diffusion-fundamentals.org

The Open-Access Journal for the Basic Principles of Diffusion Theory, Experiment and Application

Diffusion of humans out of Africa and the phonemic diversity cline

Joaquim Pérez-Losada[1]*, Joaquim Fort[1,2]

[1]Complex Systems Laboratory, University of Girona, Girona, Spain
[2]Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain
*Presenting author: joaquim.perez@udg.edu

(Received: 2023/06/24, Revised version: 2023/07/14, Published online: 2023/07/14)

Introduction

The study of the origin of human language(s) is particularly challenging due to the lack of direct data on ancient languages [1]. Languages spoken in Africa tend to have the largest phonemic inventories, while those in South America and Oceania have the smallest [2]. Atkinson [2] proposed a serial founder effect (SFE) mechanism to explain this global phonemic cline. According to his hypothesis, small populations of humans migrated carrying only part of the languages used in their original locations and phonemes would be lost. However, Atkinson did not present any model based on diffusion equations neither simulations to back his claim. Our previous simulations [3] suggested that a phonemic serial founder effect could explain the global phonemic cline, assuming both low initial phonemic inventories and a natural rate of increase in phonemic inventory size. Here we explore the alternative hypotheses that languages at the origin of the range expansion had large phonemic inventories, and that low-density populations lose phonemic diversity [1]. The latter assumption aligns with empirical evidence showing that cultural diversity decreases with population size and that smaller populations have higher rates of word loss. We aim to determine if Atkinson's proposed SFE model can generate the observed spatial phonemic cline. Additionally, we investigate the stability of the phonemic signal over long periods of time.

A Serial Founder Effect (SFE) model based on population diffusion

We used UPSID [4] database of languages and phonemes. Our model simulates the dispersal of hunter-gatherer tribes from a single origin, in a two-dimensional grid representing the Earth's surface with 1,000 x 1,000 nodes, and the subsequent evolution and spread of languages spoken by these tribes. The distance between adjacent nodes (r) is set at 50 km to reflect pre-industrial population dispersal patterns. The model considers a population density of 1.2 people/km^2 for hunter-gatherer tribes, with each tribe consisting of approximately 300 individuals speaking a common language. Multiple tribes can speak the same language, and a node can accommodate up to 10 tribes. The language of each tribe is encoded as a binary string representing the presence (1) or absence (0) of phonemes. We assume that languages spoken in low-density populations, as encountered at the forefront of a wave of expansion, experience phonemic evolution through random phoneme loss. Each time step represents one generation, with value T = 32 years. The new population number density $p(x, y, t + T)$ in each cell and time, is computed through four steps: (i) Dispersal: a fraction p_e of tribes remains in the original node (for ethnographic data, $p_e = 0.38$), while the remainder disperse randomly to neighboring nodes, i.e. [5]:

$$p(x, y, t + T) = R_0 \Bigg[p_e p(x, y, t) + (1 - p_e)\left(\frac{1}{4}p(x - r, y, t) + \frac{1}{4}p(x + r, y, t) + \frac{1}{4}p(x, y - r, t) + \frac{1}{4}p(x, y + r, t)\right)\Bigg] \quad (1)$$

(ii) Reproduction: new tribes are generated at a net fecundity rate (from ethnography, $R_0 = 1.4$), with the resulting number adjusted to the nearest integer, (iii) Vertical transmission: information on the number of phonemes in a language is copied from the parent tribe to the child tribe without any changes, and finally (iv) Mutation: phonemic loss occurs in low-density cells, where a randomly selected phoneme in the parent language is converted from "1" (present) to "0" (absent) in the child language. This process allows new languages to emerge at nodes with a population density below saturation.

The model runs for a total of 72,960 years, representing approximately 2,280 generations, which corresponds to a reasonable estimation of the time span since the Out-of-Africa dispersal. Two initial conditions are considered: (a) initial languages have around 71 phonemes, similarly to present-day languages with the largest phoneme inventories (e.g., click languages); (b) the initial languages have a range of 35 to 40 phonemes, reflecting the languages spoken today in the region of origin of the out-of-Africa dispersal.

Results and discussion

Figure 1 displays the number of phonemes versus distance from the likely origin of the Out-of-Africa dispersal, for a phonemic loss time of 80 generations per phoneme. As mentioned above, two scenarios are considered: languages with a similar number of phonemes to present-day click languages (66–76 phonemes, open circles, blue solid line), and languages spoken today in the regions of origin of the Out-of-Africa dispersal (35–40 phonemes, squares, red solid line). For the second case, the slope and intercept of the simulated cline closely matched the observed values (green dash-dotted and dashed lines). Note that the initial number of phonemes had minimal impact on the slope of the simulated clines. This finding implies that regardless of the phonemic diversity and mixture of languages at the onset of the Out-of-Africa dispersal, the simulated slope agrees with the observed one. Also, the simulations demonstrate that the SFE signal for phonemic diversity does not disappear even over a deep temporal scale of 300,000 years, contradicting claims that such a signal would rapidly diminish.

Conclusions

A SFE model, combined with the assumption that languages have a higher probability of phonemic loss in low-density populations, can generate the observed global cline. These findings challenge previous assertions regarding the disappearance of SFE signals.

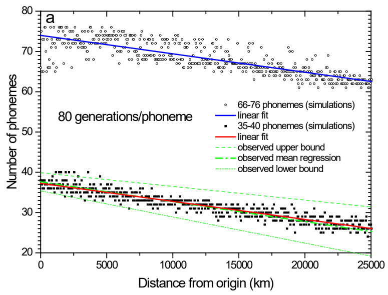

Fig. 1. Simulated phonemic cline (i.e., number of phonemes versus distance from Atkinson's putative origin of the Out- of-Africa dispersal). Two instances of the same model are shown in which at the onset of the Out-of-Africa dispersal, initial languages had 66–76 phonemes (open circles) and 35–40 phonemes (squares). The observed global phonemic cline is also shown (green dash-dotted and dashed lines). Adapted from Ref. [1].

References

[1] J. Pérez-Losada, J. Fort, A serial founder effect model of phonemic diversity based on phonemic loss in low-density populations, PLoS ONE 13(6) (2018) e0198346.

[2] QD. Atkinson, Phonemic diversity supports a serial founder effect model of language expansion from Africa, Science 332 (2011) 346–349.

[3] J. Fort, J. Pérez-Losada, Can a linguistic serial founder effect originating in Africa explain the worldwide phonemic cline?, J R Soc Interface 13 (2016) 20160185, 1–9.

[4] I. Maddieson, K. Precoda, Updating UPSID, UCLA working papers in Phonetics 74 (2011) 104–111.

[5] J. Fort, J. Pérez-Losada, and N. Isern, Fronts from integrodifference equations and persistence effects on the Neolithic transition, Phys. Rev. E 76 (2007) 031913, 1-10.