COMMUNICATIONS IN COMPUTATIONAL PHYSICS
Vol. **3**, No. 2, pp. 271-294

Commun. Comput. Phys.
February 2008

REVIEW ARTICLE

Birth, Survival and Death of Languages by Monte Carlo Simulation

C. Schulze[1], D. Stauffer[1,*] and S. Wichmann[2]

[1] *Institute for Theoretical Physics, Cologne University, D-50923 Köln, Euroland.*
[2] *Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.*

Received 1 April 2007; Accepted (in revised version) 16 July 2007

Available online 27 September 2007

Abstract. Simulations mostly by physicists of the competition between adult languages since 2003 are reviewed. The Viviane and Schulze models give good and reasonable agreement, respectively, with the empirical histogram of language sizes. Also the numbers of different languages within one language family is modeled reasonably in an intermediate range. Bilingualism is now incorporated into the Schulze model. Also the rate at which the majority shifts from one language to another is found to be nearly independent of the population size, or to depend strongly on it, according to details of the Schulze model. Other simulations, like Nettle-Culicover-Nowak, are reviewed more briefly.

AMS subject classifications: 91.C15

PACS: 89.20.-1, 89.65.-s
Key words: Language competition, Schulze model, Viviane model, agent-based modeling.

Contents

*Corresponding author. *Email addresses:* stauffer@thp.Uni-Koeln.DE (D. Stauffer), wichmann@eva.mpg.de (S. Wichmann)

1 Introduction

While the emergence and learning of human languages has been simulated for decades on computers [1], and while a later economics Nobel laureate also contributed to linguistics long ago [2], the competition between existing languages of adults is a more recent research trend, where physicists have tried to play a major role. The modeling follows the principle of survival of the fittest, as known from Darwinian evolution in biology, and indeed many of the techniques have been borrowed from simulational biology [3]. This emphasis from physics on the competition of existing languages for adult humans started with Abrams and Strogatz [4] and was then followed by at least six groups independently [5–10]. More recently, of course, reviews [3, 11] and conferences brought them together, and others followed them [12–15].

Today about 7000 different languages (as defined by linguists) are spoken, and on average every two weeks or so one of them dies out [16]. On the other hand, the split of Latin into different languages spoken from Portugal to Romania is well documented. In statistical physics, we can describe and explain the pressure which air molecules of a known density and temperature exert on the walls. But we cannot predict where one given molecule will be one second from now. Similarly, the application of statistical physics tools to linguistics may describe the ensemble of the seven thousand or so presently existing languages, but not the extinction of one given language in one given region on Earth. Fig. 1 shows how many languages exist today, as a function of the number of speakers of that language. A statistical theory of language competition thus first of all should try to reproduce such results, in order to validate the model. If it fails to describe this fact, why should one trust it at all? Or as stated by linguist Yang on page 216 of [18]: It is time for the ancient field of linguistics to join the quantitative world of modern science.

This review starts with our own model for numerous languages in Section 2, followed by a review of the alternative model of Viviane de Oliveira and coworkers [10]. Then we review more briefly the many other models which at present do not allow for the simulation of thousands of different languages. Work paying special attention to sociolinguistic modeling, i.e., using respectively Barabási-Albert networks and Social Impact Theory, is reviewed at the end of Subsection 2.1 and the beginning of Section 4, and Section 5 develops a more purely linguistic point of view on the whole simulation enterprise.

2 Schulze model

2.1 Definition

Our own simulations, also called the Schulze model, characterize each language (or grammar) by F independent features each of which can take one of Q different values; the binary case $Q=2$ allows the storage in bit-strings. Three basic mechanisms connected with probabilities p, q and r are common to all variants:

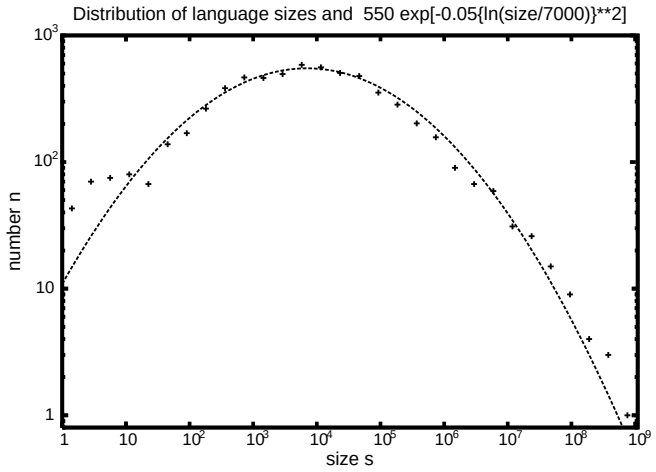

Figure 1: Empirical variation of the number N_s of languages spoken by s people each. For better presentation, the language sizes s are binned in powers of two. Data from Ethnologue [21], as plotted in [22]. The parabola corresponds to a log-normal distribution; we see deviations from it for the smallest sizes [17].

i. With probability p at each iteration and for each feature, this feature is changed (or mutated in biological language). This change is random or not, depending on process ii).

ii. With probability q the mutation/change under i) is not random but instead transfers the value of this feature from another person in the population. This transfer is called diffusion by linguists. With probability $1-q$, the change is random.

iii. With probability $(1-x)^2 r$ (also $(1-x^2)r$ has been used instead) somebody discards the mother language and takes over the whole language (all F features) from another person in the population. Here x is the fraction of people speaking the old language. This flight is called shift by linguists.

Linguistically these three types of modification may correspond to the analog of biological mutations, to the transfer of loanwords from one language into another, and to learning the new language, for instance by immigrants. We *assume* that the F features are modified independently of each other. Since [19] linguists have been aware that many grammatical traits are interrelated in the sense that the presence of one trait typically implies the presence of another across many languages. This type of interdependence could be built into a model, but unless the purpose of the investigation is to specifically look at effects of such interdependence there is no need for this kind of complication. It has recently been shown in [20] that mutually independent features can be identified, and we may allow ourselves the assumption that the features of our model belong to this set. For diffusion and language shift (using the terminology of linguists) one could build in a greater probability of targeting a language which is more similar to one's own than other languages [12], but in reality people are more likely to shift to or be influenced by

the languages that are important for their socio-economic well-being than languages that just happen to be closely related.

Several variants are possible: One can use one joint population where everybody can meet everybody for transfer and shift; or we put people onto a square $L \times L$ lattice or more complicated network, where diffusion and/or shift are possible only from a randomly selected neighbour. People may migrate on this lattice, a phenomenon which physicists would call diffusion. The population can be fixed, meaning that at every iteration all adults are replaced by their children. Or it can grow by a suitable birth and death process; in this case the shifting probability can include also a factor proportional to the population. If one dislikes having three free parameters p,q,r one may set $q = 0$, $r = 1$ without much loss in results.

For the number F of features, from 8 to 64 were used in simulations. Real languages contain many thousands of words for everyday use, and thus one should identify one feature rather with an independent grammatical element (like the order of subject, object and verb in a sentence) than with a word. F for real languages was estimated as about 30 [23] or about 40 to 50 [18] such choices, and the *Word Atlas of Language Structures* [24] lists 138 features with up to $Q = 9$ values. These grammar sizes thus correspond roughly to what has been simulated. According to [25] the average rate of change in normal linguistic typological features, i.e., excluding a few extraordinarily unstable ones, is 16 % per 1000 years.

2.2 Main results

If we start with everybody speaking one language (or with just one Eve), then at low p this language still dominates and is spoken by more than half of the total population, with the remaining people speaking a minor and short-lived variant of this dominating language. At high p, on the other hand, the whole population soon fragments into many languages, roughly such that everybody selects nearly randomly one of the Q^F possible languages. This corresponds to the biblical story of the Tower of Babel. We thus have dominance for small p and fragmentation for large p, with a first-order phase transition or jump at some threshold value which depends on the other parameters and details of the model.

If instead we start with everybody speaking a randomly selected language, then for high p this situation remains. For low p, however, after some time one language by random accidents happens to grow to a sufficiently large size such that it then grows rapidly to be spoken by more than half of the population. Thus a transition from initial fragmentation to final dominance happens in Fig. 2. The threshold value is different from the one for the opposite direction from dominance to fragmentation: we have hysteresis as is common for first-order transitions, Fig. 3. Empirically, this transition to one dominating language was observed on the American continent; for instance, within the last five centuries, two thirds of the native Brazilian languages have died out. And in the last half century we observed the rise of English in physics research publications. While 85

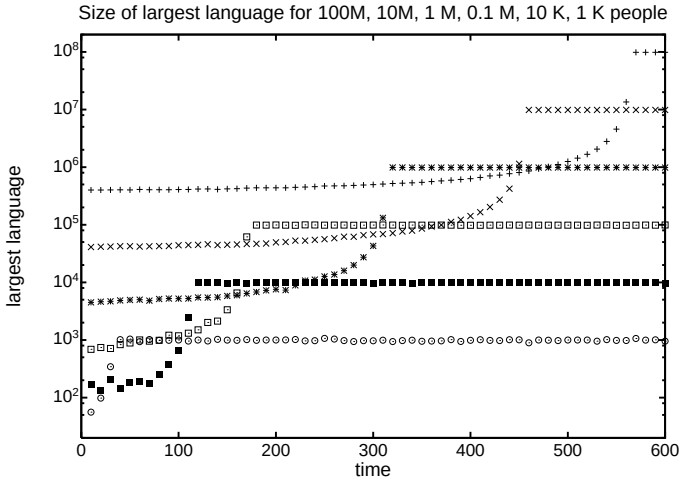

Figure 2: Variation with time of the number of people speaking the most widespread language, for various population sizes. The larger the population is, the longer is the time until the transition from fragmentation to dominance takes place. $Q=2, F=8, p=0.06, q=0.94$; from [11].

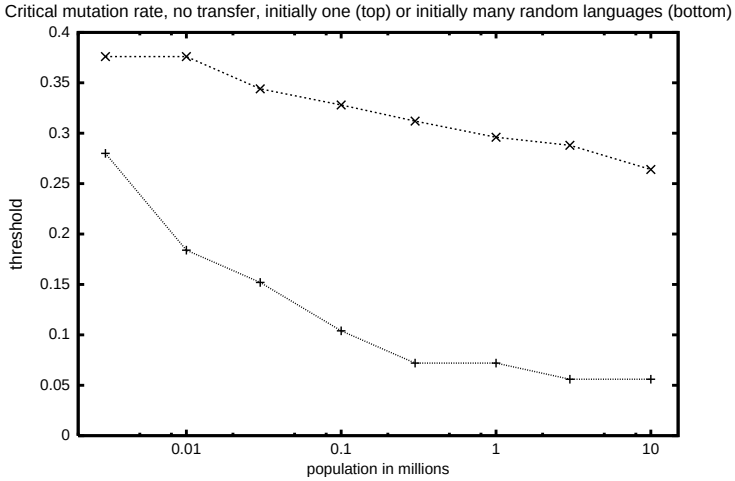

Figure 3: Dependence of the mutation threshold for the phase transition on the population size; upper data from dominance to fragmentation, lower data from fragmentation to dominance. Above the curves we arrive at fragmentation, below at dominance. From [11].

years ago, physicist Bose sent his paper from India to Einstein in order to have it translated from English into German (which led to Bose-Einstein condensation), after World War II physics research was usually published in English, first in Japan, from the 1960's in (West) Germany, a decade later in France, from the 1990's in Russia; finally, China has witnessed a surge in physics papers written in English since 2000.

The time needed to go from fragmentation to dominance increases roughly logarithmically with population size, at least in the binary case $Q=2$ without lattice. Thus a

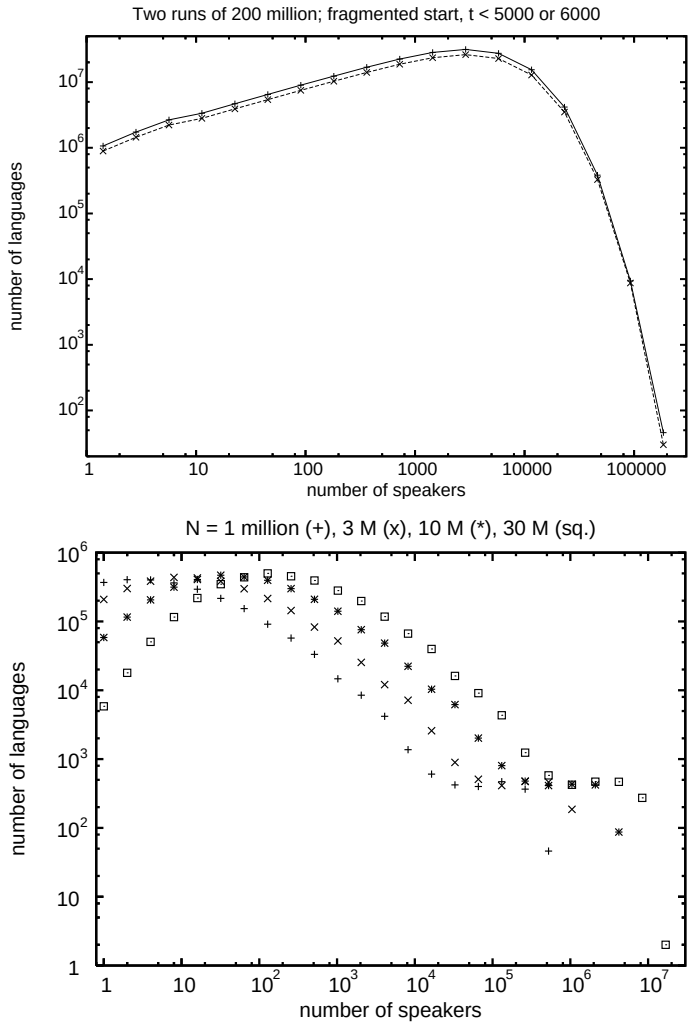

Figure 4: Language size distribution N_s. Top: without lattice, with random multiplicative noise, mutation rate 0.0032 per bit-string, not in equilibrium [22]. Bottom: On scale-free network in equilibrium [32].

mathematical solution for an infinite population might never get this transition. In other words, proper models should be agent-based [26], with independently acting individuals; one should not average over the whole population, using differential equation for the concentrations. Such simulations have been standard in computational physics for half a century (Monte Carlo and Molecular Dynamics), while mean field approximations average over many individuals and can give somewhat or completely wrong results. (The transition from fragmentation to dominance may require a shift probability $(1-x)^2 r$ instead of $(1-x^2)r$.)

The language size distribution to be compared with Fig. 1 is shown in Fig. 4a. To get it, we looked at non-equilibrium results and introduced random multiplicative noise,

since otherwise the language sizes were too small and their distribution too irregular. Fig. 4b avoids these tricks and instead places the people on a directed scale-free network, discussed below.

No lattice or other spatial structure was employed in the above simulations. On a lattice one can look at language geography [27–29]. North and South of the Alps, different languages are spoken, and a similar separation is caused by the English Channel. Genetic and linguistic boundaries in Europe mostly coincide, and about two thirds of them agree with natural boundaries like a mountain chain or sea [30]. We simulated this effect on a lattice [31] with contact only between nearest neighbours and a horizontal barrier separating the upper from the lower half. The shift from a small to a large language happens across the barrier only with a small crossing probability c. For $c=0$ one thus has two completely separated halves of the lattice, and trivially the languages which evolve as dominating are different on both sides of the border. With $c=1$ the border has no effect, and only one language dominates. Fig. 5 shows how often for small but finite c two separate dominating languages may coexist; already quite small c suffices, particularly for large lattices, to unify the two regions into only one with the same language dominating on both sides.

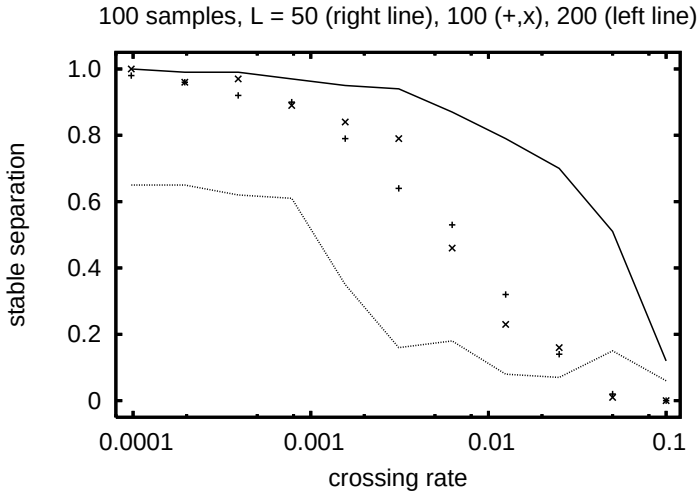

Figure 5: Fraction of cases when a semi-permeable barrier allowed two different languages to dominate on its two sides in the Schulze model.

A variant of this barrier program was used to simulate cases like Alsace in France, west of Germany. Many people west of the border speak German but belong to France and want to belong there. Assuming that 80 percent west of the barrier (= Rhine river) speak German, will contacts with Germany to the east induce them to have political views like Germans instead of French? Since now only two languages are involved we simulate them by one bit only and use the $F=8$ features, represented by numbers from 1 to $Q=3$ or 5, to define the political opinions of the people. With probability 0.2 at each

iteration, each German-speaking Alsatian selects the opinions of a randomly selected German from east of the border; this long-range contact sets in only after two dominating sets of opinion (= national identity) on both sides of the border were established. We find that the population may switch political allegiance, but except for linear lattice dimension $L<20$ this happens only after many thousands of iterations. Sometimes long metastable plateaus are formed in the fraction of people belonging to the two opinion sets.

Going back to Fig. 4b, there we employed a directed Barabási-Albert scale-free network, used by linguists before [33]. These networks are grown from a small fully connected core such that each new network member selects m already existing members as teachers. The more people have selected a certain teacher before, the higher the probability is that this teacher will again be selected. Information only flows from the teacher to the person who selected this teacher, not in the opposite direction [35].

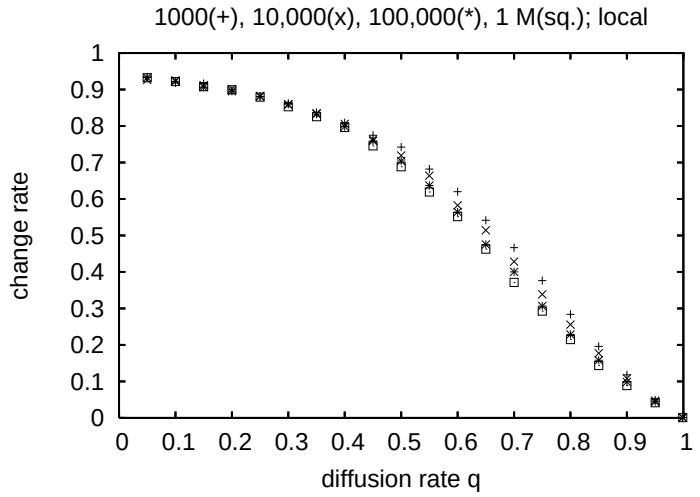

Figure 6: Variation with diffusion q of the rate at which the dominating language is replaced on a scale-free network; $p=0.5, r=0.9$, see [34].

This directed network is also used in Fig. 6 which shows how often the dominating language is changed into another dominating language. There is only a minor change when the population size increases drastically, in agreement with some empirical evidence [34]. Here, diffusion takes place between network neighbours. If instead diffusion is possible between arbitrary network nodes, the rate of changes goes down drastically with increasing population size [34], in agreement with other empirical evidence [36].

When a region is conquered by people speaking another language, we assume that at each iteration each person with probability s shifts to the conquering language. The time needed for the conquering language to become dominating is inversely proportional to s for directed Barabási-Albert networks, but goes to infinity on the square lattice for decreasing s at some critical value s_c [37].

2.3 Bilingualism

Several authors studied the possibility that people speak more than one language [5,6,38], and here we do the same for the Schulze model on the square lattice, with $F=8$ features of $Q=3$ different values, using only interactions to nearest neighbours [31]. For this purpose we modify the shift process.

Before, languages spoken by a fraction x of the four neighbours were dropped in favour of the language of a randomly selected neighbour with probability $(1-x)^2 r$ ($r=0.9$). Now we do this at lattice site i only if none of the four neighbours of i speaks the mother language of site i: $x=0$; then with probability r we replace the mother language of i by the mother language of a randomly selected neighbour. Otherwise, for $x>0$, with the above probability $(1-x)^2 r$, site i learns as an additional "foreign" language a randomly selected (foreign or mother) language of a randomly selected neighbour. If in the latter case $x>0$, site i has already learned a foreign language before, then this old foreign language is replaced by the new foreign language.

These are the learning and replacement events if everybody can speak at most two languages. If instead the number of languages for each person is restricted by an overall upper limit, then for $x>0$ the last-learned foreign language is replaced by the newly selected neighbour language. If this upper limit is set equal to one, we go back to the model of monolingual speakers. In all cases, x is the fraction of neighbours of i speaking as their mother tongue the mother language of site i. For language diffusion we took $q=0.9$ throughout, and for language change mostly $p=0.01$. Thus one iteration may correspond to about one human generation.

We start with a fragmented distribution of mother languages and no foreign languages, except that one particular language is spoken initially by ten percent of the people, randomly spread over the lattice. Then we check if this "lingua franca" finally (after at most 10^5 iterations) is spoken by about everybody: transition from fragmentation to dominance of initially favoured language. (If another language dominated we count this case as fragmentation.)

For ten 50×50 lattices, this transition happened up to $p=0.04$ if bilingualism is allowed, while for monolinguals it happens up to the larger changing rate $p=0.09$: Bilingualism makes dominance of one language less stable against continuous changes; see also [38]. For Q=5 instead of 3 this limit moves from 0.04 to 0.05, while for Q=3,F=16 it is about 0.03. Fig. 7 shows for 8000×8000 lattices the time dependence of the fraction of people speaking the largest and the second-largest language, separately for mother tongue and foreign languages; the comparison for the case without bilinguals is restricted to the mother language and shows that dominance is reached faster without bilinguals.

In these simulations, after a short time everybody speaks two languages, and if up to ten languages are allowed, then again after a short time everybody speaks ten languages. This is nice but unrealistic. In order to take into account that people forget again foreign languages which were learned but not used, or give up learning a foreign language when the need for it dissipates, we assume that at every time step each speaker (more precisely,

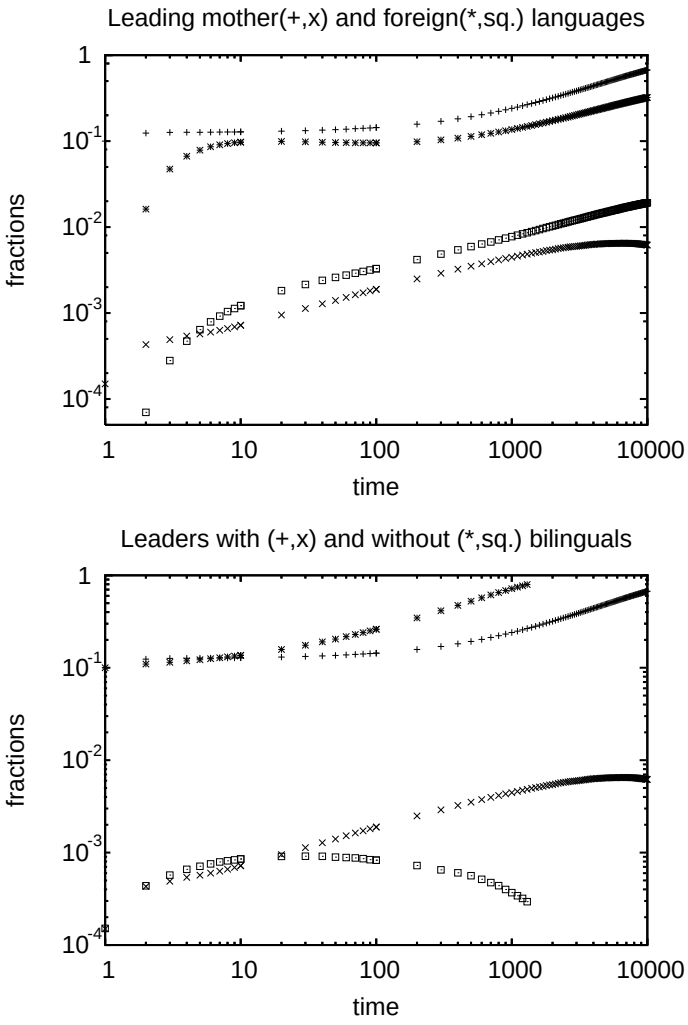

Figure 7: Largest and second-largest languages in the Schulze lattice at $p=0.01, q=r=0.9$ in a 8000×8000 lattice without migration. Part a includes bilinguals, part b compares only the mother languages with and without bilinguals.

each lattice site) may give up the last-learned foreign language if none of the neighbours at that time speaks this language. This forgetting happens with a probability between zero and five percent, fixed for each site randomly and independently at the beginning.

In addition we included migration via exchanging locations: A speaker or family exchanges residence with a randomly selected neighbour, and both carry their languages with them. This happens at each iteration with a probability d; physicists call d the diffusion constant. Fig. 8 shows that appreciable migration can drastically speed up the growth of the lingua franca from having an initial advantage of being spoken by ten percent of the population to being dominant.

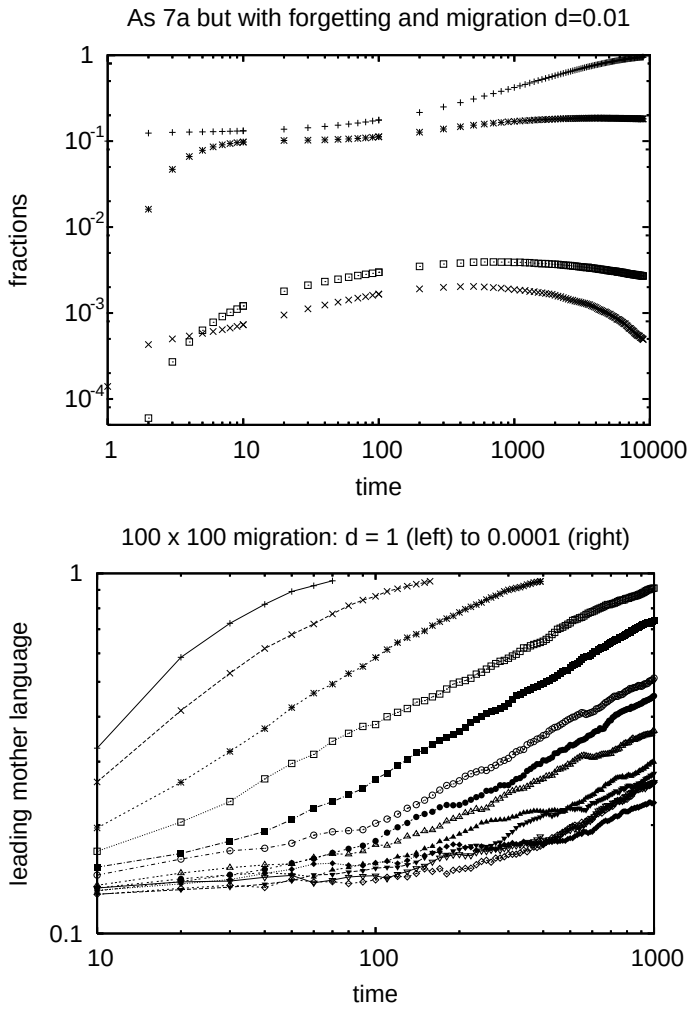

Figure 8: Part a: As Fig. 7a, but including a migration probability d of 1 % and a site-dependent forgetting probability between zero and 5 %, on a 6000×6000 lattice. Part b shows the drastic speed-up of dominance if the migration probability d is enhanced: $d=1,0.5,0.2,0.1,\ldots,0.0005,0.0002,0.0001$.

3 Viviane model

3.1 Definition

The model of Viviane de Oliveira et al. [10] has become known as the Viviane model (following the Brazilian naming practices). It simulates the colonization of an uninhabited continent by people. Each site j of an $L\times L$ square lattice can later be populated by c_j people; this carrying capacity c_j is an integer, selected randomly between 1 and some upper limit $m\sim10^2$. Initially only one site i is occupied by c_i people.

Then at each time step, one randomly selected empty lattice neighbour j of occupied sites becomes occupied with probability c_j/m by c_j people. Thus after some time the whole lattice becomes occupied and the simulation stops. In contrast to the Schulze model, the Viviane model is a growth process and not one eventually fluctuating about some equilibrium.

Languages have no internal structure and are simply numbered 1, 2, 3,,, with 1 being the number of the language spoken on the originally occupied site. All people within one lattice site speak the same language. First, if a new site has been colonized the language spoken there is taken from one of the occupied neighbours k, proportional to the fitness F_k of that neighbour site k. This fitness is the total number of people anywhere in the lattice speaking the language of k, except that it is bounded from above by a maximum fitness M_k fixed randomly between 1 and some $M_{\max} \sim 10^3$. Second, mutations (language change) are made with probability α/F_j on the freshly occupied site j only, from the selected language of neighbour k. A mutation means that a new language is created which gets a new number not used previously.

In this way, the flight (shift) from small languages and the mutations (change), which were two separate processes in the Schulze model, are combined into one process; and this process also is a transfer (diffusion) process which in the Schulze model was dealt with separately. Thus here we have only one free parameter α, the mutation coefficient, instead of three parameters p,q,r in the Schulze model.

Variants also allow mutations later, after a site is occupied. Or a language is characterized by a string of F bits ($Q=2$ in the Schulze notation) and only different bit-strings count as different languages [39]. Or the capacities c_k are not homogeneously distributed between 1 and m but more often small than large, with a frequency proportional to $1/c_j$, as long as it is not larger than the maximum m. A lot of computer time then is saved if after the occupation of the new sites one selects two of its occupied neighbours and takes the language from the one with the bigger capacity; if only one neighbour is occupied then its language is taken over.

3.2 Results

In contrast to the Schulze model, the Viviane model gives languages spoken by 10^9 people if a sufficiently large lattice is used. The language size distribution N_s has a maximum at moderately small language sizes s. However, instead of a round parabola as in Fig. 1, the log-log plot of N_s versus s gives two straight lines meeting at the maximum, meaning one power law for small s (where N_s increases with s) and another power law for large s (where N_s decreases with s). So, not everything is solved yet.

Crucial progress was made by Paulo Murilo de Oliveira (not the same family as Viviane de Oliveira), who introduced the above-mentioned modifications [39]: Languages (grammars) are characterized by bit-strings ($Q=2$) of length $F \simeq 13$ and count as different only when their bit-strings differ; the carrying capacities c are selected with a probability $\propto 1/c$, and the newly colonized site gets the fitter language of two previously occupied

neighbours. Now the distribution is roughly log-normal, Fig. 9, with enhancement for very small sizes; the total population and the total number of languages can be made close to the present reality, the maximum of the parabola in a double-logarithmic plot (with binning by factors of two in s) is near $s \simeq 10^4$, while the largest language is spoken by 10^9 people, similar to Mandarin Chinese.

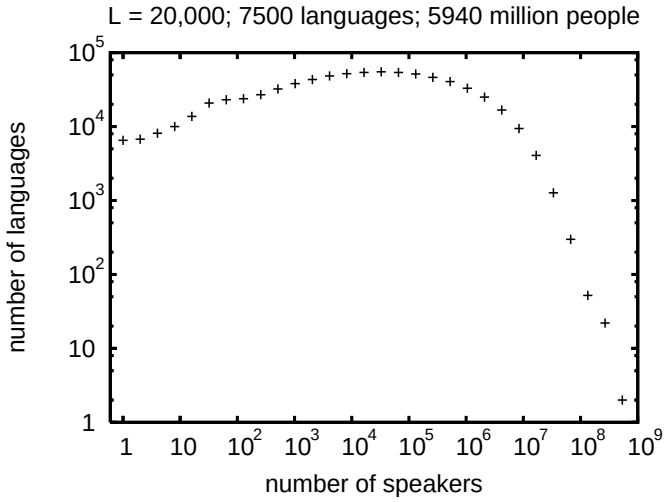

Figure 9: Variation of the number N_s of languages spoken by s people each in the Viviane model, as modified and published in [39].

Fig. 10 shows for both the modified Viviane model and the Schulze lattice model that languages in general are less similar to each other if they are widely separated geographically, in agreement with reality [28, 29]. Note the difference in scales: One lattice constant (distance between nearest neighbours) corresponds to about one kilometer in the Viviane model and 1000 kilometers in the Schulze model, if Fig. 10 is compared to reality [28]. The bottom part of Fig. 10 shows the variation more quantitatively, averaged over ten large lattices: if $d(r)$ is the difference at distance r, then the bottom part shows $[d(\infty)-d(r)]/d(\infty)$ semilogarithmically. In this way we see slight deviations from a simple exponential decays towards zero.

Also the classification of different languages into one family, like the Indo-European languages, has been simulated with moderate success. Following the history of the mutations during the colonization, a language tree like can be constructed in the unmodified version (Fig. 11.15 in [11]). One can imagine that this is Latin, splitting up into Romanian, Italian, Spanish and French, with Spanish then splitting into Castilian, Galician and Catalan, and Catalan mutating further into Majorcan. (Many small branches were omitted for clarity.) More quantitative information is obtained from the modified bit-string version [39]. The mutated language on a newly occupied site starts a new family if it differs in two or more bits from the bit-string characterizing the historically first language of the old family. The size distribution of language families in Fig. 11 agrees in its cen-

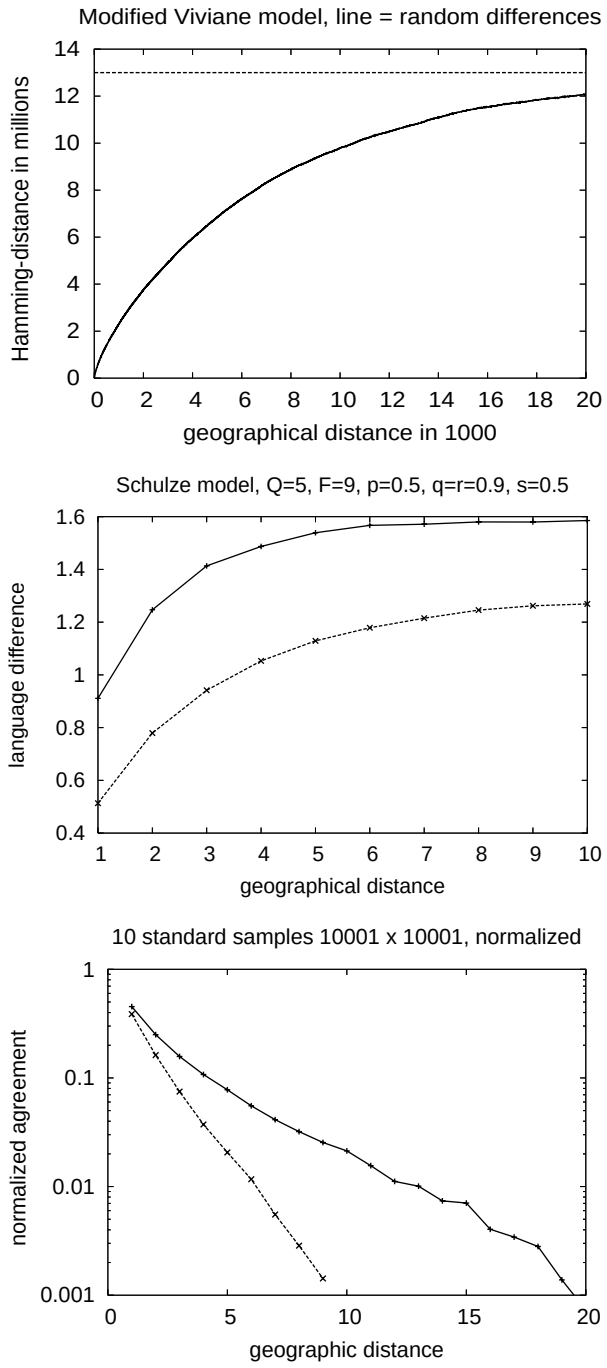

Figure 10: Language differences (in arbitrary units [28]) as a function of geographical distance in the Viviane model (top) and the Schulze model (centre). The horizontal line corresponds to completely uncorrelated languages, that means if all feature values (= bits) were selected randomly. In the central and bottom part, + and x correspond to start with fragmentation (+) and with dominance (x). From [28, 32]. See text for bottom part.

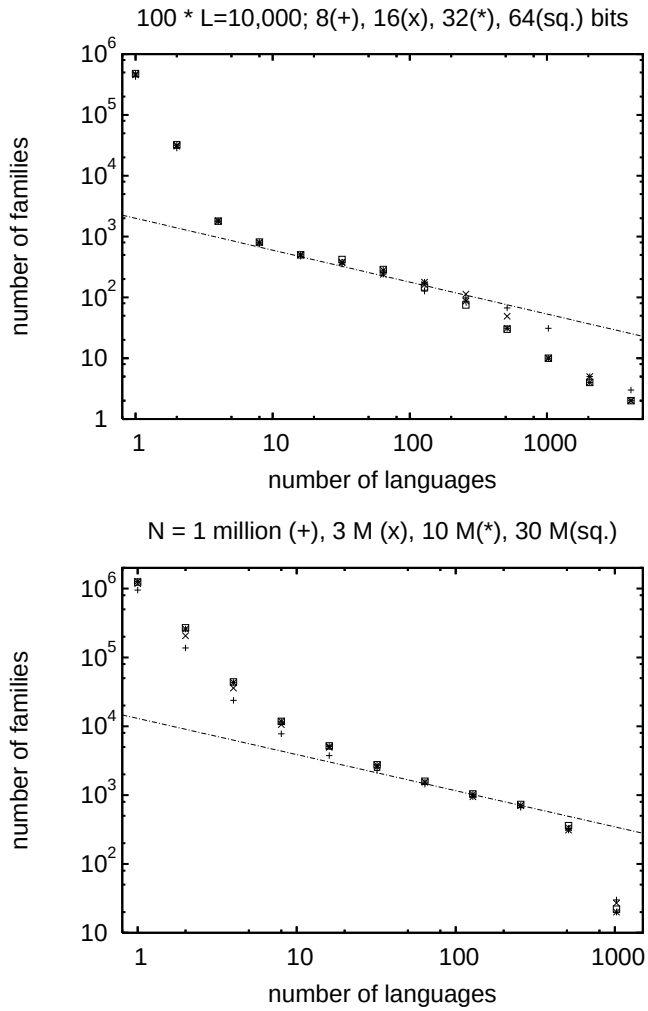

Figure 11: Number of families as a function of the number of different languages in this family [32]. Top: Modified Viviane model for various lengths of the bit-strings. Bottom: Schulze model with $Q=5, F=8$ on directed Barabási-Albert scale-free network, $p=0.5, q=0.59, r=0.9$ for various population sizes.

tral part with the empirically observed [40] exponent –0.525 and is independent of the length $F=8$, 16, 32 or 64 of the bit-strings for the Viviane model and independent of the population size for the Schulze model.

(Mathematical note on rank plots: If the number n_ℓ of families containing ℓ languages decays for large ℓ as $1/\ell^\tau$, then the number $n>\ell$ of families containing at least ℓ languages decays only as $1/\ell^{\tau-1}$, and the same is true for the number of families with ℓ to 2ℓ languages, as plotted in Fig. 11: $\tau=1.525$. If $r=1$ denotes the largest family, $r=2$ the second largest, etc, we identify the rank r with $n>\ell$ and invert the above relation to $\ell \propto 1/r^{1/(\tau-1)} \propto 1/r^{1.9}$ as found empirically [40].)

4 Other models

Years before physicists invaded en masse the field of linguistics, Nettle [36] already wrote down a differential equation for the number L of languages,

$$\frac{dL}{dt} = 70/t - L/20,$$

where time t is measured in millennia. For long times, only one language (mathematically: zero languages) will remain; however that time lies far in the future. A more detailed splitting mechanism was introduced by Novotny and Drozd [42] for the emergence of new languages from one mother language, and this gave a log-normal distribution of language sizes, in agreement with reality except presumably at the smallest sizes [17]. In the same spirit of looking at languages as a whole, ignoring the individuals, are the very recent models of Tuncay [14], who coupled a splitting mechanism with random multiplicative noise in the size of the growing population, plus an extinction probability, and found the desired roughly log-normal size distribution for the simulated languages. He also checked the lifetimes of the simulated languages, and the language families. An "early" attempt to apply the Ising model of statistical physics to linguistics [43], see also [44], had little following.

The same Nettle [36] as well as Culicover and Nowak [45] applied the Social Impact Theory of Latané [46] to language change. Each site on a square lattice has a variable $\sigma_i = \pm S_i$ where the positive number S_i gives the social status (influence) of that person. The dynamics then follows the majority influence of the neighbours, similar to an Ising model:

$$\begin{aligned} &\sigma_i(t+1)/S_i \\ &= \text{sign}\,[b_+ N_+^{a-1} \sum_k \sigma_k(t)/r_{ik}^2 - b_- N_-^{a-1} \sum_k{}' \sigma_k'(t)/r_{ik'}^2], \end{aligned}$$

where N_+ is the number of lattice sites with positive and N_- the one with negative σ, a is a free exponent giving nonlinear influence, the b are two bias factors, r_{ij} is the geometric distance between the lattice sites i and j, and k runs over all sites with positive σ while k' runs over the remaining sites with negative σ. Thus σ becomes positive if the positive influence is larger than the negative one, and it becomes negative in the opposite case.

In the special case $a=1, b_+ = b_i$ the above rule reduces to

$$\sigma_i(t+1)/S_i = \text{sign} \sum_j \sigma_j(t)/r_{ij}^2,$$

with j running over all sites. This rule becomes an Ising model if there are no status differences, $S_i = 1$ for all i. The two choices for σ can be two opinions, two languages, two language features, or any other binary option.

Nettle introduces an age structure into this model through five age intervals. Children in the first interval have no influence on others but are influenced by them, adolescents influence others and are influenced by them, and adults (ages 3 to 5) no longer change but they influence others. When they die at age 5, they are replaced by children of age 1. (The learning at ages 1 and 2 has an error rate assumed to be typically 5 %.)

As a result, without differences in S and b, he finds the two possibilities about equally often ("paramagnet") for $a \leq 1/2$, while one choice dominates ("ferromagnet") for $1/2 < a \leq 1$. He finds that once a clear majority adheres to one choice, this choice remains the majority, and that for a change in the majority choice one thus needs differences in status S and bias b. However, the standard two-dimensional Ising model shows such switches of the sign of the spontaneous magnetization without any of these complications, Fig. 12, provided the temperature is only slightly below the critical Curie temperature. There the rate of change decreases exponentially with increasing lattice size.

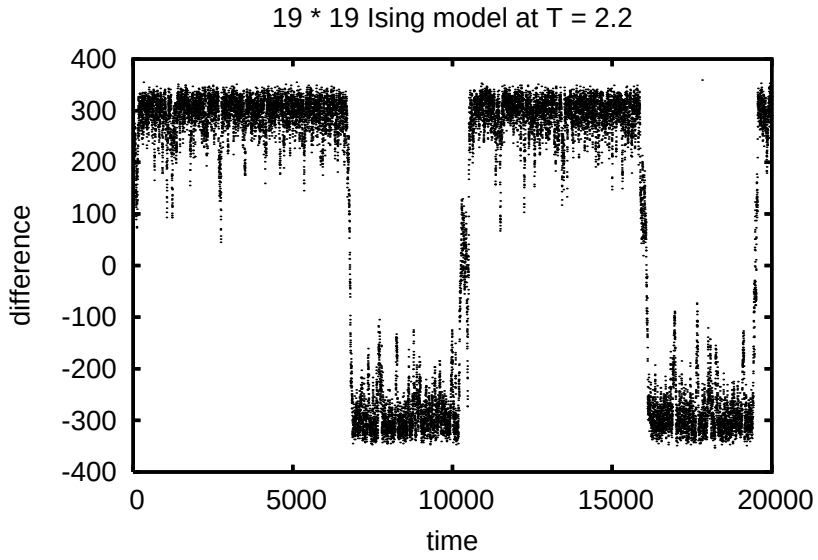

Figure 12: $N_+ - N_-$ in an Ising model on a small square lattice at a temperature (noise level) three percent below the critical temperature. For temperatures above this critical temperature, $N_+ \simeq N_-$.

Also Culicover and Nowak [45] use the Social Impact Theory, confirm Nettle's results, and study clustering and correlations on the square lattice. They generalize Nettle's version to three, instead of only one, binary variables. Instead of an influence decaying as the inverse squared distance, they assume it to be constant up to a maximum distance, and zero beyond that distance; and only a fixed number of teachers are selected randomly from that neighbourhood. At the end not all 8 possible languages exist, since some have died out.

Numerous coupled differential equations were studied by scientists coming from theoretical chemistry, mathematics and computer science [48] for the purpose of language

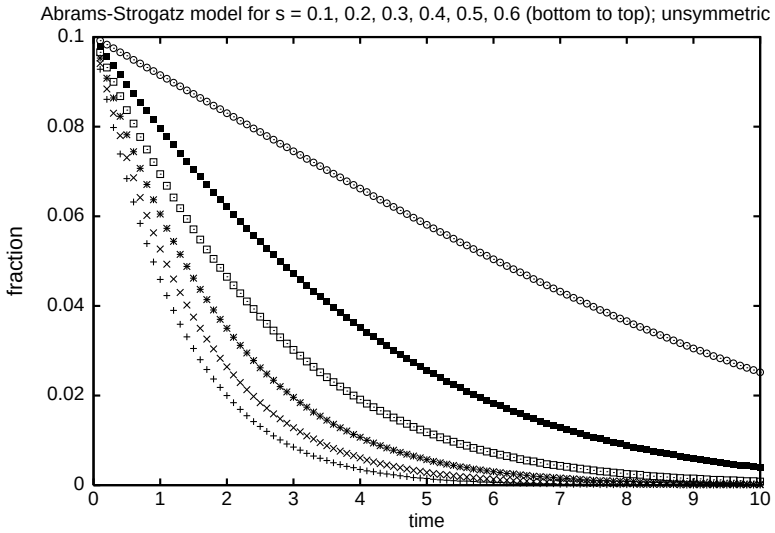

Figure 13: Abrams-Strogatz model: Fraction of people speaking language X, versus time, for an initial concentration of only ten percent, $a=1.31$, and various prestige values s of this language. For $s \geq 0.7$ this language finally wins over the other language Y. From [3].

learning by children. They have also been applied [3] to the competition of up to 8000 languages of adults, but since the original authors have to our knowledge not followed this re-interpretation of their learning model we now refer to [3,48] for details and results.

It was the population dynamics of Abrams and Strogatz [4] which started the small avalanche of physics papers on language competition. They assume two languages X and Y, spoken by the fractions x and $y=1-x$ of a fixed population with a time dependence

$$dx/dt = yx^a s - xy^a(1-s),$$

with a status or prestige variable S which is close to one if X has a high prestige and close to zero for low prestige of X. The neutral case is $s=1/2$. The exponent $a=1.31$ was fitted to some empirical data of how minority languages decay in size. If a is replaced by unity we arrive at the logistic equation of Verhulst from the 19th century, which was applied to languages by Shen in 1997, as cited in [47].

Fig. 13 shows the resulting $x(t)$ if X is spoken initially by a minority of ten percent only. Then for low, neutral or slightly higher status s of its language X, the fraction decays further towards zero, but for a higher status like $s=0.7$ it finally wins over and is spoken by everybody (not shown). This may correspond to the influence of a colonial power; indeed in France today most people speak French as a result of the Roman conquest of more than two millennia ago, and in the New World many of the native languages have become extinct in the last five centuries since the Europeans arrived there.

This Abrams-Strogatz approach was soon generalized to a lattice by Patriarca and Lepännen [4], and later to populations with bilingual speakers [5,38], coexistence of the two languages [8] as well simulations based on individuals [13].

Such agent-based simulations were also made by Kosmidis et al. [6] who gave each person a string of 20 bits. The first 10 belonged to one language, the last 10 to another language. In this way they were able to simulate people speaking, more of less correctly, one or two languages. One could also interpret their model as one for English which, due to the conquest of England by the Normans in 1066, has a mixture of Germanic (Anglo-Saxon) and French words [52].

Finally, Schwämmle [9] also used bit-strings, but to describe biological ageing through the Penna model. The child can learn the language from the mother, the father, or both, thus also allowing for bilinguals. This model accommodates the fact that languages are learned easier in youth than at old age. In this way it builds a bridge between language competition and language learning [48].

5 How physics may inform linguistics: prospects for future research

As the research described above has progressed a larger design has become apparent, which consists in an empirical side looking for quantitative distributions involving languages [28, 40, 49] and a development of models simulating similar quantitative distributions. The hope is that as more and more quantifiable relations in and among languages are discovered and simulation models are developed which can adequately replicate these distributions, the simulation models will of necessity become more and more adequate as models of actual languages, and could therefore be employed for purposes beyond the ones for which they were designed. For instance, the revised Viviane model, which was designed to capture the distribution of speaker populations and the population of languages within families [32], could potentially be employed for investigating absolute rates of language change, an issue with which linguists are very much concerned [50], inasmuch as knowledge of how fast languages change could provide us with a way to date prehistoric events involving people speaking given reconstructed languages. Thus, a strand of research where linguists and physicists can and will continue to cooperate is the search for quantifiable distributions on the one hand and the fine-tuning of models which can adequately simulate an increasing range of such distributions.

One anonymous referee formulated a point of criticism which many readers of the literature reviewed here may share: "There appears to be a substantial gap between the sociolinguistic processes being modeled and the theoretical mechanisms that underpin the models of them." It is true that the models are all simplifications, but reductionism is at the heart of any scientific enterprise. It should also be stressed that even if there is such a field as sociolinguistics, where people study the social conditions for variation and change within languages, this does not mean that all is known about these conditions. Just like no good explanation for why apples fall to the ground when you drop them has been found within physics, there are many apparently simple problems in linguistics that have not been solved. Still, it is correct that more complicated scenarios as well as

hypotheses about how the behavior of languages depends on various social parameters could and should be taken into account. We review some of these issues in the following.

Apart from some exceptions [5, 6, 38], most work on language competition has assumed monolingual speakers. Since most of the world's population is bi- or multilingual this is clearly not adequate. Language shift will normally involve transitional bilingualism, or bilingualism may persist for centuries without the majority language necessarily replacing minority languages. Diglossia, i.e., the use of different languages for different purposes [51], may help sustain bilingualism. Current models can be extended to investigate under which conditions bilingualism may persist or get reduced to monolingualism. Different kinds of situations can be modeled, such as the replacement of certain, but not all, languages within the domain of the Roman empire, the development of so-called linguistic areas, where several languages share a number of features (e.g., the Balkans, India, Mesoamerica), multilingualism caused by linguistic exogamy (the northwest Amazon region), the shift from one to another lingua franca with retention of minority languages (Mayan immigrants in urban United States shifting from Spanish to English but retaining Mayan languages), etc., and may be applied to situations where prehistoric interaction has left linguistic traces but where the nature of the interaction is unknown (e.g., the sharing of linguistic features around the entire coast of the Pacific Ocean). Section 2.3 above therefore also presented a new extension of the Schulze model to bilingualism.

An area where physicists may wish to try out their hands more is that of language change [34]. Simulations may help linguists come to terms with realities that are accessible through empirical research only in small fragments. Languages develop and change through the interaction of multitudes of agents using large lexical inventories and complex grammars [52]. The kinds of regularities that linguist can identify, such as the regularity of sound changes or directed paths of grammaticalization (roughly, the process whereby separate words become part of the morphology, cf. [53]), are mostly accessible only to a retrospective view, through the comparison of language stages dozens or hundreds of years apart. What lies in between is a flux whose behavior is not easy to understand. 19th century historical linguists, with their focus on regular sound changes, lived in a universe of clean equations such as Latin p = English f (as in *pater* = *father*). Finding such "sound laws" is still important in the methodology of historical linguistics [54]. However, the advent of 20th century sociolinguistics (e.g., [55]), with its focus on the social mechanisms behind sound changes, complicated the picture, much as the picture gets complicated when one moves from clean Newtonian physics to modern statistical physics. Unlike physicists, linguists investigating the way that languages change have taken little recourse to simulations that might help them understand the complexity of how language change or shift percolates within a community. For instance, a leading sociolinguist has argued that "networks constituted chiefly of strong ties function as a mechanism to support minority languages, resisting institutional pressures to language shift, but when these networks weaken, language shift is likely to take place" (p. 558 in [56]). This hypothesis is based on just a few case studies, and such case studies are extremely costly and cannot even begin to cover the multitude of different situations that

actually obtains. In addition to the strength of network ties, other important parameters are presumably the size of the group speaking the minority language, geography, prestige of one as opposed to the other language, economic gain involved in shifting language, age- and gender-determined mobility, and maybe more. The behaviors of such parameters can be investigated in simulations (e.g., for geography see [11, 31, 41]). Network theory in many different guises has also been applied to the study of linguistic variation among speakers of a single language (see pp. 116-136 in [57] for a review). It has been found, for instance, that individuals nurturing local speech varieties tend to be the ones most integrated in the local societies through different networks; and it has been suggested that a type of social organization characterized by overlapping close-knit networks will inhibit change, while mobility will lead to more change, explaining why some languages (e.g., Icelandic) seem more conservative than others (e.g., English). Such a hypothesis eminently lends itself to be tested in simulations. On a wider scale, changes may spread between different cities, skipping intermediate territory, as in the case of certain vowel shifts in the northern United States [58]. Network analysis seem to have superseded the more traditional sociolinguistic works that followed in the wake of [59], where the focus was on correlating variation in speech with parameters such as class and gender. Still, hypotheses such as the one according to which women cater towards prestige variants in their speech [60] is of potential interest; the effects of such a potential factor may again be investigated through simulations.

Finally, more work needs to be done towards the integration of the modeling of language competition by physicists reviewed here and the modeling of language evolution by computational linguists [45, 61–65]. While physicists have been adept in modeling the interaction among agents but have operated with languages represented only by numbers or bit-strings, computational linguists offer elaborate grammar models. With more complex models of the interior structure of languages carried by agents, research need not be limited to a focus on language competition, but could be extended to issues of language structure itself.

Acknowledgments

Improved family statistics were obtained for the Viviane model by P. M. C. de Oliveira, D. Stauffer, S. Wichmann, S. Moss de Oliveira, submitted, e-print 0709.0868 on arXiv.org, and for the Tuncay model by C. Tuncay, submitted to Int. J. Mod. Phys. C 19 in press (2008). Endangered languages were modelled by J. W. Minett, W. S. Y. Wang, Lingua, in press (2007).

References

[1] W. S. Y. Wang, J. W. Minett, Trans. Philological Soc. 103 (2005) 121, http://www.isrl.uiuc.edu/amag/langev/

[2] R. Selten, J. Pool, S.64, in: R. Selten (Ed.), Game Equilibrium Models IV, Springer, Berlin-Heidelberg, 1992.
[3] D. Stauffer, S. Moss de Oliveira, P. M. C. de Oliveira, J. S. Sá Martins, Biology, Sociology, Geology by Computational Physicists, Elsevier, Amsterdam, 2006.
[4] D. Abrams, S. H. Strogatz, Nature 424 (2003) 900; M. Patriarca, T. Leppänen, Physica A 338 (2004) 296.
[5] J. Mira, A. Paredes, Europhys. Lett. 69 (2005) 1031.
[6] K. Kosmidis, J. M. Halley, P. Argyrakis, Physica A 353 (2005) 595; K. Kosmidis, A. Kalampokis, P. Argyrakis, Physica A 366 (2006) 495 and 370 (2006) 808.
[7] C. Schulze, D. Stauffer, Int. J. Mod. Phys. C 16 (2005) 781 and Phys. Life Rev. 2 (2005) 89.
[8] J. P. Pinasco, L. Romanelli, Physica A 361 (2006) 355.
[9] V. Schwämmle, Int. J. Mod. Phys. C 16 (2005) 1519 and 17 (2006) 103.
[10] V. M. de Oliveira, M. A. F. Gomes, I. R. Tsang, Physica A 361 (2006) 361 and 368 (2006) 257.
[11] C. Schulze, D. Stauffer, in: B. K. Chakrabarti, A. Chakraborti, A. Chatterjee (Eds.), Econophysics and Sociophysics: Trends and Perspectives, Wiley-VCH, Weinheim, 2006, pp. 311. For shorter reviews see AIP Conference Proceedings 779 (2005) 49; Comput. Sci. Eng. 8 (2006) 86.
[12] T. Tşileanu, H. Meyer-Ortmanns, Int. J. Mod. Phys. C 17 (2006) 259.
[13] D. Stauffer, X. Castelló, V. M. Eguíluz, M. San Miguel, Physica A 274 (2007) 835.
[14] Ç. Tuncay, Int. J. Mod. Phys. C 18 (2007) 1061, 1203 and issue 10, e-print 0705.1838 on arXiv.org
[15] G. J. Baxter, R. A. Blythe, W. Croft, A. J. McKane, Phys. Rev. E 73 (2006) 046118.
[16] D. Crystal, Language Death, Cambridge University Press, Cambridge, 2000.
[17] W. J. Sutherland, Nature 423 (2003) 276.
[18] C. Yang, The Infinite Gift–How Children Learn and Unlearn the Languages of the World, Scribner, New York, 2006.
[19] J. H. Greenberg, Some universals of grammar with particular reference to the order of meaningful elements, in: J. H. Greenberg (Ed.), Universals of Language, MIT Press, London, 1963, pp. 73-113.
[20] E. W. Holman, Approximately independent typological features of languages, submitted.
[21] B. F. Grimes, Ethnologue: Languages of the World, 14 ed., Summer Institute of Linguistics, Dallas, TX, 2006, www.ethnologue.org
[22] D. Stauffer, C. Schulze, F. W. S. Lima, S. Wichmann, S. Solomon, Physica A 371 (2006) 719.
[23] E. J. Briscoe, Language 76 (2000) 245.
[24] M. Haspelmath, M. Dryer, D. Gil, B. Comrie (Eds.), The World Atlas of Language Structures, Oxford University Press, Oxford, 2005.
[25] S. Wichmann, E. W. Holman, Assessing temporal stability for linguistic typological features, 2007, submitted.
[26] F. C. Billari, T. Fent, A. Prskawetz, J. Scheffran (Eds.), Agent-based Computational Modelling, Physica-Verlag, Heidelberg, 2006.
[27] H. Goebl, Mitt.Österr. Geogr. Ges. 146 (2004) 247 (in German).
[28] E. W. Holman, C. Schulze, D. Stauffer, S. Wichmann, Linguistic Typology, in press.
[29] L. L. Cavalli-Sforza, W. S. Y. Wang, Language 62 (1986) 38.
[30] G. Barbujani, R. R. Sokal, Proc. Natl. Acad. USA 87 (1990) 1816.
[31] C. Schulze, D. Stauffer, Physica A 379 (2007) 661.
[32] D. Stauffer, C. Schulze, S. Wichmann, in: S. Kolling (Ed.), Beiträge zur Experimentalphysik, Didaktik und computergestützten Physik, Logos-Verlag, Berlin, 2007 (Festschrift Patt).

[33] J. Ke, T. Gong, W. S.-Y. Wang, talk at 5th EVOLANG conference, Leipzig, Germany, March 31, 2004; Int. J. Mod. Phys. C 19 (2008) in press.
[34] S. Wichmann, D. Stauffer, C. Schulze, E. Holman, (2007), preprint.
[35] R. Albert, L. A. Barabási, Rev. Mod. Phys. 74 (2002) 47; X. Castello, R. Toivonen, V. M. Eguiluz, J. Samaräki, K. Kaski, M. San Miguel, e-print 0705.2560 on arXiv.org
[36] D. Nettle, Lingua 108, 95 and 119 (1999); Proc. Natl. Acad. Sci. USA 96 (1999) 3325.
[37] C. Schulze, D. Stauffer, arXiv:0707.0072.
[38] X. Castelló, V. M. Eguíluz, M. San Miguel, New J. Phys. 8 (2006) 308.
[39] P. M. C. de Oliveira, D. Stauffer, F. W. S. Lima, A. O. Sousa, C. Schulze, S. Moss de Oliveira, Physica A 376 (2007) 609.
[40] S. Wichmann, J. Linguistics 41 (2005) 117.
[41] C. Schulze, D. Stauffer, Adv. Complex Syst. 9 (2006) 183.
[42] V. Novotny, P. Drozd, Proc. Roy. Soc. London B 267 (2000) 947.
[43] N. Prévost, The physics of language: toward a phase transition of language change, Ph.D. thesis, Simon Fraser University, Vancouver, 2003.
[44] Y. Itoh, S. Ueda, Physica D 198 (2004) 333.
[45] P. Culicover, A. Nowak, Dynamical Grammar, Oxford University Press, Oxford, 2003.
[46] B. Latané, Am. Psychol. 36 (1981) 343.
[47] W. S. Y. Wang, J. Ke, J. W. Minett, in: C. R. Huang, W. Lenders (Eds.), Computational Linguistics and Beyond, Academica Sinica, Institute of Linguistics, Teipeh, 2004.
[48] M. A. Nowak, N. L. Komarova, P. Niyogi, Nature 417 (2000) 611; N. L. Komarova, J. Theor. Biol. 230 (2004) 227.
[49] S. Wichmann, E. W. Holman, Pairwise comparisons of typological profiles, submitted. For the proceedings of the conference Rara & Rarissima – Collecting and interpreting unusual characteristics of human languages, Leipzig (Germany), 29 March - 1 April 2006; e-print 07040071 at arXiv.org
[50] J. Nichols, Diversity and stability in languages, in: B. D. Joseph, R. D. Janda (Eds.), The Handbook of Historical Linguistics, Blackwell Publishing, Malden/Oxford/Melbourne/Berlin, 2003, pp. 283-310.
[51] C. A. Ferguson, Word 15 (1959) 325.
[52] S. G. Thomason, T. Kaufman, Language Contact, Creolization, and Genetic Linguistics, University of California Press, Berkeley, 1988.
[53] B. Heine, U. Claudi, F. Hünnemeyer, Grammaticalization. A Conceptual Framework, University of Chicago Press, Chicago, 1991.
[54] L. Campbell, Historical Linguistics: An Introduction, MIT Press, Cambridge, 1999.
[55] P. Trudgill, Sociolinguistics. An Introduction to Language and Society, 4th ed., Penguin, London, 2000, pp. 549-527.
[56] L. Milroy, Social networks, in: J. K. Chambers, P. Trudgill, N. Schilling-Estes (Eds.), The Handbook of Language Variation and Change, Blackwell, 2002.
[57] L. Milroy, M. Gordon, Sociolinguistics. Method and Interpretation, Blackwell, Malden, 2003.
[58] W. Labov, The three dialects of English, in: P. Eckert (Ed.), New Ways of Analyzing Sound Change, Academic Press, New York, pp. 1-44.
[59] W. Labov, The Social Stratification of English in New York City, Center for Applied Linguistics, Washington, DC, 1966.
[60] W. Labov, Language Variation and Change 2 (1990) 205.
[61] A. Cangelosi, D. Parisi (Eds.), Simulating the Evolution of Language, Springer-Verlag, Berlin, 2002.

[62] E. J. Briscoe (Ed.), Linguistic Evolution through Language Acquisition: Formal and Computational Models, Cambridge University Press, Cambridge, 2002.
[63] M. Christiansen, S. Kirby (Eds.), Language Evolution, Oxford University Press, Oxford, 2003.
[64] B. de Boer, Computer modelling as a tool for understanding language evolution, in: N. Gontier, J. P. van Bendegem, D. Aerts (Eds.), Evolutionary Epistemology, Language and Culture: A Non-Adaptationist Systems Theoretical Approach, Springer, Dordrecht, 2006, pp. 381-406.
[65] P. Niyogi, The Computational Nature of Language Learning and Evolution, MIT Press, Cambridge & London, 2006.