SIMULATION MODEL FOR THE EVOLUTION OF LANGUAGE WITH SPATIAL TOPOLOGY

CECILIA DI CHIO
Department of Computer Science, University of Essex
Colchester, CO4 3SQ, UK
cdichi@essex.ac.uk

PAOLO DI CHIO
Dipartimento di Sistemi e Istituzioni per l'Economia, University of L'Aquila
L'Aquila, 67100, Italy
pdc@ec.univaq.it

In this paper, we present an agent-based simulation model for the evolution of language. This is based on a previous model proposed by the authors and inspired by Nowak's simplest mathematical model. We extend our previous work with the introduction of a significant characteristic: a world where the languages live and evolve, and which influences interactions among individuals. The main goal of this research is to present a model which shows how the presence of a topological structure influences the communication among individuals and contributes to the emergence of clusters of different languages.

1. Introduction

The genetic and linguistic systems follow two parallel evolutionary trajectories, i.e. they co-evolve (Cavalli-Sforza, 2000). Isolation, either social or geographic, causes independent evolution and genetic differentiation. The same happens with languages: isolation reduces cultural exchanges and the languages of isolated populations becomes more and more different. The study of the emergence of these isolated clusters of languages has been the motivation for our research. In order to achieve our goal, we have used the evolutionary theory of games, and in particular the theory of *evolutionary language games*, together with *agent-based simulation models*.

The theory of evolutionary language games arises from the union of *evolutionary game theory* (Maynard Smith, 1982) and the theory of *language games* (Wittgenstein, 1953).

The simulation model we present is an extension of our previous model (Di Chio & Di Chio, 2005), which was inspired by Martin Nowak's simplest mathematical model (Nowak & Krakauer, 1999), (Nowak, Plotkin, & Krakauer, 1999), (Nowak, 2000). In this new model, we introduce a characteristic that we think is crucial for a more realistic simulation of the evolution of language: a topological structure (the *world*) where the (individuals who speak different) languages live and evolve. This structure influences the interactions among the individuals and contributes to the emergence of clusters of different languages.

This model is an agent-based simulation. These kinds of models are characterised by a certain number of agents which can control their own behaviour according to their perception of the environment they live in. The goal of an agent-based simulation is to create agents which are able to interact with the environment in an intelligent way. For this reason, these simulations are widely used in modelling artificial intelligence and artificial life. Examples of agent-based simulation systems are cellular automata, ant systems (Bonabeau, Dorigo, & Theraulaz, 1999) and particle swarm systems (Kennedy & Eberhart, 2001).

The rest of this paper is organised as follow. In the next section we describe our simulation model. In section 3 we present some results, and we conclude in section 4.

2. The simulation model

Nowak's mathematical model describes quite accurately the emergence of a linguistic system but, at the same time, it is based on simple assumptions. In particular, there is no environment able to influence the communication among the individuals. Since isolation is one of the main reasons for the differentiation of languages and the emergence of linguistic groups, we developed a simulation model adding to Nowak's a *world*: an environment with a topological structure where the (individuals which speak the) languages live and evolve.

The world where the agents will live is a 2-D discrete grid whose x and y dimensions are exogenous parameters and which is topologically equivalent to a torus. Agents represent individuals as well as languages and do not move around, but are in a fixed *cell* location. They produce offspring, which will be generated and put into the environment according to a certain set of rules. As in the previous model (Di Chio & Di Chio, 2005), the size of the population is constant in time and each new generation completely replaces the old one.

Following Nowak's notation, each language $\mathcal{L}$ is defined by a relationship between a finite set of n *objects*, and a finite set of m *sounds* (the *vocabulary* of the language). The matrices $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{Q}$ retain their meaning. The payoff of the language game between agents a_h and a_k will be written as

$$\pi(a_h, a_k) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (p_{i,j} q'_{j,i} + p'_{i,j} q_{j,i}) \qquad (1)$$

The similarities with the mathematical model end here. The computation of fitness, the generation of offspring and the positioning of the newborn agents, now take into account the presence of the world. Let us examine each of these issues in detail.

> The fitness function is modified in order to be influenced by the distances between the individuals, in such a way that the contribution to the fitness of the agent a_h is higher for closer individuals it plays with. This mirrors the real world situation, where communication is more likely to happen between individuals which are closest to each other (using some suitable metric).
>
> The number of offspring that each agent generates will be proportional to the agent's fitness, but the factor of proportionality changes. It is no longer the global fitness (the fitness of the whole population) but a "locally global" fitness, i.e. the fitness of a suitable neighbourhood of the generating agent. To avoid too abrupt a separation among agents, we adopt a fuzziness in the definition of neighbourhood, weighting the fitness of pairs of agents with a smooth function. This is the key point which (eventually) leads to the emergence of many groups of individuals, and therefore to many cluster of languages.
>
> We position the newly generated agents following one of two strategies: (a) put all offspring in (a *list* in) the same cell as the parent or (b) put offspring in *neighbouring* cells. These have been chosen to mirror a more (the latter) or less (the former) strong isolation process.

Under these assumptions, we no longer have the emergence of a single, (possibly) optimum language spoken by the whole population. Instead, what we observe is the emergence of a certain number of *clusters* of different languages. This happens because nearby agents have fitnesses high enough to generate offspring, which in turn will be near to each other.

More formally, let $d(a_h, a_k)$ be the euclidean distance between the agents a_h and a_k and $\rho(a_h, a_k) = e^{-d(a_h, a_k)}$ the function of d we will use to weight the payoffs. The fitness ϕ for a_h is given by

$$\phi_{a_h} = \sum_{k \neq h} \pi(a_h, a_k)\rho(a_h, a_k) \tag{2}$$

We then need to weight each individual's global fitness in a similar way. Thus, not only will the communicative ability be influenced by the distance, but also the number of offspring.

To compute the number of offspring, we have to take into account the "locally global" fitness. If $\Phi(a_h)$ is the global fitness relevant to the individual a_h, and A_h is a suitable neighbourhood of a_h, we have

$$\Phi(a_h) = \sum_{a_k \in A_h} \phi_{a_k} \tag{3}$$

The number of offspring s_{a_h} for a_h is proportional to the ratio between the individual's fitness and the global fitness, that is

$$s_{a_h} = n_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} \propto \frac{\phi_{a_h}}{\Phi(a_h)} \tag{4}$$

where n_{A_h} is the number of agents in A_h. We do not know A_h, but we can "fuzzify" it and write (for the global fitness)

$$\Phi(a_h) = \sum_k \phi_{a_k} \rho(a_h, a_k) \tag{5}$$

and for n_{A_h}

$$\tilde{n}_{A_h} = \sum_{k \neq h} \rho(a_h, a_k) \tag{6}$$

In each generation, the population size N is constant. Therefore, we have

$$\sum_h s_{a_h} = \sum_h n_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = N \tag{7}$$

whilst

$$\sum_h \tilde{n}_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = M \tag{8}$$

Thus (to retain population size N per generation) the actual number of offspring for each individual is given by

$$s_{a_h} = \frac{N}{M} \sum_{k \neq h} \rho(a_h, a_k) \tag{9}$$

where N/M is a normalisation factor.

At each generation, the offspring of the same language will be close to each other, their fitnesses will be higher, and they will leave more offspring. This is a phenomenon which happens locally and therefore we expect to observe the process of language clustering.

Starting from a population of many different languages (i.e. from many different populations, each one made of just one language), the simulation shows how these languages spontaneously move (closer or further away) until the emergence of independent populations. This happens without any form of "artificial" constraint but thanks just to communication.

2.1. *Implementation of the model*

To implement the simulation model, we have used the *Swarm* platform (Swarm Development Group, 2000) and the *Objective-C* programming language, as we did for our previous model (Di Chio & Di Chio, 2005).

This simulation has a model swarm called `LangGameModelSwarm`. This creates the lists of the present, past and newborn languages, generates the offspring and manages the language game (the turns in the game for the languages). The agent `Language` deals with actions intrinsic to the language, such as creating the matrices $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{Q}$ and sampling $\mathcal{P}$, playing the language games (i.e. computing the fitness), computing its own coordinates in the world and calculating the distance between itself and the other agents. The agent `LangSpace` represents the world which contains the languages. Finally, there are two different observer swarms, which perform similar actions, the only difference being that one makes the observation in a graphical model through the use of a GUI and therefore also deals with the function to manage the graphical interface. For more details on the implementation see (Di Chio, 2004).

3. Results

We have conducted several experiments to ensure the robustness of the model[a]. The settings for the parameters of the simulation are summarised in table 1.

Table 1. Parameters settings.

Parameter	Value
(objects, sounds)	$(5, 5)$, $(10, 10)$, $(25, 25)$
Population size	100 individuals
Sampling parameter k	1, 4, 7, 10, 25
Generations	100
Iterations	20

We have run different experiments according to the positioning of the offspring in the world (neighbouring cells lookup or list) and whether the fitness is weighted or not with the distance.

The graphs in figure 1 show the results of the simulations (with the smallest vocabulary and sampling parameter values 1 and 25) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the population is replaced with neighbourhood lookup.

Those in figure 2 show the results of the simulations (with the same parameters as before) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the new population is positioned in a list.

The last two graphs (fig. 3) show the configuration of clusters in detail. In particular, we can observe that, if the replacement is with neighborhood lookup, it is possible to have clusters with more than one language, whilst if the population is positioned in a list, there is just one language in each cell.

---

[a] All the simulations have been run on a 2.4GHz Intel Pentium 4® CPU with 512MB RAM on the RedHat® Linux 9.0 operating system.

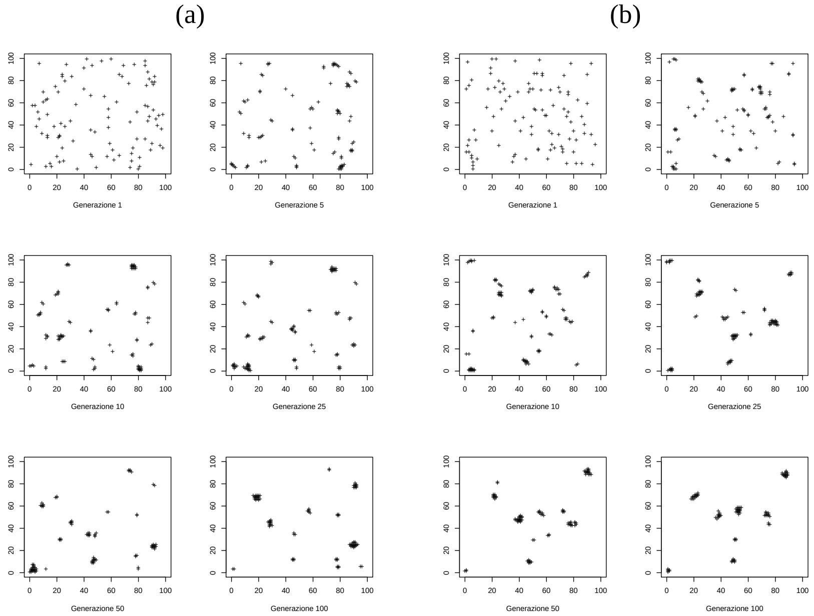

Figure 1. Simulation model with (objects,sounds) $= (5, 5)$ (a) $k = 1$ (b) $k = 25$. Distance influences both ϕ and Φ. Population replaced with neighbourhood lookup.

For a more comprehensive set of graphs, as well as a complete list of clusters and theirs characteristics, refer to (Di Chio, 2004).

3.1. *Analysis of the results*

As the simulation results show, it is clear how important the presence of a topological structure is for the behaviour of the languages. We can in fact observe, by varying parameters, the emergence of different clusters of different languages.

The replacement with neighbourhood lookup causes the clusters to continually evolving. This happens because, by positioning the new individuals in the cells around their parents, the dimensions of the cluster are continuously varying, and therefore the distance among individuals in different clusters changes from one generation to the other. These variations help in the emergence of new languages in new positions (i.e. positions different from the starting ones).

On the other hand, positioning the new population in lists is a way to clearly highlight the process of cluster creation. Since all the offspring of an individual are placed in the same cell, the spatial dimensions of the clusters are constant (and equal to 1 cell). Therefore, in this situation we will not observe the emergence of new languages in new positions, but only the disappearance of isolated (weaker) languages.

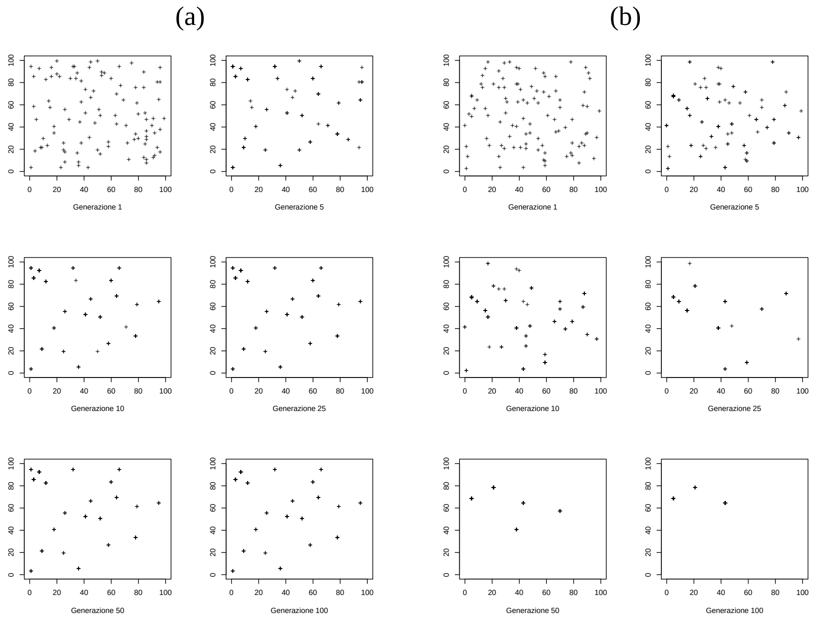

Figure 2. Simulation model with (objects,sounds) $= (5, 5)$ (a) $k = 1$ (b) $k = 25$. Distance influences both ϕ and Φ. Population positioned in a list.

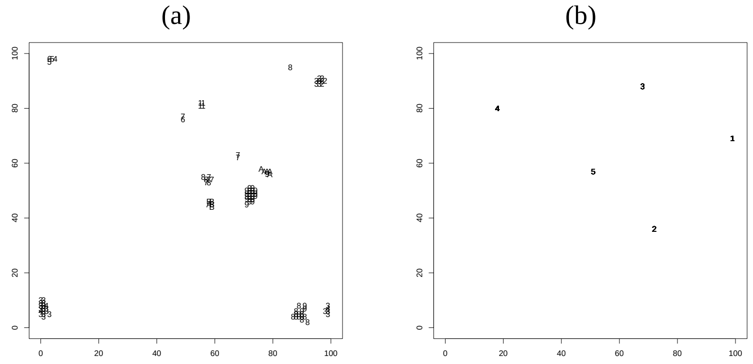

Figure 3. Simulation model with (objects,sounds) $= (5, 5)$ and $k = 7$ (a) Neighbourhood lookup, 11 languages and 14 clusters (b) List, 5 languages and 5 clusters.

4. Conclusion

We have presented a simulation model for the evolution of languages based on the theory of evolutionary language games. The model, inspired by a mathematical model due to the biologist Martin Nowak, adds a topological structure in which the languages live. We have then studied how clusters of different languages emerge and evolve in the world, thanks to the influence of the environment on the communication among individuals.

Our results have shown the emergence of different configurations, according to the parameters acting on the system, e.g. the influence of the environment on the offspring generation and the way that the new languages are introduced to the world.

There are a number of interesting future directions we would like to explore:

- allow multiple parents and overlapping generations (population size no longer constant);
- separate the individual from the language, allowing an individual to speak more than just one language;
- study other linguistic phenomena such as dialects or pidgin/creole languages;
- expand our model to let agents move around, much like in a particle swarm system.

References

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems.* Oxford University Press.

Cavalli-Sforza, L. (2000). *People, genes and languages.* University of California Press.

Di Chio, C. (2004). *Modelli di simulazione evolutiva per lo sviluppo del linguaggio.* (Tesi di Laurea, University of Roma ”La Sapienza”)

Di Chio, C., & Di Chio, P. (2005). *A simple simulation model for the evolution of language.* (Under revision)

Kennedy, J., & Eberhart, R. C. (2001). *Swarm intelligence.* Morgan Kaufmann.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge University Press.

Nowak, M. (2000). Evolutionary biology of language. *Philosophical Transactions: Biological Sciences*, *355*, 1615-1622.

Nowak, M., & Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 8028-8033.

Nowak, M., Plotkin, J., & Krakauer, D. (1999). The evolutionary language game. *Journal of Theoretical Biology*, *200*, 147-162.

Swarm Development Group. (2000). *Brief overview of swarm.*

Wittgenstein, L. (1953). *Philosophical investigation.* Blackwell.