialects in some of the results of our earlier model of the coevolution of physiology and language (Livingstone and Fyfe, 1998, 2000). The model used does not include any apparent social or adaptive benefits which might give rise to such diversity, and the result was in an unexpected result which was not pursued.

That the result was at odds with other work, claiming the need for social function, gave impetus for further investigation. To begin with, we used essentially the same model as out earlier work, removing some features no longer required (principally removing evolution of the agents themselves – as now it only the evolution of the language used by the agents that we wish to examine) and adding others. However, the implementation of the individual language learner was not modified at all.

The following section details our implementation of language learners, after which we present the population model and results. In a later section we discuss and compare our results against those of Nettle.

A Simple Language Agent

The language used by the agents in our model is a greatly simplified one, the agents themselves capable of only a small repertoire of signals. A very small set of arbitrary 'meanings' exists, common to all agents. The agents produce conventionalized signals to indicate the current meaning, and listener agents attempt to interpret what the original meaning was from the received signal. Each agent participates in many interactions, sometimes as signal producer, sometimes as listener.

An agent is implemented by a fully connected Artificial Neural Network (ANN), with two layers of nodes – three internal state nodes to hold the 'meanings' and three signal nodes (plus a signal bias node).

The internal state is a sparse bipolar vector (+ or –1 at each node, only one being set to +1 for any one meaning). The signals are non-sparse bipolar vectors (arbitrary vectors with + or –1 at each node). This representation allows three possible meanings and 2^3 (= 8) possible distinct signals. This provides some degree of redundancy in the representation of meaning, the significance of which becomes clear when viewing the results of the model.

In signal production, the meaning is presented to the internal nodes, and fed-forward through the ANN to generate a signal for that meaning. Each output is thresholded to a bipolar value. Signals are interpreted in a similar way – the incoming signal is presented at the signal layer and fed-back to generate a meaning vector. A winner-take-all comparison at the internal state layer determines which one of the nodes has the greatest activation, and this node is set to +1, the remaining to –1.

Using a standard ANN learning algorithm to train the signal production of the agents would allow the agents to learn to use the same signals for given meanings, but would not train the agents to use *distinct* signals for each meaning (c.f. Fyfe and Livingstone, 1997).

Oliphant (1997) overcomes this problem by using the signal production behavior of the population to train the signal reception behavior of the agents and the signal reception behavior of the population to train the transmission behavior. We make use of a similarly inverted training algorithm for our ANN.

Similarly to Steels' 'naming game' (Steels, 1996), one agent takes a turn as a teacher, another as a student. A meaning is presented to the teacher, which then produces a signal.

This signal is presented to the signal layer of the student and fed-back to produce a generated meaning. The difference between the original meaning, *x*, and the generated meaning, *x'*, us used for updating the weights. This is shown in the equation below.

$$\Delta w_{ij} = h\left(x_i - x_i'\right) y_j$$

As is shown, no learning only occurs when the learner misclassifies the signal.

Spatially Organized Populations of Language Learners

As well as using a very simple ANN to represent each agent – each with six nodes, excluding bias – we use a very simple population structure.

Agents are placed in a single row. The ends of the row are *not* connected – so with the exception of the agents at the row ends, each agent has two immediate neighbors. Communication between agents – for learning or for evaluating the success rate of the learned signaling schemes – is limited by pre-determined neighborhoods based on distance. An agent may communicate with any other agent in its neighborhood.

As an example consider the following. During the training period, agent A has been selected to be the student in one round. The agent which will act as teacher is selected from the agents within the neighborhood of A. The chance that each of the agents within the neighborhood has of being selected is determined by a normal-distribution curve centered on A. Agent B, immediately next to A, is more likely to be selected than agent C, several positions removed.

This creates a population where communication is localized but which does not have any explicit group boundaries.

We also adopt a population aging structure considerably reduced from that used by Nettle. Agents pass through only two life stages, child and adult, opposed to five.

Children and adults populate separate rows of identical size. As a child, agents learn only from adults – the position within each row being used in the determination of the neighborhoods. After training, the existing adults are removed, the current child population is aged and a new row of children populated. A variation on this, where the children have an additional short period of signal acquisition where they learn only from their peers, leads to no significant differences in the recorded results.

We have tested this model using a variety of initial and environmental conditions. The principal factors that are varied between different tests are:

- Neighborhood size. By varying the standard-deviation of the normal curve which is used to determine the likelihood of a neighbors selection, the effective neighborhood size can be varied from one that only includes the immediate neighbors to one that includes the entire population. The default neighborhood includes only a small number of agents to either side of the currently selected agent.

  Additionally, the neighborhood can be changed to cover the entire population with a uniform distribution.
- Initial conditions. The first generation has no previous generation to learn from, and so must be initialized somehow. It is possible to initialize the population such that all agents use the same signaling scheme. Alternatively, and by default, random values may be used. Training of the following generation can then begin (or of the current, if peer-to-peer learning is to be used).
- Noise. By default the communication between agents is noise free and the perfect replication of signaling schemes across generations is possible – something that is not possible in the human language learning for several reasons. First, the linguistic evidence presented to children is insufficient to allow perfect replication. It is also possible that errors in language use or comprehension, on the part of learners or the speakers who provide them with evidence , may help drive language change (Steels and Kaplan, 1998), or that individual variations in language use (Chambers, 1995) might have some impact.

  In the model we use noise to represent these extra forces that may act on language replication from generation to generation. With a small probability, noise may cause individual bits in the signal to be inverted (+1 or –1 or vice versa).

Additional factors that can be varied include the population size, the number of meanings and/or signals used or the learning rate used in the ANN learning rule.

Results

Here we present only an overview of the results and highlight some advantages and disadvantages of the approach used in the model – most of the results have been previously presented in more detail elsewhere (Livingstone and Fyfe, 1999a and b).

Under the default conditions distinct dialects are readily apparent in each generation of the population, and the dialects themselves evolve over generations, gaining and losing speakers eventually being replaced by other dialects.

Within the populations dialect continua form whereby chains of mutually intelligible signal schemes span across the population, linking dialects that may, or may not, be mutually intelligible (as shown in Figure 1). Between distinct dialects, the point where signal use changes tends to be different for each internal state, forming indistinct boundaries between the dialects (similar to the position presented in Figure 2).

In some cases convergence is observed, where the whole population learns the same signaling scheme. Once this occurs (in the cases where this has been observed, it happens only after a very large number of generations has passed – on the order of several hundreds of thousands of generations) signal diversity is unable to return to the population. Similarly, when starting with a converged signaling scheme, diversity never emerges.

The addition of a small amount of noise is sufficient to introduce variation into the signaling schemes, which rapidly, within tens or hundreds of generations, creates diversity and distinct dialects across the population. Noise is also able to prevent convergence occurring.

Increasing the neighborhood size tends to reduce the number of distinct dialects. At the limit, where any agent might interact with any other with uniform probability, a single global dialect emerges. Some variation exists inside this dialect, with different sets of signals being used for each meaning. For a given meaning an agent may use any one of the signals from the set, but each agent is able to correctly interpret the signals in use. This result is in accordance with the view that dialect differences flow automatically from any non-uniformity in the interactions of speakers in a community (c.f. Sweet, 1888, para 189).

By using the three bit binary representation for signals, eight different signals were possible. The advantages this provides for visualization of the results (as described below) prevented us from changing this in different runs.

Review and Evaluation

Our model has some important advantages, as well as limitations.

The signal representation allows a signal to be plotted as a color pixel, and this makes visualization of the results very easy – not just for comparing the dialects within a generation, but across many hundreds of generations. More important is the ease with which the results can be compared to the patterns of human linguistic diversity, and the similarities with these observations.

This increases our confidence in the claim that social motivation, a factor absent in our model, is not a requirement for the evolution of diversity in human language.

A severe limitation of the model is the constrained representations of signals and meanings. Meaning are pre-determined, and limited by the three possible internal states. The signals are similarly restricted, this time to eight possible discrete forms.

The ability of agents to interpret two or more signals as having the same meaning is an advantage, however. Similar signals are more likely to be interpreted the same, and this brings a degree of redundancy to the signal schemes, a feature absent in most other models. The importance of this is supported by the weight given to the role of linguistic junk by Lass (1997), who considers redundancy to be a key feature of language in the processes of language change.

However, the results of this model are opposed by results from three different models used by nettle, even before other models are considered.

Additional corroboration for our results is highly desirable, and this is what we seek to gain, as detailed in the remainder of this chapter.

Emergence of Dialects in Spatially Organized Populations of Vowel Learning Agents

Corroboration can be gained in the same manner as used by Nettle – by using alternative models that provide qualitatively similar results. However, rather than develop a new model from scratch, we chose to base our new model on some existing work, de Boer's model of emergent vowel systems (see chapter 3).

We use an implementation of individual agents based closely on de Boer's original, adding a more structured population model. Initially we use a linear arrangement of agents – similar to the one used in our previous model.

The agents used here have some additional advantages, beyond simply allowing us to re-test our results. The signals exists in a continuous domain, in which there is a very wide range of possible signals. The number of signals an agent learns is not fixed, allowing for much more open-ended evolution of signaling schemes.

The phonological model used is quite realistic, and produces emergent vowel systems very like those that appear in human language. This may aid the 'believability' of our results as the individual sound systems can be closely compared to real sound systems, not just the overall patterns of diversity and change.

It is also possible to use the agents in a variety of population models, representing different social conditions, which makes the model potentially extremely flexible. These are not explored here, however.

A disadvantage of using the phonological model in comparison to the previous model is that visualization may not be quite so easy. In particular, a clear view of the dialect used by each speaker, when displaying the vowel systems used by the whole population, may not be achievable. This is not an insurmountable problem, as we shall see next.

de Boer's Model of Emergent Phonology

As in the model used previously, de Boer's agents learn through repeated interactions and learning. The model has been thoroughly documented, enabling an independent re-implementation to be attempted (de Boer, 1997, 2000, de Boer and Vogt, 1999). This was done, and testing determined that it performed qualitatively the same as de Boer's own implementation.

In de Boer's implementation, small populations of twenty vowel learners exist and interact with each other according to a uniform random selection. No form of spatial or social structure is used.

An example of an emergent vowel system from our implementation of the model is shown in Figure 3. Superimposed onto this are the approximate positions of the major vowel sounds of the English language.

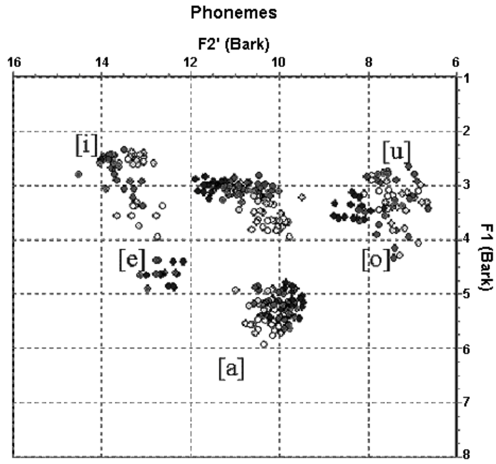

Figure 3, An emergent phonology. Clusters appear in areas of the phoneme-space where multiple agents have learned the same vowels. The approximate positions of some the major vowels of English have been superimposed on the graph.

Experimental Setup

The model was enhanced to include a larger population of agents, spread across a spatial array. 100 agents were arranged in a single row, the ends of which are not connected, as before. The algorithm for agent learning is not altered from de Boer's original, other than to enable the neighborhood-based selection of partners. Once an agent has been selected a partner is required for learning. The partner is selected from a position along the line on either side of the original, within a limit of ten agents distant. Agents near the ends of the line have as a consequence fewer other agents to communicate with. The selection of partners is from a uniform random distribution within the neighborhood.

Table 1 sets out the parameter settings for the run detailed in the following results. As before, a number of runs were performed for a variety of parameter settings, some of these are discussed below.

Parameters			
Population	100	Training rounds	25,000
Neighborhood size	10	Noise (%)	15

Table 1. Parameter settings for testing the evolution of dialects with the emergent phonology model

The large number of training rounds is used to provide a good chance of each and every agent receiving somewhere on the order of 200 training examples, or more.
The chosen neighborhood selection limit was determined by the size of the population divided by ten. This figure ensures that there is reasonable distance between the far ends of the population. The actual neighborhood size is double this, as agents can communicate with others on either side up to this distance.
In de Boer's work the effect of varying noise is well documented. Smaller noise values allow the emergence of vowels systems with more vowels than occur with larger noise values, which tend to have fewer, larger, clusters. The value we use is in the mid-range of values used by de Boer.

Results

Running the model produces what appears to be a rather messy result, shown in Figure 4, compared to the previous (Figure 3). This could represent a fairly noisy system with 4 or perhaps 6 different vowels, the clusters being indistinct and seemingly overlapping. By breaking the population into groups, and displaying only the different vowels used by agents within the groups, one or two groups at a time, a better picture of the vowel use can be gained.

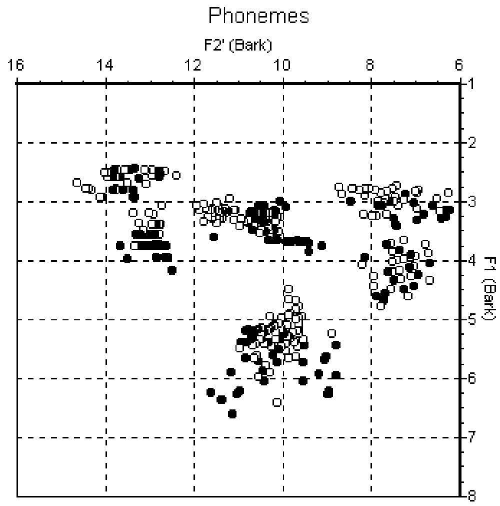

Figure 4. The emergent vowel system of the population.

The diagrams in Figure 5 show the same emergent vowel system. The population has been split into five arbitrary continuous groups. The first twenty agents if the population form the first group, the next twenty are placed in group two, and so on.

Note, these groups exist only for the purpose of displaying the results – and no related boundaries exist during the negotiation of the phonology.

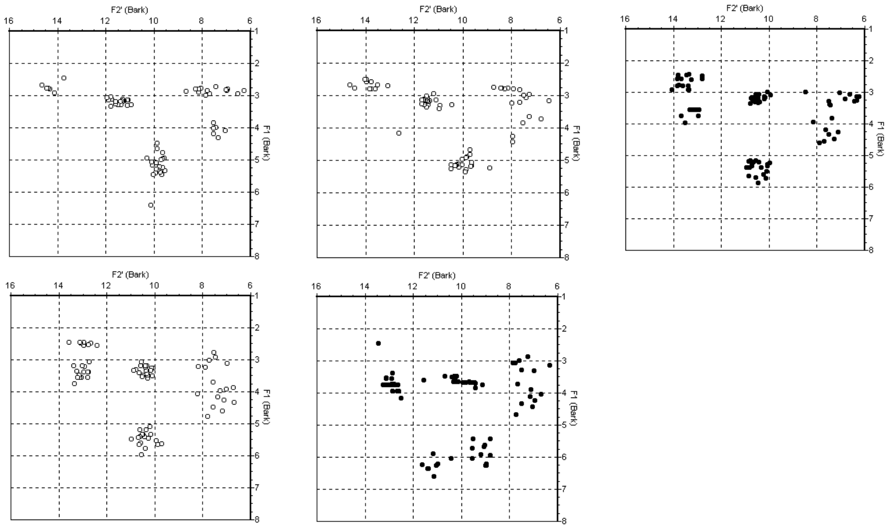

Figure 5. The emergent vowel system of the population. Each diagram shows the phonemes used by a different continuous sub-group of the population.

Some of the individual diagrams remain a little unclear – it is not always obvious whether one or other of the clusters represents one or two vowels. Even within a single group, there is some distance, and it is possible that different agents within a group have learned slightly different sets of vowels. The groups themselves have not been chosen with regard to how close the vowel systems of the individuals within the group are – rather, the groups are a completely arbitrary division of the population. As such we should not expect that within groups the vowel systems should be very similar or that there must be distinct differences between adjacent groups.

Nevertheless, it would appear that most groups have developed a four or five vowel system, which is largely shared amongst the agents within a group, with gradual shifts between the groups. Figure 6 emphasizes the differences that exist across the population. In this figure, the phonemes used by the agents of the first and the last groups are shown together (as white and black dots respectively).

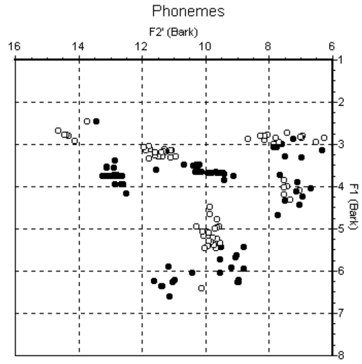

Figure 6. The emergent vowel system of the population.

As with the previous work, a dialect continuum has emerged in the population. Minor changes exist within neighborhoods, allowing successful communication therein. Across the population, more major shifts and differences exist. Although the model is entirely unrelated the qualitative result is the same – the negotiation of a communication scheme/phonology within a population, where neighborhoods limit interactions, gives rise to emergent dialects without any requirement for any need or motivation to create the different dialects.

Future Goals and Research

In this chapter we have shown how different micro-simulation models have been used to explore issues in the evolution of linguistic diversity – primarily in the requirement of social motivation for the emergence of dialects.

Yet, more detailed work is required before this issue is settled. Such is the evidence of the extent to which dialect can affect social behavior, that the idea that not just some, but *all*, language change and diversity is socially motivated will not be easily dispelled.

The arguments are unlikely to be progressed by building further models which find the same results as found already – whether for or against the need for motivated change. Instead, future work would be better to build on the existing models.

A direction is suggested by the idea of a language ecology.

Language ecology (Haugen, 1971, Mülhäusler, 1996) considers languages as existing in an environment, the environment being the human society in which they exist. Different societal organizations and population mixes form different ecologies, these being the conditions under which languages evolve. Mufwene (1997) extends this to consider also the particular mix of languages and variants present, and their typological features, adding the structural states of languages to the list of relevant ecological conditions.

Artificial language ecologies may be modeled in a micro-simulation by embedding phonological or grammatical language learners in models of different social structures. Modifying the existing language acquisition rules to incorporate social motivation may lead to different simulated outcomes. If previously observed language changes, say the results of language contact under particular conditions, can be successfully reenacted with such models, then the experiments might aid in our understanding of the influence social motivation has had on the evolution of dialects and language diversity.

Conclusions

We hope to have shown, in our own work, that social function is not required for the evolution of linguistic diversity as well as to have indicated what will be a fruitful direction in which to take future model based research. Rather than elaborate on this, we conclude this chapter with a brief consideration of some of the issues important in the development of simulation models of language diversity.

As also noted by Briscoe (2000), implementing language learning by means of a single update, without modeling the individual learning interactions which occur as part of a stochastic process, can lead to results that differ quite significantly. This may lead to quite opposite conclusions, as has been shown by the our work versus that of Nettle. Language learning involves many stochastic events, and the learning environment is different for every learner – and it makes sense to capture this in our models. Contact has a significant role to play in cultural evolution of language, not just contact between speakers of different languages but the contact between speakers of different lects of the same language, a role often underestimated.

Micro-simulation is the only feasible approach whereby models can be developed which include the effect of contact on the evolution of languages.

While it may be appropriate to build models with complex linguistic ecologies, it may be less so to use models in which the appropriate use of dialect brings direct adaptive benefits. As in seen in Arita and Koyama (1998), where dialect diversity becomes subsidiary to the evolution of (non-)cooperative behavior, the introduction of genetic evolution can produce results highly unlikely in the cultural evolution of language. As genetic evolution has little role to play in the evolution of languages, models would be best to eliminate it entirely.

Finally, it is worthwhile to consider what the reason for using a simulation-model is in the first place. There is much evidence of language change gathered in linguistic atlases, and a great many studies of local and national dialects the world over. Why do we need artificial evidence?

Simulation allows the development of models which can incorporate different aspects of society as well as different aspects of language and language learning. By testing the models under different conditions the role played by these can be investigated more closely in a model – it is not possible to toggle social influence on and off in real life, like it is in a model. If the real world evidence is able to support more than one interpretation, then maybe our simulated evidence can provide the balance of evidence required to come down on one side.

References

Arita, T. and Y. Koyama (1998). Evolution of Linguistic Diversity in a Simple Communication System. Published in the Proceedings of Artificial Life IV, UCLA, C. Adami, R.K. Belew, H. Kitano, C.E. Taylor (eds.), MIT Press.

Arita, T. and C. E. Taylor (1996). A simple model for the evolution of communication. Evolutionary Programming V, L. J. Fogel, P. J. Angeline, T. Bäck (eds.), MIT Press.

Axelrod, R. (1997). “The dissemination of culture: A model with local convergence and global polarization.” Journal of Conflict Resolution **41 (1997)**: 203-226.

de Boer, B. (1997). Generating Vowel Systems in a Population of Agents. Presented at the Fourth European Conference on Artificial Life, ECAL 97, Brighton, UK.

de Boer, B. and P. Vogt (1999). Emergence of Speech Sounds in a Changing Population. Published in the Proceedings of the Fifth European Conference on Artificial Life, ECAL 99, Lecture Notes in Artificial Intelligence, Volume 1674, D. Floreano, J.-D. Nicoud, F. Mondada, (eds.) Springer.

de Boer, B. (2000). Emergence of sound systems through self-organization. The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form. C. Knight, M. Studdert-Kennedy and J. R. Hurford, Cambridge University Press**:** 177-198.

Briscoe, E. J. (2000). Macro and Micro Models of Linguistic Diversity. Presented at The 3rd International Conference on the Evolution of Language, Paris, April 3rd-6th.

Cavalli-Sforza, L. L. and M. W. Feldman (1978). “Towards a Theory of Cultural Evolution.” Interdisciplinary Review **3**: 99-107.

Chambers, J. K. and P. Trudgill (1980). Dialectology, Cambridge University Press.

Chambers, J. K. (1995). Sociolinguistic Theory, Blackwell.

Crystal, D. (1987). The Cambridge Encyclopedia of Language, Cambridge University Press.

Dunbar, R. I. M. (1993). “Co-evolution of neocortex size, group size and language in humans.” Brain and Behavioural Sciences **16**: 681-735.

Fyfe, C. and D. Livingstone (1997). Developing a Community Language, presented at the 4th European Conference on Artificial Life, ECAL 97, Brighton, UK.

Grimm, J. (1822). Deutsche Grammatik, Part I: Zweite Ausgabe. Göttingen, Dieterich.

Haugen, E. (1971). The Ecology of Language. In Ecology of Language, Essays by Einar Haugen. A. S. Dil (ed). Stanford, CA, Stanford University Press**:** 325-339.

Hutchins, E. and B. Hazelhurst (1995). How to invent a lexicon: the development of shared symbols in interaction. Published in Artificial Societies: The computer simulation of social life. N. Gilbert and R. Conte (eds.), UCL Press.

Kirby, S. (1998). Fitness and the selective adaptation of language. Approaches to the Evolution of Language. J. R. Hurford, M. Studdert-Kennedy and C. Knight, Cambridge University Press**:** 359-383.

Labov, W. (1972). Sociolinguistic Patterns. Philadelphia, University of Pennsylvania Press.

Lass, R. (1997). Historical Linguistics and Language Change. Cambridge, UK, Cambridge University Press.

Latané, B. (1981). "The psychology of social impact." American Psychologist **36**: 343-365.

Livingstone, D. and C. Fyfe (1998). A Computational Model of Language-Physiology Coevolution. Presented at The 2nd International Conference on the Evolution of Language, London.

Livingstone, D. and C. Fyfe (1999a). Modeling the Evolution of Linguistic Diversity. Published in the Proceedings of the Fifth European Conference on Artificial Life, ECAL 99, Lecture Notes in Artificial Intelligence, Volume 1674, D. Floreano, J.-D. Nicoud, F. Mondada, (eds.) Springer.

Livingstone, D. and C. Fyfe (1999b). Dialect in Learned Communication. Presented at the AISB '99 Symposium on Imitation in Animals and Artifacts, K. Dautenhahn and C. Nehaniv (eds.), Edinburgh, The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Livingstone, D. and C. Fyfe (2000). Modeling Language-Physiology Coevolution. In The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form. C. Knight, M. Studdert-Kennedy and J. R. Hurford, Cambridge University Press: 199-215.

Milroy, J. (1993). On the social origins of language change. In Historical Linguistics: Problems and Perspectives. C. Jones (ed.), London, New York, Longman.

Milroy, L. (1980). Language and social networks. Oxford, Basil Blackwell.

Mufwene, S. S. (1997). What research on creole genesis can contribute to historical linguistics. In Historical Linguiustics 1997. M. S. Schmid, J. R. Austin and D. Stein (eds.). Amsterdam, John Benjamins.

Mülhäusler, P. (1996). Linguistic ecology: Language change and linguistic imperialism in the pacific region. London, Routledge.

Nettle, D. (1999a). Linguistic Diversity, Oxford University Press.

Nettle, D. (1999b). "Using social impact theory to simulate language change." Lingua **108**: 95-117.

Nettle, D. and R. I. M. Dunbar (1997). "Social Markers and the Evolution of Reciprocal Exchange." Current Anthropology **38**(1): 93-99.

Niyogi, P. and R. C. Berwick (1995). A Dynamical Systems Model for Language Change, A.I. Memo No. 1515 (C.B.C.L. Paper No. 114), Massachusetts Institute of Technology, March 1995.

Maeda, Y., T. Sasaki, et al. (1997). Self-Reorganizing of Language triggered by "Language Contact". ECAL 97, Brighton, UK.

Oliphant, M. (1997). Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language, University of California, San Diego.

Pagel, M. (2000). The history, rate and pattern of world linguistic evolution. In The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form. C. Knight, M. Studdert-Kennedy and J. R. Hurford, Cambridge University Press**:** 391-416.

Steels, L. (1996). Emergent adaptive lexicons. Published in From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour, P. Maes, M. Mataric, J.-A. Meyer, J. Pollack and S.W. Wilson (eds.), MIT Press, Cambridge, MA.

Steels, L. and F. Kaplan (1998). Stochasticity as a Source of Innovation in Language Games. Published in the Proceedings of Artificial Life IV, UCLA, C. Adami, R.K. Belew, H. Kitano, C.E. Taylor (eds.), MIT Press.

Stoness, S. C. and C. Dircks (1999). Investigating Language Change: A Multi-Agent Neural-Network Based Simulation. Published in the Proceedings of the Fifth European Conference on Artificial Life, ECAL 99, Lecture Notes in Artificial Intelligence, Volume 1674, D. Floreano, J.-D. Nicoud, F. Mondada, (eds.) Springer.

Sweet, H. (1888). The History of English Sounds. Oxford, Clarendon Press.

Trudgill, P. and J. Cheshire, Eds. (1998). The Sociolinguistics Reader, Volume 1: Multilingualism and Variation. London, Arnold

Werner, G. M. and M. G. Dyer (1991). Evolution of Communication in Artificial Organisms. In Artificial Life II. C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen (eds.), Addison-Wesley.