1. Introduction

Quentin Atkinson (2011) proposes a theory with the important implication that human language originates and expands from Africa. His theory, however, does not seem fully convincing as it stands because, first, he basically tests only a single logical consequence of his theory (whereas normally theories are tested against a variety of such consequences), and secondly, the body of data he provides may also be reasonably accounted for in different ways. Therefore, additional evidence is required to confirm or disconfirm the theory in a more conclusive manner.

Atkinson’s article in *Science* is centered around the positive correlation between the size of the phonemic inventory of a language and the size of the population speaking that language, in the sense that small populations have fewer phonemes. The intuition behind a similar idea is not new: it was suggested by Trudgill (2004) in the specific form that languages with small populations have either very small or very large inventories, and languages with larger populations favour “medium-sized populations”. In testing Trudgill’s claim, I showed that this specific correlation does not really hold in a database of 428 languages (Pericliev 2004), but a part of this correlation, pertaining to the favouring of small inventories by small-sized populations, was later shown by Hay & Bauer (2007) to hold statistically, including in my dataset.

Atkinson replicates Hay & Bauer’s finding, and tries to fit it into a serial founder effect, by analogy to human genetics where it is currently believed that genetic and phenotypic diversity declines with distance from Africa, as predicted by a serial founder effect in which successive population bottlenecks reduce diversity, suggesting an African origin of modern humans. Analogously, Atkinson finds that the number of phonemes used in a global sample of 504 languages, derived from *WALS* (Haspelmath et al. (eds.) 2008), de-
creases with the decrease of the size of the populations speaking them and fits this into a serial founder effect model of expansion from an inferred origin in Africa. This result, he assumes, points to parallel mechanisms shaping genetic and linguistic diversity and supports an African origin of modern human languages.

It will be clear that to assess the magnitude of Atkinson’s contribution we should evaluate the extent to which the linguistic data on phonemic diversity he presents manages to confirm the origin and expansion of human language from Africa. Presenting purely linguistic grounds for an African origin of language would constitute a remarkable discovery, even if one common ingredient of important discoveries, viz. their unexpectedness (cf. Pericliev 2010: 1–12), is lacking here, as the African origin of language is actually a logical consequence of two current tenets: (i) humans originate from Africa, (ii) humans and their language evolve in a parallel manner.

Some commentators in discussion fora on the internet have noted that different (presumably better) conceptions of the notion of phonemic diversity or, alternatively, the same concept, but computed from a database different from WALS, may lead to results different from those reported by Atkinson. Here, I will try to show briefly that even if we grant the validity of Atkinson’s analysis, the fact that phonemic diversity may fit a serial founder effect model does not prove convincingly that a serial founder effect with inferred origin in Africa must really have taken place in language as a whole, allowing for such inferences to be made. I will give two basic arguments to this effect.

2. First argument

In the first place, in order to present convincing support for a theory positing some effect holding in language, one would expect, first, a number of various consequences from this theory to be derived and subsequently tested, and, secondly, a justification to be made of the effect that is posited to hold.

Regarding the first point, what we find in Atkinson’s case is only a single consequence to be drawn from the theory pertaining to the diminishing phonemic diversity: if a founder effect holds then it follows that the diversity (i.e., the phoneme inventory) in the founder population would be a subset of the diversity (i.e., phoneme inventory) of the parent population; therefore, the number of phonemes in the founder population would be smaller than the number in the parent population. One does not normally try to support a whole theory with such scanty data because further data may disconfirm the theory, as it happens all too often in scientific practice. It would not do to ignore potentially disconfirming data by the argument that the latter are not affected by the serial founder effect, whereas phonemic diversity is, because in this case argumentation should be provided to show why exactly this is the case.
Additionally, it is worth noting that Atkinson’s data on phonemic diversity alone does not yield strong support for an “out-of-Africa” origin of human language, owing to the fact that it does not seem difficult to produce heaps of potential datasets conforming to a serial founder effect, but with inferred origins other than Africa. We do not need to enter into detailed discussion with exact computations to show this. We note first that Atkinson assigns a number (computed in some way and of no interest for present purposes) as the “phonemic diversity”, e.g., to large geographical areas (Africa, Asia, Europe, North America, South America, and Oceania), and shows that the phonemic diversity is largest in Africa, smaller in Asia, and smallest in South America and Oceania, corresponding to currently assumed large human population migrations. However, in practical terms any arbitrary global distribution D of some linguistic feature F may be used to partition the globe into geographical areas (possibly unusual and not corresponding to current continents) with statistically significant variation of F. This being the case, for D there will necessarily exist some serial founder effect model, since the assigned diversity of geographical areas is a number and these numbers can be ordered in successively diminishing order (conforming to a serial founder effect). Among the huge number of such distributions fitting a serial founder effect, there will certainly be such whose inferred origin is other than Africa. Then, the question arises: Which of these potentially different results must one trust? If one chooses to prefer Atkinson’s phonemic diversity over other features, this choice should basically be made on the grounds that phonemic diversity agrees, or fits, the trusted “out-of-Africa” origin and expansion of humans, which means that, in effect, the argument should be based on what actually the article is intended to prove – a clearly circular argument.

Regarding the second requirement for presenting convincing supporting evidence for a theory positing some effect having taken place in language, one should provide a justification of this effect, showing the rationale behind our empirical findings. However, again, we do not find sufficient argumentation of this kind in Atkinson’s article, and one wonders why the number of phonemes must diminish in the founder population. If one looks at the concept in genetics, it makes sense why diversity should decrease with the strongly diminishing number of people in the founder group in comparison with the parent group, and case-studies (Iceland, Easter Island, etc.) have been presented to prove the reliability of the effect having really taken place. The linguistic example, with the diminishing size of phoneme inventories after a founder event, in contrast, does not seem very convincing. A group of people, speaking a language L, irrespectively of how small this group is, will quite probably carry the whole phonemic inventory to the new place of residence, and not only a part of this inventory; indeed, as speakers of L they must know all the phonemes of L. Of course, we can reasonably assume that some allophones will be lost in
the founder population, as some people using certain allophones would not be members of the new population, but this does not mean loss of phonemes and it is not immediately clear whether this will lead subsequently to a phoneme loss or not.

3. Second argument

In the second place, my other concern with the article is that a single body of data can in principle be explained in a great variety of different ways, some of these explanations completely distorting the essence of the actual event being explained. This happens both in everyday life and in science. Thus, on a summer business trip abroad I went to a shop to buy towels and bathroom mules, and the shopkeeper wished me a nice holiday at the seaside, whereas the actual explanation was far more prosaic, for these accessories had actually been lost from my luggage during the air flight, and I just needed them for other purposes, and of course had no intention of going to the seaside on vacation at all. Science typically offers similar traps and one is reminded in this context of a popular anecdote among physicists: an experimental physicist brings in data which the theoretical physicist readily explains and when the experimentalists finds a mistake in the data, the theoretician notes that these new data can even more easily be explained. Similarly, we should be very careful in the case under consideration. Indeed, Atkinson is careful and considers and rejects some alternative explanations (demographic factors, etc.) but of course these factors do not exhaust all reasonable alternatives. Of the huge number of possible other alternatives, some of these more plausible than others, one readily comes to mind, viz., that it is not that a serial founder effect has taken place, but rather that the data should be explained by the principle of evolution from simple to complex. This basic principle of the evolution of life forms makes very good sense in linguistics, it has its proponents in the field, and is additionally observed, e.g., in child language acquisition. If we posit this principle rather than the serial founder effect, then, in broad outline, we should get the inverse picture of that drawn by Atkinson, viz. it will turn out that human language originates in Oceania (or South America) where languages have the smallest (and therefore simplest) phoneme inventories, while Africa has the biggest (and therefore the most complex) inventories. If we add to this the fact that the oldest sites have been found (inexplicably) in South America, this makes an alternative which – although of course I am not seriously considering it – is at least some illustration of the trickiness of similar explanations.

4. Conclusion

In conclusion, Atkinson’s article, as it stands, does not seem to present convincing linguistic evidence for an “out-of-Africa” origin and expansion of hu-
On phonemic diversity and the origin of language in Africa

man language. What is needed to make the argument compelling are diverse additional data, a good illustration that a founder effect is indeed operative in language, and perhaps an exclusion of some reasonable alternative explanations for the evidence presented. This being said, I must say that I nevertheless find the article interesting and important in addressing a fundamental problem of language evolution, in applying sophisticated techniques to the problem, which are not in the customary analytical repertoire of linguists, and most of all in stimulating a debate in the linguistic community. We may at this moment be far from finding an adequate answer to this fundamental problem, but such a debate is welcome, as this is how science progresses.

Received: 31 May 2011

References


Pericliev, Vladimir. 2004. There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. Linguistic Typology 8, 376–383.
