arXiv:physics/0509018 v1   2 Sep 2005

# Phase transition in a sexual age–structured model of learning foreign languages

V. Schwämmle

*Instituto de Física, Universidade Federal Fluminense, Av. Litorânea, s/n - Boa Viagem,*
*24210-340, Niterói, RJ, Brasil.*
*Institute for Computer Physics, University of Stuttgart, Pfaffenwaldring 27, D-70569 Stuttgart,*
*Germany.*
*veit@if.uff.br*

The understanding of language competition helps us to predict extinction and survival of languages spoken by minorities. A simple agent–based model of a sexual population, based on the Penna model, is built in order to find out under which circumstances one language dominates other ones. This model considers that only young people learn foreign languages. The simulations show a first order phase transition where the ratio between the number of speakers of different languages is the order parameter and the mutation rate is the control one.

*Keywords*: language; ageing; numerical model; phase transition

PACS Nos.:89.75.-k,89.65.-s

## 1. Introduction

Recently, increasing attention has been paid on the understanding of linguistic systems by computational and analytical methods. Physicists, mathematicians, computer scientists and biologists apply their tools on the investigation of language capability, language change and language competition[1]. Especially the similarity to biological systems opens the field of linguistics to methods of the far better understood evolutionary systems. For an overview see refs.[1,2,3].

It is believed that a concept of "Universal Grammar" is fixed by some way in our genetic code enabling humans to learn languages fast during our childhood[4]. Much attention has been paid on the evolution of this trait and the competition between different grammars[5,6]. In these models the different grammars compete by giving higher fitness to better competitors. Because of that it is possible to apply directly models already successfully used in evolutionary biology.

The competition between languages has been investigated in the last years analyzing the stability of an system of two or more languages[3,7,8,9,10,11] as well as their size distribution[12,13]. In our approach we use a computational model *without*

2   *Veit Schwämmle*

giving different fitness to agents with different language traits and without neglecting bilinguals. Models of that kind can be found in refs.[3,10,11,12].

This makes our model different from biological systems. The knowledge of a different language does not imply a higher death probability (we are aware that, unfortunately, there are quite many exceptions of that rule). In order to present an age–structured model we build our one on the Penna model[14,15]. By the treatment of the genes like bits the model yields a rather realistic age–structure of a sexually reproducing population, and is additionally suitable for large population sizes due to its low computational cost. The usage of an age–structured model enables us to merge an age–dependent learning procedure with a model of language competition in order get more insight into their relation. We will focus on the question whether a phase transition exists, as found in the models of refs.[3,6], between the state of dominance of one language and an uniform distribution (also called fragmentation).

The paper is organized as follows: the next section explains the model, the following section presents the results of the model with respect to its parameter space. The last section discusses the results and proposes further extensions of the model.

## 2. The model

The model in this paper combines population genetics with age–dependent language learning. The presentation of the genetical part, with the purpose to yield a stable population having its characteristic age–structure, will be provided only shortly to the reader. For a more detailed description of the Penna model and its applications we refer to refs.[14,15].

**The sexual Penna model** is a individual–based model where every agent is represented by its age and two strings (diploid) of 32 bits. The bit–strings are read in parallel. Every iteration after birth a new position on each of the bit–strings becomes visible. A bit set to one corresponds to an harmful allele. At the beginning of a simulation five positions are randomly chosen to be dominant ones. If the position is dominant one bit of the two bit–strings suffices to make the position represent a deleterious mutation. On the other positions both bits need to be set to one for deleterious effects. At the age at which an agent has its third deleterious mutation it dies. In order to avoid population explosion an additional factor limits its growth: the Verhulst factor $1 - N/K$ gives the probability of an agent to survive an iteration. Here $N$ is the current population size and $K$ the so–called carrying capacity defining the maximum size of a population. After reaching the reproduction age $R = 10$ a female agent every iteration chooses randomly a male with age equal or older than $R$ in order to produce one offspring. For each parent its two bit–strings are randomly crossed and recombined. Both parents contribute one string to the offspring. The offspring suffers one deleterious mutation at a randomly chosen position on each of its both new strings. This high mutation rate lets the population

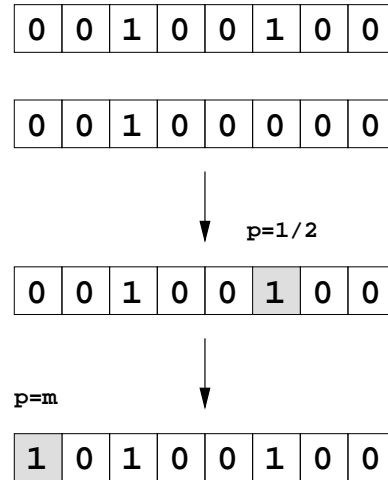reach a stationary distribution fast. The two bit–strings are initialized randomly.

| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|

| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|

p=1/2

| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|

p=m

| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|

Fig. 1.   The language string of the child is built by the two parent's language strings. If both parents speak a language, the child speaks it as well. If only one of the two parents speaks one, the child speaks it with 50% probability. After composition of the child's bit–string mutations with probability $m$ generate a new language spoken by the child on a randomly chosen position.

**The language trait** of an agent, that means its current ability to speak certain languages, is modelled in a similar way as the genetic strings of biological ageing.

The whole structure, a language should consist of, is neglected, and language is treated as an unit. Each agent contributes one additional bit–string of $L$ bits representing its capability to speak a maximum of $L$ languages. So an agent who has for instance its third bit set, speaks currently language $l = 3$ (eventually among others). An agent can learn or forget a language only in youth during the first $c$ iterations after birth. We call $c$ the *maximum learning age*. Older agents are not able to change their knowledge on languages. The interaction between a young agent and a randomly chosen teaching agent works as follows: At first a random position on the language bit–string is chosen. The agent learns the language of that position if it is spoken by the teacher. If the agent already speaks more than one language it forgets the language of the chosen position if the teacher does not know it. In the other cases no process of learning/forgetting is carried out. Let us illustrate the outcome of the interaction in computational terms: the language trait of an agent speaking zero or one language has an OR operation with the teacher's one at the chosen position, otherwise an AND operation. The young agents's language traits are actualized in this way by $f$ different teachers at different positions per iteration. $f = 0.5$ means that the language traits are actualized once per iteration with a probability of 50%. At birth an offspring obtains its language trait as illustrated

4   *Veit Schwämmle*

in Figure 1. If both parents speak the same language the offspring will speak it as well. In the case that only one of the parents speaks a language the child learns it with a probability of 50%. Additionally, the child can know a new language at birth with probability $m$ (i.e. a randomly chosen bit position is set to one). We call this probability $m$ of erroneous learning the *mutation rate* in the following because of its analogy to biological systems. The bit–string is initialized by chance at the beginning of a simulation.

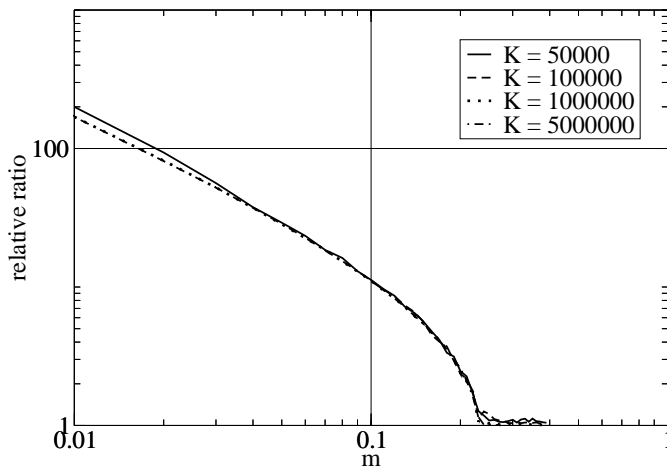## 3. Results

### 3.1. *Two languages*



Fig. 2.   Ratio of the two language sizes versus mutation rate. The size is the number of agents speaking that language. The figure shows a clear phase transition. Different population sizes alter neither the shape nor the position of the critical point.

This section will concentrate at first on the results of simulations with a trait of two languages ($L = 2$). The simulations are carried out for at least $20,000$ iterations or more for large population sizes. All results present the final stationary state of a simulation.

A phase transition can be observed for the ratio between the number of speakers of different languages. The ratio is averaged over the last $1,000$ time steps. Figure 2 shows the ratio versus the mutation rate, determining the control parameter, for different carrying capacities $K$. The simulations are carried out with a maximum learning age of $c = 4$ and $f = 0.5$. Each point represents the outcome of a separate
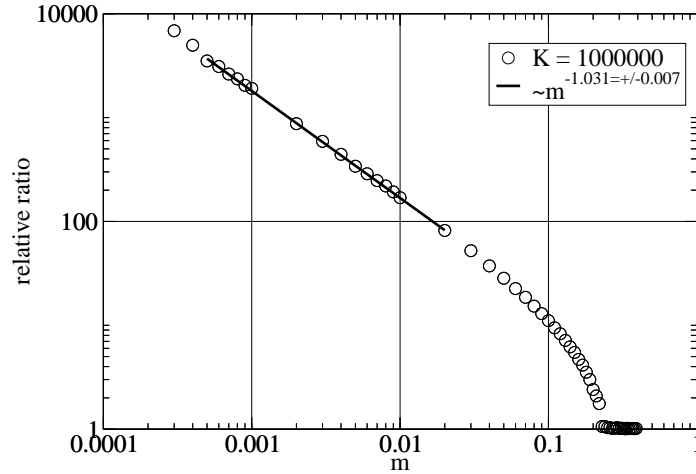
Fig. 3.   The language ratio shows a power law with an exponent of about minus one for small mutation rates.

simulation. The language ratio decreases fast at the critical point which is situated between $m = 0.2$ and $m = 0.3$. The transition separates the phase/state of about equal sizes for the two languages for small values of $m$ from the phase/state where one language dominates clearly the other one. We can observe in Figure 2 as well that the population size does not alter the shape of the transition. Thus this transition can be found for arbitrary population numbers and therefore corresponds to a real physical phase transition.

The language ratio increases strongly for low mutation rates. A power law with an exponent of minus one is exhibited in Figure 3. This exponent is the same for all simulations presented here. The parameters in the simulations are $c = 4$, $f = 0.5$ and $K = 1000000$, i.e. the population consists of about $80,000$ agents.

In Figure 4 the shape of the phase transition is compared for different values of $c$ and $f$, parameters defining the amount of interaction between the agents. The carrying capacity $K = 100,000$ fixes the population number at around $8,000$. The shape of the transition remains the same but is shifted to higher values of the mutation rate for increasing $c$ and $f$. The shift is a linear one with respect to each of both parameters. The shift decreases for values of $c > 10$ as expected due to the smaller number of individuals surviving up to ages much older than the reproduction age $R = 10$ (not shown). The phase transition disappears and the only final state of the model is the dominant one for very large values of $f$.
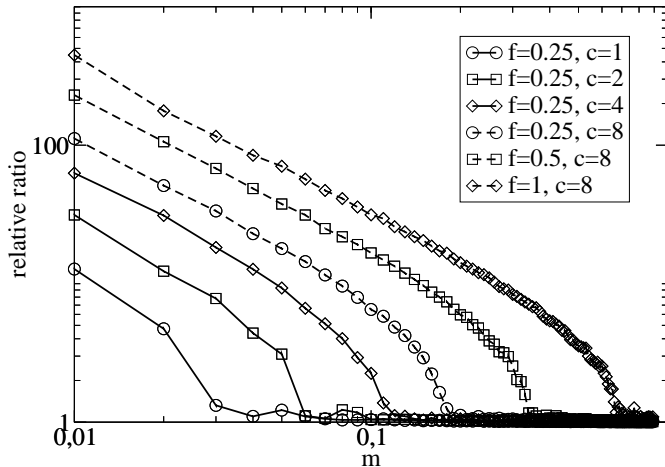
6    *Veit Schwämmle*



Fig. 4.    The curve of the phase transition is shifted to large mutation values for increasing learning time during childhood. The shift increases linearly with an increasing maximum learning age $c$ as well as with increasing amount $f$ of interaction per iteration.

### 3.2.  *Many languages*

In the following the results of simulations with a language trait consisting of more than two languages are compared to the previous ones. The language ratio $R$, the measure for a possible domination of one over the others, is now defined by

$$R = \frac{N(l_{max}) \cdot (L - 1)}{\sum\limits_{l=1, l \neq l_{max}}^{L} N(l)}, \tag{1}$$

where $l_{max}$ is the most spoken language, $L$ the number of languages, and $N(l)$ the number of agents speaking language $l$. The simulations are carried out with $c = 5$ and $K = 100,000$. We used $f = 0.5, 1, 2, 4$ for $L = 2, 4, 8, 16$, respectively in order to keep the interaction per language constant. Figure 5a compares the shape of the transition for different numbers of languages. The position of the critical point is shifted to smaller values of $m$ for more languages. This means that the more languages compete against each other, the lower is the possibility to have a scenario where one language dominates the other ones. Figure 5b compares the same simulations carried out with the initial state where all agents speak language $l = 1$. The critical point is shifted to larger values of the mutation rate. This dependence of the final state on the initialization, leading to a hysteresis, shows that the phase transition is of first order. Here the jump from one phase into the other is located at the intersection of the transition curve and the curve for the ratio of two languages.

*Phase transition in a sexual age–structured model of learning foreign languages*   7
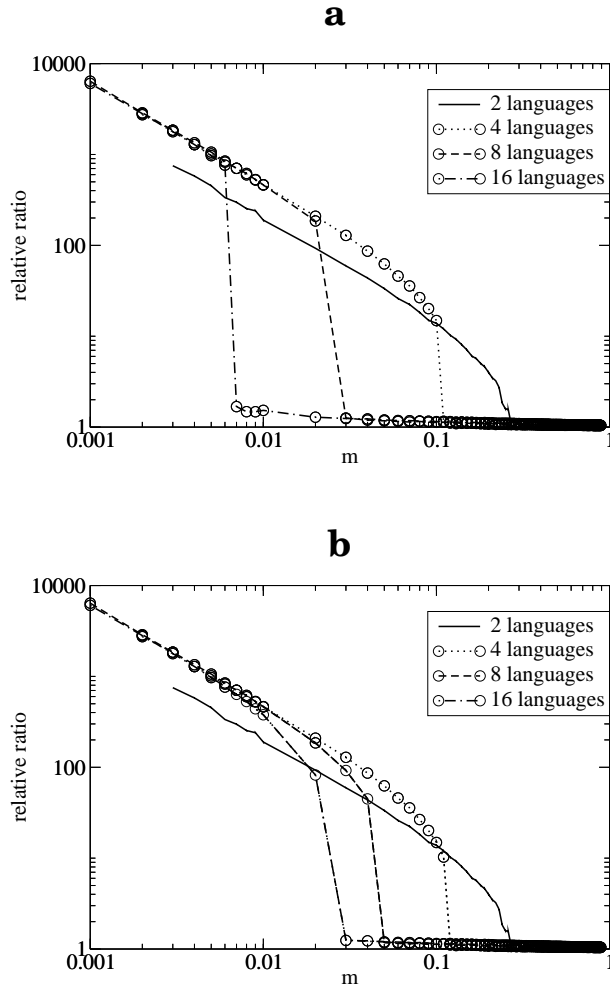
**a**



**b**



Fig. 5.   Ratio of languages versus mutation rate for different numbers of $L$ languages. The critical point moves to smaller values of the mutation rate for increasing $L$. The phase transition is more abrupt. **a**: For numbers of languages larger than two the curves collapse into one for values of $m$ in the left part where one language dominates the other ones. **b**: The initialization with one language leads to a different result. The transition shows a hysteresis for the simulations with more than two languages increasing with $L$.

Finally, we present a histogram of the Hamming distances[16] between all agents for simulations with 16 languages for different values of the mutation rate (Figure 6). The parameters are the same as before. The Hamming distance is defined by the number of bits by which the two bit–strings differ from each other. Or, in other words, the number of bits which need to be changed to turn the language trait of

an agent into that of the other one. Figure 6 shows that the peak of the histogram of Hamming distances turns from zero (dominance) to one (no dominance) at the critical point (which here is located between $m = 0.6$ and $m = 0.7$).
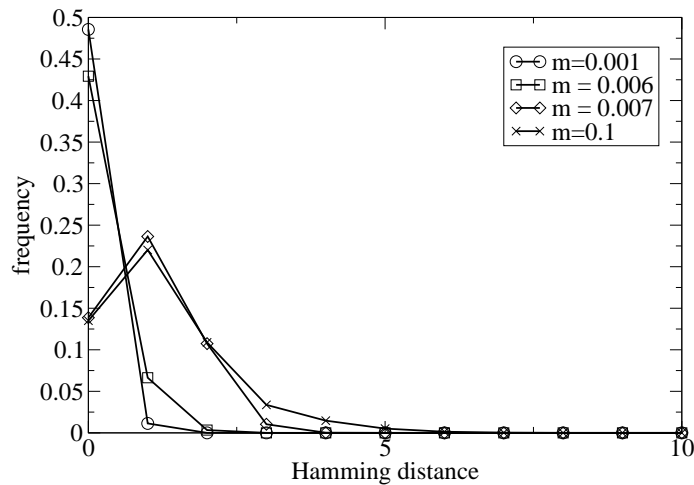


Fig. 6.   Histogram of the Hamming distances between the agents in a simulation of 16 languages for different values of the mutation rate.

## 4. Discussion and conclusions

In our model of language competition a phase transition similar to the one for the competition of different grammars in ref.[6] and the one for the competition of languages of ref.[3] is observed. Nevertheless, working without fitness improvement for agents better adapted by their languages trait but with a sexual population with random mating makes the simulations useful to characterize current language competition. The phase of an uniform distribution of the languages (fragmentation) suggests a scenario where these languages compete equally against each other maintaining a stable state of coexistence. In the other state one language dominates clearly the other ones. The dependence of the phase transition point on the amount of language interaction as well as on the number of competing languages lets us make the following conclusions: There is no optimal value of maximum learning age for a language. Its increase shifts the critical point to larger values of the mutation rate. Thus a larger maximum learning age prevents the evolution of many languages where contact among speakers of different languages is strong. The model shows also that a larger number of different languages decreases the possibility of dominance

of one of them and makes the transition jump higher. Hence, the extinction of a language in a multilingual system can lead to a transition to the state where one language will begin to dominate all the other ones.

The phase transition analyzed in the presented model is of first order for simulations with many languages as found in the model of ref.[3,6]. The differences are that in our case, the transition is completely independent of the population size. We do not know if the simplification of a language to a single bit leads to this discrepancy or if it is the usage of a model with sexual reproduction. Further analysis has to be done.

The model brought insight into the mechanisms of language competition but still much has to be done. The results are similar to the ones of ref.[3]. Nevertheless, our model presents another approach to understand language competition. The design of an analytical model, for instance by diffusion equations similar to the ones used in ref.[8], without fitness would reveal if it is possible to obtain similar results with a mean field theory. The model presented in this paper can be extended by putting it on a lattice in order to investigate how the transition behaves with respect to a geographic distribution of languages. Another extension would be the substitution of random mating by assortative mating[9], a concept used frequently in the theory of biological speciation, in order to give different priorities to monolingual parents and bilingual ones. In order to understand language invasion, the model can be extended to include social structures.

## Acknowledgements

## References

1. W. S-Y. Wang and J. W. Minett. *Trends in Ecology and Evolution*, 20:263, 2005.
2. "Evolution of Language". *special section in Science*, 303:1315, 2004.
3. D. Stauffer and C. Schulze. *Physics of Life Reviews*, 2:89, 2005.
4. N. Chomsky. *Rules and Representations*. Columbia University Press, New York, 1980.
5. M. A. Nowak, N. L. Komarova, and P. Niyogi. *Science*, 291:114, 2001.
6. N. L. Komarova. *Journal of Theoretical Biology*, 230:227, 2004.
7. D. M. Abrams and S. H. Strogatz. *Nature*, 424:900, 2003.
8. M. Patriarca and T. Leppänen. *Physica A: Statistical Mechanics and its Applications*, 338:296, 2004.
9. K. Kosmidis, J. M Halley, and P. Argyrakis. *Physica A*, 353:595, 2005.
10. J. Mira and A. Paredes. *Europhysics Letters*, 69:1031, 2005.
11. V. Schwämmle. *Int. J. Mod. Phys. C*, 16:issue 10, 2005. physics/0503238.
12. C. Schulze and D. Stauffer. *Int. J. Mod. Phys. C*, 16:781, 2005.
13. V. de Oliveira, M. A. F. Gomes, and I. R. Tsang. *Physica A*, page in press, 2005. physcics/0505197.
14. T. J. P. Penna. *J. Stat. Phys.*, 78:1629, 1995.
15. D. Stauffer, S. Moss de Oliveira, P. M. C. de Oliveira, and J. S. Sá Martins. *Biology, Sociology, Geology by Computational Physicists*. Elsevier, 2006. in preparation.

10    *Veit Schwämmle*

16.  T. Tessileanu and H. Meyer-Ortmanns. *Int. J. Mod. Phys. C*, 17:issue 3, 2006.