Support for a linguistic serial founder effect originating in Africa

J. Perez-Losada[1] and J. Fort[1,2]

[1] Complex Systems Lab and Physics Department, Universitat de Girona, Girona, (Spain)

[2] ICREA, Passeig Lluís Companys 3, 08010 Barcelona, Catalonia, (Spain)

Introduction

Present African languages have some of the largest phonemic inventories in the world. In contrast, languages with the smallest inventories are spoken in South America and Oceania[1]. The latter were the last regions to be occupied by modern humans after the out-of-Africa range expansion[2]. Atkinson[1] hypothesized that during the out-of-Africa human dispersal about 70 ky ago[3], phoneme inventory size could have decreased due to a serial founder effect. A serial founder effect had been previously proposed to explain the observed decrease in *genetic* diversity with distance from Africa[4]. Thus, although the correlation of *phonemic* diversity on distance from Africa is not as strong as that of *genetic* diversity on distance from Africa, both clines might in principle be due to the same process, namely the out-of-Africa expansion[5].

Perreault and Mathew[6] have proposed the possibility that some phonemic inventories may increase over time. They argued that the fact that languages with more speakers tend to have more phonemes[7] suggests that new phonemes are more likely to appear in large populations. Perreault and Mathew[6] considered the population that arrived to the coastal region spanning from southern India to the Malay Peninsula during the out-of-Africa dispersal of modern humans from Africa, that dispersed across Mainland Southeast Asia (densely populated, region B) and also into the Andaman Islands (scarcely populated, region C), located in the Indian Ocean. Perreault and Mathew[6] noticed that, despite both regions were colonized at similar times and from the same place, the languages presently spoken in region B have more phonemes, than the corresponding value in region C. They assumed that for the very large population in region B, some mechanism of phoneme accumulation might have gradually increased the phonemic inventories of the new languages arising from diversification. They estimated a linear phonemic accumulation rate between 0.26 and 0.38 phonemes/ky.

Here we simulate the human dispersal out of Africa about 70 ky ago[3]. Our idea is very simple[8]. If the hypothesis of a serial founder effect were valid, using reasonable assumptions, anthropologically realistic parameter values for growth and dispersal, an archaeologically and genetically realistic value for the time spent from the onset of the out-of-Africa range expansion, and an empirical value for the phonemic accumulation rate, we should be able to obtain a worldwide distribution of phonemic diversity similar to the observed one. On the contrary, if such an approach leads to a distribution different from the observed one, then we will conclude that a simple serial founder model cannot explain the observed pattern.

The observed cline

First, we checked that the presence of a global cline in phonemic diversity was present in the dataset. In Fig. 1 the number of phonemes of each language is plotted as a function of its average distance from the putative African origin proposed by Atkinson[1]. The negative correlation between number of phonemes and distance is statistically significant (r = -0.313, p<0.001). Thus, there is a cline.

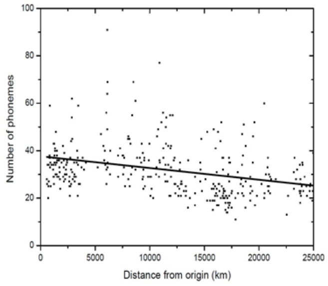

Fig. 1 Plot of the number of phonemes of 366 present languages versus their distances from the most likely origin of the out-of-Africa dispersal. Phonemic data have been obtained from the UPSID database. The linear fit (straight line) has slope= $-(3.4\text{-}6.5)\cdot10^{-4}$ phonemes/km, intercept=35.4-39.9 phonemes (95% confidence-level intervals), r = -0.313, n = 366. The slope is very highly significantly different from zero (P < 0.001).

The model

We have represented the surface of the Earth by a square grid of 1,000 x 1,000 nodes. Speaker community size is fixed for each tribe (400 people), but several tribes may speak the same language (and tribes speaking the same language may be located either at the same spatial cell or at different ones). The language of each tribe is represented by a binary string of 60 digits equal to "0" (absence of a particular phoneme) or "1" (presence of that phoneme). We assume that the original language(s) spoken by the modern humans who started the out-of-Africa dispersal had few phonemes.

We also assume that the last out-of-Africa dispersal started about 70 ky ago[3]. In our simulations, each time step T corresponds to one generation. For each time step and at each occupied node, the following sequence of three processes (dispersal, reproduction and vertical transmission) is repeated:

(i) Isotropic Dispersal. A randomly-selected number of tribes, equal to the nearest integer to the product of the initial number times $(1-p_e)/4$, disperses into each of the 4 neighboring nodes. The rest of the population stays at the original node.

(ii) Reproduction. We use a net reproductive rate (births minus deaths) $a = 1\,\%$ per year = 0.01/y, which has been estimated from archaeological data[9].

(iii) Vertical transmission. The string of values "0" or "1" (indicating the absence or presence of a phoneme) from a randomly-selected old tribe are all passed to a randomly-selected new tribe in the same node.

New languages are formed by the phonemic accumulation process, by changing a value "0" (absence of a phoneme) into "1" (presence of that phoneme). Two subsequent accumulation steps are separated by a longer time interval (e.g., every 82 generations, see below) than steps (i)-(iii) above, which take place once per generation.

In our model we assumed that only languages with high speaker *densities* increase their number of phonemes. From the beginning of the simulation, the processes of accumulation of phonemes takes place in all cells where the population density has already reached carrying capacity, by randomly adding one extra phoneme to each language (i.e. a digit "0" is turned into "1") at the prescribed rate of one phoneme 82 generations (higher bound) or 120 generations (lower bound). These bounds correspond, respectively, to the upper (0.38 phonemes per ky) and lower (0.26 phonemes per ky) bounds estimated by Perreault and Mathew[6].

Results

Figures 2-3 show the results of the simulations, i.e. the number of phonemes of each language as a function of its distance from the origin after 70 ky[3] of the onset of the out-of-Africa dispersal. The two plots display a cline of decreasing phonemic diversity with increasing distance from the presumed origin of the out-of-Africa expansion. The slopes of all three clines are very highly significantly different from zero (P<0.001).

The intercept of the linear fit in Fig. 2 is lower than in Fig. 3, as expected because fewer phonemes are added per unit time in Fig. 2 than in Fig. 3. The intercept according to the observed data (caption to Fig. 1) indicates that, on average, present African languages have about 35.4-39.9 phonemes (95% confidence-level interval). This range is consistent with the 39.5-40.1 phonemes predicted by the intercept of the linear fit to the cline simulated using the upper bound of the observed accumulation rate (Fig. 3). We note, however, that the lower bound of the observed accumulation rate (Fig. 2) yields an intercept (31.2-31.6 phonemes) that is too small to be consistent with the observations (35.4-39.9 phonemes, from Fig. 1).

Acknowledgements

Funded by MINECO (FIS-2012-31307 & SimulPast-CSD-2010-00034), FBBVA (Neodigit-PIN2015E) and ICREA (Academia award, JF).

Results

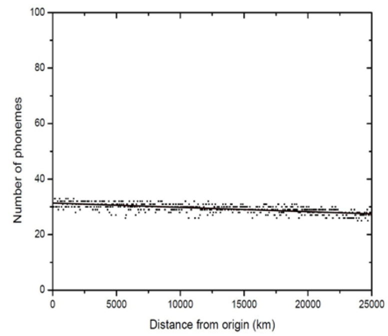

Fig. 2 Simulated number of phonemes at present ($t = 2{,}280$ generations≈70 ky) versus distance from the most likely origin of the out-of-Africa dispersal, for the observed lower bound of the rate of phonemic accumulation (0.26 phonemes per ky, i.e. 1 phoneme added each 120 generations). The linear fit (straight line) has slope= $-(1.4\text{-}1.7)\cdot10^{-4}$ phonemes/km, intercept= 31.2-31.6 phonemes (95% confidence-level intervals), r = -0.671, n = 501, P < 0.001.

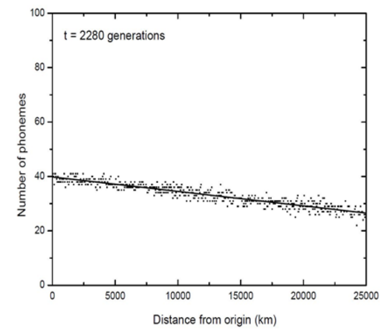

Fig. 3 Simulated number of phonemes at present ($t = 2{,}280$ generations≈70 ky) versus distance from the most likely origin of the out-of-Africa dispersal, for the observed upper bound of the rate of phonemic accumulation (0.38 phonemes per ky, i.e. 1 phoneme added each 82 generations).). The linear fit (straight line) has slope= $-(5.1\text{-}5.5)\cdot10^{-4}$ phonemes/km, intercept= 39.5-40.1 phonemes (95% confidence-level intervals), r = -0.929, n = 501, P < 0.001.

Conclusions

It has been proposed that a serial founder effect could have caused the present observed pattern of global phonemic diversity. Here we have presented a model that simulates the human range expansion out of Africa and the subsequent spatial linguistic dynamics until today[8]. It does not assume copying errors, Darwinian competition, reduced contrastive possibilities, or any other specific linguistic mechanism.

We have shown that the decrease of linguistic diversity with distance (from the presumed origin of the expansion) arises under three assumptions, previously introduced by other authors: (i) an accumulation rate for phonemes; (ii) small phonemic inventories for the languages spoken before the out-of-Africa dispersal; (iii) an increase in the phonemic accumulation rate with the number of speakers per unit area.

Numerical simulations show that the predictions of the model agree with the observed decrease of linguistic diversity with increasing distance from the most likely origin of the out-of-Africa dispersal. Thus the proposal that a serial founder effect could have caused the present observed pattern of global phonemic diversity is viable, if three strong assumptions are satisfied.

References

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

[2] Stewart, J. R. & Stringer, C. B. Human evolution out of Africa: the role of refugia and climate change. Science 335, 1317-1321 (2012).

[3] Oppenheimer, S. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Phil. Trans. R. Soc. B 367, 770-784 (2012).

[4] Deshpande, O., Batzoglou, S., Feldman, M. W., Cavalli-Sforza, L. L. A serial founder effect model for human settlement out of Africa. Proc. Ry. Soc. B 276, 291-300 (2009).

[5] Henn, B. M., Cavalli-Sforza, L. L. & Feldman, M. W. The great human expansion. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17758-17764 (2012).

[6] Perreault, C. & Mathew, S. Dating the Origin of Language Using Phonemic Diversity. PLoS One 7, e35289 (2012).

[7] Hay, J. & Bauer, L. Phoneme inventory size and population size. Language 83, 388–400 (2007).

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

[9] Conolly, J., Colledge, S. & Shennan, S. Founder effect, drift, and adaptive change in domestic crop use in early Neolithic Europe. J. Arch. Sci. 35, 2797-2804 (2008).