Schelling-Voter Model: An Application to Language Competition

Inés Caridi[a], Francisco Nemiña[b], Juan P. Pinasco[c,*], Pablo Schiaffino[d]

[a]*Instituto del Cálculo and CONICET, Facultad de Ciencias Exactas y Naturales,*
Universidad de Buenos Aires, Ciudad Universitaria, Pab. II
Int. Guiraldes 2160 (1428) Buenos Aires, Argentina
[b]*Depto. de Física, Facultad de Ciencias Exactas y Naturales,*
Universidad de Buenos Aires, Ciudad Universitaria, Pab. I
Int. Guiraldes 2160 (1428) Buenos Aires, Argentina
[c]*Depto. de Matemática, IMAS-CONICET, Facultad de Ciencias Exactas y Naturales,*
Universidad de Buenos Aires, Ciudad Universitaria, Pab. I
Int. Guiraldes 2160 (1428) Buenos Aires, Argentina
[d]*Facultad de Ciencias Económicas, Universidad de Palermo, and*
Depto. de Historia, Universidad Torcuato Di Tella,
Av. Figueroa Alcorta 7350 (1428) Buenos Aires, Argentina.

Abstract

In this work we analyze the language competition problem by using an interacting agent-based model which interpolates the classical Schelling and Voter models. Briefly, an agent may change its place of residence or his language when he is surrounded by more individuals of the other kind than the ones he can tolerate. We analyze this dynamic process in terms of the free space to move in, the pressure to change the language, and the propensity to change location. We identify the different regimes and the relationship with the language competition problem.

Keywords: Schelling Model; Voter Model; segregation; language competition
PACS: 89.65.Gh, 82.20.Wt

*Corresponding author
 Email addresses:* `ines@df.uba.ar` (Inés Caridi), `nemina@gmail.com` (Francisco Nemiña), `jpinasco@dm.uba.ar` (Juan P. Pinasco), `ines@df.uba.ar` (Pablo Schiaffino)
 URL: `http://mate.dm.uba.ar/∼jpinasco/` (Juan P. Pinasco),
`https://sites.google.com/site/pablolschiaffino/` (Pablo Schiaffino)

1. Introduction

A language evolves with its speakers, and then it is well adapted to natural environment, history, traditions and customs. Usually, there is no translation between two languages for key concepts in a given culture, and the loss of those languages will have several negative consequences. Several examples can be gathered from all around the world: among the Amuesha tribe in Peruvian Amazon, the diversity of crops was reduced due to the loss of knowledge associated to the language endangerment (see UNESCO [1]), and there exist almost 3000 endangered languages in the world.

The study of language dynamics was started by Nettle in [2], who considered several agents interacting in a given network, and changing their languages according to certain rules. These kinds of models are related to the voter model introduced by Holley and Liggett [3], and extensively studied by researchers from complex systems and statistical mechanics. We refer the interested reader to [4, 5] for more references and details.

In those agent-based models, extinction of one of the competing languages is predicted. A different approach based on ordinary or partial differential equations started with the work of Abrams and Strogatz [6]. Their mean field model of language competition predicts the extinction of one of the competing languages. The possibility of coexistence of two languages in the same region was first pointed out by Pinasco and Romanelli in [7], including resources for each population as in Lotka-Volterra models. Previously, Patriarca and Leppanen [8] showed that two languages, each one with its own region of influence, can survive with a tiny zone of coexistence in the border between them. Later, Kandler and Steele [9] showed that extinction of one of the languages is the only stable outcome of Lotka-Volterra models, even when there is spatial dependence under stronger assumptions on both populations: mainly, one population can't extract resources from the other one, a very restrictive condition that excludes slavery and exploitation, together with any kind of symbiosis beneficial to both groups. Also, Patriarca and Heinsalu [10] showed that coexistence is possible due to natural barriers (e.g. islands, mountains).

On the other hand, there are several phenomena related to language dynamics that those models fail to reflect. One of them is the existence of ghettos, and the formation of clusters of people of similar origins in big cities (like Little Italy in New York, or Chinatown in different cities around the world). Although the segregation in ghettos was imposed in the past by

2

some central authority, today they emerge in Europe and America as a result of several individual decisions. Also, the existence of large areas with a huge number of languages, mainly West and Central Africa, Central America, South-East Asia, and the Pacific Island [1, 11, 12], was not completely explained: the existence of natural barriers explains the case of Mexico or the Pacific Island (see [10]); however, the African diversity of languages despite the absence of barriers remains to be explained.

In this work we propose a different mechanism of language competition, a mix between the Voter model, and the Schelling segregation model [13, 14]. In his model, Schelling showed how individual residential preferences impacted on segregated neighborhood patterns, and how even modest levels of racial preferences can be amplified into high levels of global segregation. In his original model, the agents are black or white, and they are located in a square lattice with a proportion of empty sites. Each agent has a fixed tolerance, which is a bound of the rate of people of the opposite color that agent tolerates among his neighbors, and when this bound is reached, that agent decides to migrate to some empty site.

What happens if the agent can change either his color or his position? As far as we know, that option was not previously considered, since this does not make sense when we speak about racial characteristics. However, this is a clearly valid option in many social and economic situations, where a frustrated agent can choose between trying to adapt to his neighbors or migrate to a better place. A mix between Axelrod and Schelling models was proposed in [15]. In this work, two neighbor agents are selected at random and if they do not share any characteristic, one of them changes its position; this model is reduced to the Schelling model in our case. A network evolution model in [16] has a link rewiring or opinion change as a consequence of the interaction of two connected agents with different opinions, see also [17]. Let us stress that in both models, no tolerance parameters are introduced, as in Schelling-type models.

Here we are considering a time interval in which languages remain the same or change slowly, but which comprises several generations of individuals. Thus an agent which changes its language two or more times represents different individuals. Also, we can think of each agent as a small group of people (a block, a family), and we define this agents language as the predominant one among the members of the group.

In a separate work [18], we analyze several aspects of the Schelling-Voter

model. We introduce a slight modification in the spirit of [19, 20] (i.e., enabling also happy agents to change color or location with very low probability), and we study their properties with statistical mechanic techniques.

The paper is organized as follows: in Section §2 we describe the model. The results of simulations are presented in Section §3, and in Section §4 we analyze language patterns in term of different equilibria obtained in the simulations.

2. The Schelling-Voter Model

In this model, we consider a population of N agents, each one located in a different site of a square lattice of side L, and labeled as b or w (not necessarily black or white, but the language he speaks, or the social variable or belief he supports). We introduce the following global parameters:

- $1 - \rho \in (0, 1)$, the *density* of empty sites, defined as $\rho = \frac{N}{L^2}$.

- $T \in [0, 1]$, the *tolerance*, equal for all agents.

- $s \in (0, 1)$, the *prestige* of language b; the prestige of w is $1 - s$.

- $p \in [0, 1]$ the *possibility to migrate*, the same for all agents.

For a given notion of neighborhood (we have considered the Moore neighborhood of each agent, which comprises the eight sites surrounding his location) we say that an agent is *unhappy* when the ratio between the number of agents of the opposite language and the number of agents in his neighborhood is greater than T. Now, an unhappy agent can play one of two different games:

- (Schelling) The agent tries to change his location, choosing a new one at random among the empty sites in the lattice, where the agent is bound to be happy.

- (Voter) The agent tries to change his color/language.

The unhappy agent will play Schelling dynamics with probability p, and he will play Voter dynamics with probability $1 - p$. We can interpret probability p as a balance between the ease and the difficulty to migrate and to learn a new language.

In both cases the change is not mandatory, the migration can fail if the agent is unhappy in any empty space. The language change would depend on some probabilities $q_{b \to w}$, $q_{w \to b}$ which can include both the number of agents of each language (locally or globally measured), and the perceived prestige of the other language.

In this work, we focus on the results of the size of the populations speaking each language when the system stabilizes. We leave aside several measures related to statistical mechanics, see [18].

3. Simulations

We start with two similar populations in a square lattice of side $L = 50$, with free boundary conditions. We assign the same prestige $s = 0.5$ to both languages. For different values of T between 0.2 and 0.6 we obtain qualitative similar results, and we show here only the simulations for $T = 0.3$. We run simulations for p and $1 - \rho$ in the range between 0.02 and 0.98 in FORTRAN 90. At the beginning, we filled each site of the lattice with an agent with probability ρ, and labeled each one as b or w with the same probability.

In the model, we take an agent at random, then we check his tolerance. If the agent is unhappy, then he will try to migrate (with probability p) or he will try to learn the other language (with probability $1 - p$).

When the agent tries to find a new place of residence, we choose one at random among the available empty sites where the agent is bound to be happy; if the agent tries to change the language, we use the transition probabilities as in [6]

$$q_{b \to w} = \frac{W}{2(B + W)} \qquad q_{w \to b} = \frac{B}{2(B + W)},$$

where B (respectively, W) is the number of agents speaking language b in the population (resp., w).

Figure 1 shows the dependence of the size of the minority language population (S) on the empty space density $1 - \rho$ for different values of p from 0.1 to 0.9. We observe that the size of the population increases quickly with ρ, and then reaches an almost fixed value close to one-half of the total population. The behavior of the different curves is similar, the differences being due to finite size effects.

Figure 2 shows the size of the smaller population (S) as a function of p for different values of $1 - \rho$ from 0.1 to 0.9. In the case $1 - \rho \approx 0.1$ both
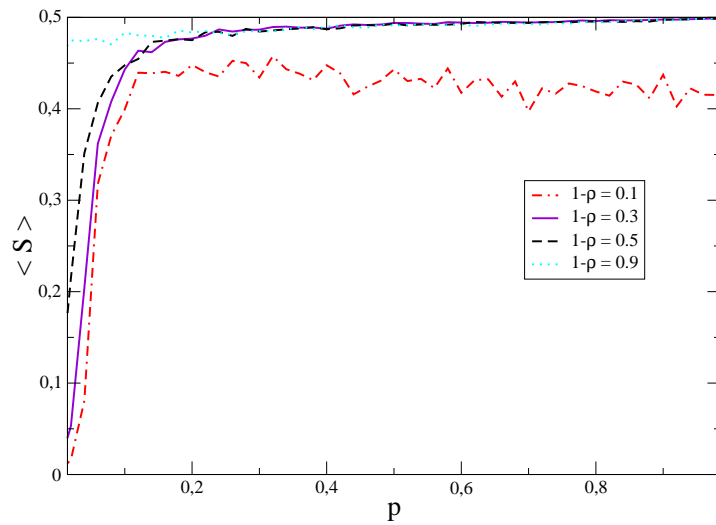
Figure 1: Size of the population speaking the minority language (S) as a function of p for different values of $1 - \rho$ and $T = 0.3$. In simulations, $N = 2500$ agents asynchronously evolve. Results are obtained by averaging 30 realizations of 50000 time steps each, or until the system
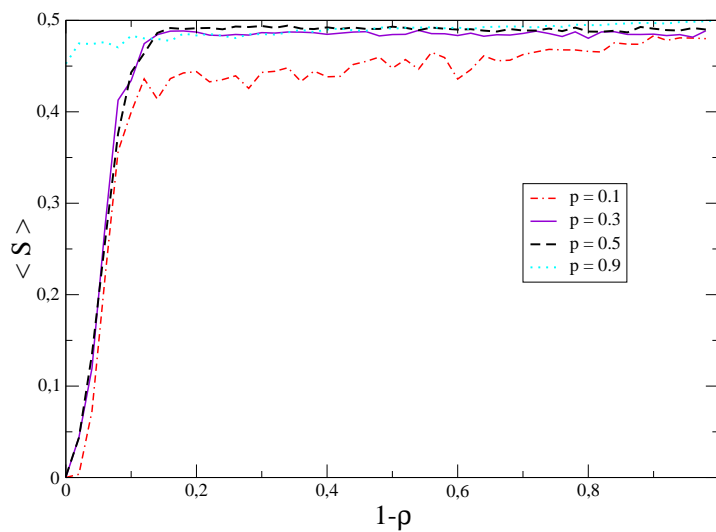

Figure 2: Size of the population speaking the minority language (S) as a function of $1 - \rho$ for different values of p and $T = 0.3$. In simulations, $N = 2500$ agents asynchronously evolve. Results are obtained by averaging 30 realizations of 50000 time steps each, or until the systems stabilizes in the sense that 99 per cent of the population is happy.
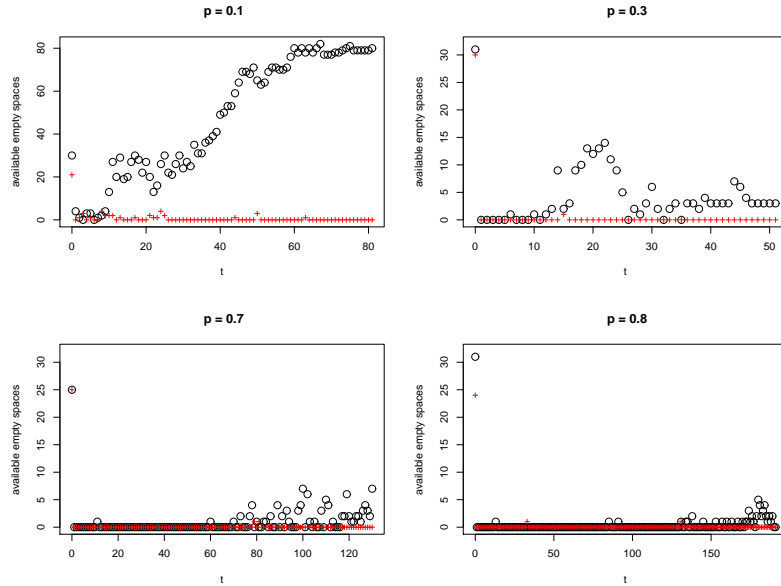
Figure 3: Amount of available empty spaces for one of the population (circles) and for the other one (plus symbols) for some values of $p = 0.1, 0.5, 0.7$ and 0.8 as a function of time. In each case, one realization is performed for the case of $1 - \rho = 0.06$ and $N = 2500$ agents. Circle symbols represent those population who obtain advantage in this particular realization.

populations results with no available empty spaces to move in, at the second time step of the dynamics, and after some rounds in which only language changes occur, empty spaces become available for one of them (see Figure 3).

Finally, Figure 4 shows the phase diagram in the variables $(1-\rho)$-p. There are three clearly identified regions:

- In region 1 we find extinction of one of the languages. For p small, they are essentially playing the voter model. When $1 - \rho$ is too small, since unhappy agents can not find comfortable places and there is no room for an interphase separating agents, they sooner or later change their language.

- In region 2 we observe another phenomenon: ghetto formation.

- In region 3 both populations coexist with similar sizes. However, we have two type of final states, the ones where both languages have a clear
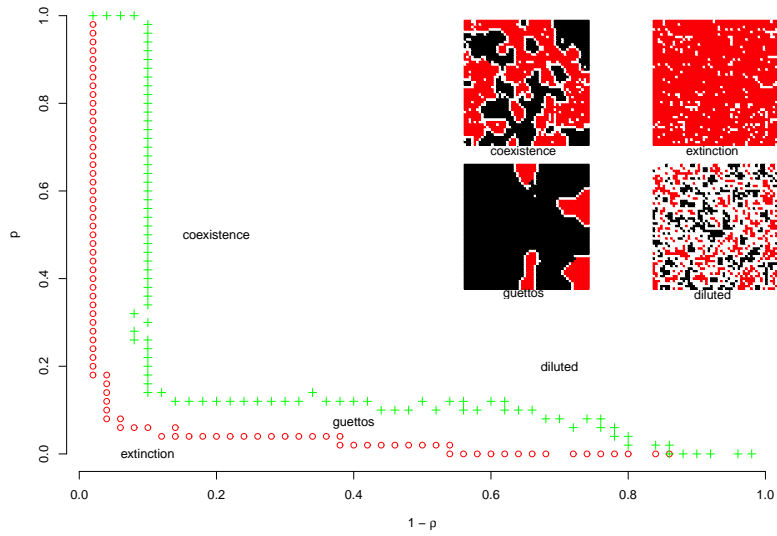
7

Figure 4: Below circle symbols there are those cases which result in the *extinction* of one population, with values of $\langle S \rangle < 0.1$. Between circle and plus symbols, there are those which result in the formation of *guettos*, with values of $0.1 < \langle S \rangle < 0.45$. Beyond plus symbols, there are those cases which result in the *coexistence* of the two populations, with values of $\langle S \rangle > 0.45$. Cases of $\rho > 0.7$ are *diluted* ones. Values of $\langle S \rangle$ were obtained by averaging S for 30 realizations of 5000 time steps each (or until the systems stabilizes) $N = 2500$ agents and $T = 0.3$ by varying p and $1 - \rho$ by steps of 0.02. The inset of the Figure shows some examples of final configurations for each region.

8

segregation, and another one where both populations remain almost like in the initial distribution, since the low population density gives few opportunities to change language or location (diluted case).

4. An Application to Language Competition

Let us focus now on the problem of language competition. We can interpret the tolerance parameter T as a mix between the acceptance of speakers of the other language, and the critical or maximum acceptable level of difficulty to interact with the neighbors of the other language group (getting a job, education for the kids, medical care, buying basic goods/services, etc.). Beyond T, the agent must decide between changing location or adopting the other language. Here, we can think that p is high when the difficulty to learn the other language is high or when it is easy to move to other sites, and p is small when it is difficult to move from one site to another one, or both languages are similar in difficulty.

Let us analyze Figure 4 and the typical configurations of the lattice in each region of the phase space when the system stabilizes.

In region 1 we can see that only one language remains. This is the case of extinct and endangered languages. A low-prestige language is faced with extinction, since it cannot survive integrated to a higher-prestige population which prefers to ignore it. There are several examples of this situation, see [1, 11], the book of Harrison [21], and the one of Nettle and Romaine [22].

Above this zone, in region 2, there are coexistence of the two populations and ghetto formation. Here, both languages survive, and they are arranged in disjoint zones. We can observe this kind of segregation in Asian and Latino communities in maps of Chicago, see [23]. Similar maps were constructed for other American cities, see [24].

From US Census Data [25], almost 30 percent of the population older than 5 years who speak Spanish, Chinese, or Korean languages at home, speak English "not well" or "not at all". So, we can safely conclude that the segregation by ethnicity observed in the maps is also segregation by language.

In region 3, there are two different phenomena. At the center of the figure we have coexistence of languages, with two segregated populations. In some sense, this is the usual distribution of nations which grow and finally reach a steady state, each one with its own language. Also, Sub-Saharan Western

9

Africa is an example of several coexisting languages with high density of people. In particular, Nigeria has more than 500 languages belonging to at least three main families (Nilo-Saharan, Afro-Asiatic, and Niger-Congo), with around 169 million people, and 950.000 km^2; the density of inhabitants is higher than the one of the European Union. The language maps show a mixed pattern of coexistence (at west, map 1 in [11]) and segregation (at east, maps 4 and 5 in [11]).

With half Nigeria's density, Benin, Ghana and Togo have several languages. The two main families (Gur, Kwa) in Ghana are clearly segregated, with a natural border around the bifurcation of the Black and White Volta Rivers. Nevertheless, at least 30 different languages coexist in each zone (see map from [11]).

Let us note that in this zone of Africa there exists a high degree of mobility of people, and small groups of people spread or merge, forming new groups, changing their languages and cultures, as D'Azevedo has described in [26]:

> Cultural pluralism, multilingualism, and multiple local traditions of origin and "ethnicity" obtain within situations that are only superficially and frequently only temporarily characterized by a predominant "tribal" orientation. (...) Throughout northern and western Liberia institutional structures and most cultural features are so generally distributed that it is no exaggeration to suggest that tribal identification is as much a matter of individual choice as of the ascribed status of birth, language, or distinctive customs.

Finally, let us consider the right hand side of the Figure 4, which correspond to zones with a low-density of population. We mention two possible examples of this situation. The Sahara Desert covers over 80% of Niger, and the country has a very low population density (around 12 persons per km^2). Around the Ténére Desert (*Ténére* means *desert* in Tuareg language), several different languages coexist, belonging to three different families (see map from [11]).

A similar picture can be found in the North part of the Kalahari's Desert basin, where several languages belonging to the Bantu and Khoisan families coexist, in Namibia and Bostwana. Namibia has the second lowest population density in the world, with 2.5 hab/km^2, and Boswana has 2.7 hab/km^2.

In both cases, the overlapping of languages of different families in the same zone, and the low density of populations suggest a coexistence in a diluted state, with few interactions.

5. Conclusions

We have presented an agent-based model which dynamics is a mix between Schelling's segregation model and the Voter model. A probability parameter p governs which dynamics is used each time an agent feels uncomfortable within its neighborhood. The agent feels uncomfortable when the proportion of neighbors of the other type exceeds a given tolerance parameter T. The other relevant parameter in the model is the density of empty locations, a low density constrains the possibility of an agent to find an acceptable site to move.

When the typical Voter model is played, an agent changes automatically its color by selecting another agent at random. Here, we introduce an additional parameter, the prestige or status of each color, following the model of Abraham and Strogatz of language competition. However, since we have set this parameter equal to one half, this represents only a change on the temporal scale, and it is twice more probable to change location than language.

We have started with similar populations, spread out at random on a square lattice, and show that the system stabilizes in few options: extinction of one of the colors/languages, ghetto formation, segregated coexistence, or diluted coexistence. Finally, we have shown several examples of those possibilities in the actual distribution of languages in the world.

[1] UNESCO Ad Hoc Expert Group on Endangered Languages, *Language Vitality and Endangerment*. UNESCO, United Nations Educational, Scientific and Cultural Organitazion (2003). See also Endangered languages http://www.unesco.org/new/en/culture/themes/endangered-languages/

[2] D. Nettle, *Using Social Impact Theory to simulate language change*. Lingua, 108 (1999) 95-117.

[3] R. A. Holley, T. M. Liggett, *Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model* Ann. Probab. Volume 3, Number 4 (1975), 643-663.

[4] C. Castellano, S. Fortunato, V. Loreto, *Statistical physics of social dynamics*, Rev. Mod. Phys. 81 (2009) 591-646.

[5] C. Schulze, D. Stauffer, S. Wichmann, *Birth, Survival and Death of Languages by Monte Carlo Simulation*, Comm. Comp. Phys. 3 (2008) 271-294.

[6] D. M. Abrams, S. H. Strogatz, *Modelling the dynamics of language death*, Nature 424 (2003) 900.

[7] J. P. Pinasco, L. Romanelli, *Coexistence of languages is possible*, Physica A. 361, 1 (2006) 355-360.

[8] M. Patriarca, T. Leppanen *Modelling language competition*, Physica A 338 (2004).

[9] A. Kandler, J. Steele, *Ecological models of language competition*, Journal of Biological Theory 3 (2008) 164-173.

[10] M. Patriarca, E. Heinsalu, *Influence of geography on language competition*, Physica A 388, 174 (2009).

[11] B. Grimes, G. Grimes, and Summer Institute of Linguistics. *Ethnologue*. Dallas, TX, USA: SiL International, 2000. See also Lewis, M. Paul, *Ethnologue: Languages of the world, sixteenth edition*, Dallas, TX, USA: SiL International. Online version: http://www. ethnologue. com (2009).
Ghana: http://www.ethnologue.com/map/GH
Niger: http://www.ethnologue.com/map/NE
Nigeria: see http://www.ethnologue.com/country/NG/maps

[12] D. Nettle, *Explaining global patterns of language diversity*, Journal of anthropological archaeology 17 (1998) 354-374.

[13] T. C. Schelling, *Models of segregation*, American Economic Review 59, 2 (1969), 488-493.

[14] T. C. Schelling, *Dynamic models of segregation*, The Journal of Mathematical Sociology 1, 2 (1971) 143-186.

[15] C. Gracia-Lzaro, L. F. Lafuerza, L. M. Flora, *Residential segregation and cultural dissemination: An Axelrod-Schelling model*, Phys. Rev. E 80, 046123 (2009).

[16] D. H. Zanette, S. Gil, *Opinion spreading and agent segregation on evolving networks*, Physica D: Nonlinear Phenomena 224 (2006) 156-165.

[17] R. Durrett, J. P. Gleeson, A. L. Lloyd, P. J. Mucha, F. Shi, D. Sivakoff, J. E. S. Socolar, C. Varghese, *Graph fission in an evolving voter model*, PNAS 109 (2012) 3682-3687.

[18] I. Caridi, F. Neminã, J. P. Pinasco, N. Saintier, P. Schiaffino, *When Schelling met Jackson, a segregation model with changing type agents*, in preparation.

[19] L. Dall'Asta, C. Castellano and M. Marsili, *Statistical physics of the Schelling model of segregation*, Journal of Statistical Mechanics: Theory and Experiment (2007) 1-9. arXiv:0707.1681

[20] L. Gauvin, J. Vannimenus, J.-P. Nadal *Phase diagram of a Schelling segregation model*, Eur. Phys. J. B, 70 (2009) 293-304.

[21] K. D. Harrison, When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge. Oxford University Press, 2007.

[22] D. Nettle, S. Romaine, Vanishing Voices. The Extinction of the World's Languages. New York: Oxford University Press, 2000.

[23] W. Rankin, *Cartography and the Reality of Boundaries* Perspecta 42 (2010) 42-45.

[24] Eric Fischer, *Race and ethnicity (2000)*, and *Race and ethnicity (2010)*, http://www.flickr.com/photos/walkingsf/sets/72157624812674967/ http://www.flickr.com/photos/walkingsf/sets/72157626354149574/

[25] US Census Bureau (2010). See Table 3A, at http://www.census.gov/hhes/socdemo/language/data/acs/appendix.html

[26] W. L. D'Azevedo, *Tribe and Chiefdom on the Windward Coast*, Rural Africana 15 (1971), 10-29.