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

1 Introduction 272
2 Schulze model 272
3 Viviane model 281
4 Other models 286
5 How physics may inform linguistics: prospects for future research 289

*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: