

Commentary: There is no demonstrable effect of desiccation

Journal of Language Evolution, 2016, 65–69
doi: 10.1093/jole/lzv015

OXFORD

Harald Hammarström*

Department of Language and Cognition, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6500 AH Nijmegen, The Netherlands

*Corresponding author: harald.hammarstrom@gmail.com

Everett, Blasi, and Roberts in their position paper (Everett et al. 2016, henceforth EBRPP) argue that the sound systems of human languages are ecologically adaptive on the grounds that, (1) human and animal behaviour is ‘generally’ adaptive and (2) that their previous work ‘supports’ the idea that ambient desiccation in the area where a language is spoken leads to the absence of lexical tone (EBRPP: 1).

The first point is that it is more reasonable to expect a priori that sound systems of human languages are ecologically adaptive, than to expect the opposite. The authors boldly claim that ‘humans adapt to their ambient conditions at every observed level’ (EBRPP: 1) and follow the theoretical perspective with a number of examples of ecological adaptation from human and animal studies (EBRPP: 2–4). But somewhere here the logical jump from ‘some’ to ‘all’ was lost on the authors. If ecological adaptation can be found on some level it does not follow that sound systems of human languages belong to an adaptive level, or that all other levels are (or should be a priori expected to be) adaptive. To be fair, formulations later are weakened to ‘nearly every observed stratum’ (EBRPP: 1) and ‘nearly every inspected level’ (EBRPP: 2), but even that seems too strong. Should the water divider for a priori categorization not be whether the climatic differences can be argued to have a *discernible impact* on the human cultural behaviour in question? That is, the a priori question need not be determined by a general rule that stipulates everything to (not?) adapt to climate, but by reasoning about the potential discernible impact given the theoretical specification of climatic differences and nature of some cultural behaviour. Had the authors approached ecological adaptivity in human language with this proviso in mind, they might have been more successful in actually finding it, such as with whistled languages (Meyer 2005: 29–86) or signed hunting registers (Divale and Zipin 1977).

The case, then, boils down to the authors’ second point, i.e. the empirical validity of the idea of ecological adaptivity of human language. The authors cite a number of studies which ‘suggest’ this is the case but without even a modicum of critical scrutiny of these studies (EBRPP: 3–4). Beyond ‘suggestive’ studies, the full weight of the case is rested on the authors’ own recent paper which allegedly (EBRPP: 5) is ‘demonstrating a robust statistical association between ambient desiccation and the absence of lexical tone (Everett et al. 2015)’. It continues ‘through various strategies, from simple intra-linguistic-family and intra-regional regressions to cross-isolate comparisons to global Monte Carlo analyses, we demonstrated that the association was clear and not the result of confounds, such as language or areal relatedness between particular data points’. (EBRPP: 5). An examination below of every claim¹ made in Everett et al. 2015 (henceforth EBRT) reveals that this is outright false.

First let me declare the specifics of the data. I use the same data on humidity² and tonality in the languages of the world³ as the EBRT (1324) paper. For the

- 1 We only scrutinize the parts that actually make a statistical claim. For example, the data in Figures 1 and 2 of Everett et al. (2015: 1324–5) do not and are thus spared commentary here. All claims regarding mean annual temperature are either not significant according to Everett et al. (2015: 1324–5) or subsumed by stronger ones involving humidity, so we only address the latter.
- 2 The mean specific humidity, here called simply ‘humidity’, per language (EBRT: 1324), kindly supplied by Seán Roberts.
- 3 The number of tones, here called simply ‘tonality’, from The Phonotactics Database of the Australian National University (ANU) with 3700 entries (subsuming a smaller database with data of less granularity in the

classification of languages into families I used the classification of Glottolog 2.6 (Hammarström et al. 2015) since it is backed by explicit references for justification.⁴

EBRT (1322–3) present arguments for why less humidity would lead to less tonality in the languages of the world. There are at least two ways to frame a prediction from this. One, which we may call the *trade-off* hypothesis, is that there is a general gradient relation such that less humidity leads to proportionately less tonality. The other, which we may call the *corner* hypothesis, is that (there is not necessarily a general relation but) an effect only in (very?) dry regions and only for complex tone (see below). EBRPP rest their case on the corner hypothesis (e.g. ‘our hypothesis that complex tonality should be disfavored in arid contexts, particularly extremely arid regions’, EBRPP: 6) which is puzzling for two reasons: (1) the statistical claims of EBRT referred to in EBRPP (5) regarding isolates, genealogical, and areal control are actually framed in terms of the trade-off hypothesis in EBRT (1325–6, SI1) and (2) the theoretical motivation given in EBRT (1–2) about throat desiccation and tone production (when taken at face value) seem to motivate the trade-off hypothesis rather than the corner hypothesis. Therefore, we take the space here to refute the claims regarding both the trade-off hypothesis and the corner hypothesis.

The trade-off hypothesis is straightforwardly tested with simple regression between humidity and tonality. If we sample one language per family (slight differences are found across different random samples, and the given values are averages) to remove genealogical dependencies, and compute this correlation, we get a Pearson’s $r \approx -0.017$ nowhere near significance ($p \approx 0.639$) and a Spearman’s $\rho \approx -0.04$ also nowhere

World Atlas of Languages Structures (WALS) also used in EBRT: 1324). Many of the data points regarding tonality in the Phonotactics Database look suspicious, but since the database in question lacks explicit sources for its data, no attempt at correction or validity assessment has been (or could be) carried out. For the purposes of the present commentary, we assume these data as given.

- 4 EBRT do not specify which language classification they used but appears to be that included in the Phonotactic Database or an extension of the classification included in WALS. In any case, the classification used is very similar, and though some details differ, none of the conclusions of the present commentary relate to any differences in the language classification. Languages in the Phonotactics Database which do not have a valid iso-639-3 code could not be systematically classified (cf. the lack of explicit sources) and were thus skipped.

near significance ($p \approx 0.811$). In other words, absolutely no relation between the two (if anything, very slightly negative), and this is before we apply even stricter independence control, involving areality!

Now for the corner hypothesis, the case is more involved, so we break it down issue-by-issue:

1. What is complex tone?

The authors divide (EBRT: 1324) tonal languages into simple tonal languages (two tones) and complex tonal languages (three or more tonal contrasts). EBRPP (11) confidently state that the prediction starts from complex tonality: ‘the prediction of our hypothesis is not that humidity broadly correlates with tonality. It is simply that desiccation yields subtle diachronic pressures against the usage of complex tonality. Therefore, we would only predict an effect in areas which include very dry climates.’ But it remains unexplained why only complex tonality would be affected, rather than any gradient tonality? How does the desiccated larynx know whether the language has simple or complex tone? Similarly, EBRT (1322–3) make many references to ‘precise’ phonation as if there was a discrete step between ‘imprecise’ and ‘precise’ phonation. In fact, the arguments in EBRT (1322–3) are squarely of the kind that make a gradient prediction between desiccation and tonality—not a discrete one—except possibly for one (EBRT: 1323): ‘Languages with phonemic tone necessitate voicing at relatively precise pitches throughout an individual’s normal range. As noted in cross-linguistic surveys of fundamental frequency, the typical pitch range for most human males is about 100 Hz (34–36). The just noticeable difference between lexical tones is about 10 Hz (37), and a cross-tone pitch difference of at least 20–30 Hz is considered marginally sufficient to achieve phonemic contrast (38). These figures suggest that languages with more than three level phonemic tones present articulatory and perceptual challenges, a suggestion that is supported by work on the acquisition of tonality (39, 40).’ Suppose we accept these numbers, then the articulatory-perceptual challenge should occur in languages with ‘more than three level phonemic tones’. With the numbers given, three level tones are fully within the articulatory-perceptual window—100 Hz divided by 20–30 Hz is 3.33 – 5.00 so the challenges should start with languages beyond three level tones or even with languages beyond five level tones. But the authors make their claims with respect to complex tone defined as three tonal contrasts, which is achieved with three level tones (or even with two level tones and one contour tone), which is *not* actually the prediction that follows from the motivating passage. Thus, even if

we assume, for the sake of the argument, that the non-gradient distinction is defensible, the actual division used by the authors is the wrong one—‘complex tone’ as four (level tones) would be the smallest number that is faithful.

2. Examining the corner hypothesis

Nevertheless, (again) for the sake of the argument, let us examine the corner hypothesis using the authors’ definition of complex tone as three tones, in spite of the motivational problems just mentioned. If the corner hypothesis is false, a random sample of complex tone languages would be expected to show the same humidity distribution as a random sample of non-complex tone languages. But if the corner hypothesis is true, a random sample of complex tone languages should lack, or at least have much fewer, languages at dry, i.e. low humidity levels. Without, or with much fewer, languages in these low humidity levels, the sample of complex tone languages should show higher mean as well as median humidity compared to a corresponding sample of non-complex tone languages. Whether these differences should be expected to be significant depends on the relative numbers of languages found in dry versus non-dry regions. At the same time as the hypothesis is explicitly characterized as not being of a gradient nature, the authors of EBRT/EBRPP never tell us what counts as a dry versus non-dry region, so we have to keep this open for the time being.

Let us first note the following. The data set has sixty-six families with (at least one language with) complex tone. If we sample one language with complex tone per family (Like EBRT, we do this by first selecting all languages with complex tone, group them by family, and then sample one per family), we get a mean humidity of 0.01416 and a median humidity of 0.01432. These means and medians are not statistically higher than any (random, tone-uninformed) subset of sixty-six languages from different families: the complex tone languages have a higher mean humidity in only 502 cases and a higher median humidity in only thirty-nine cases out of 1,000 random samples of sixty-six languages from different families. The means and medians of complex tone languages are also not statistically higher than corresponding subsets of strictly non-complex tone languages from different families: The complex tone languages have a higher mean humidity in only 603 cases and a higher median humidity in only forty-eight cases out of a 1,000 random samples of sixty-six non complex tone languages from different families. The difference in means is not significant support for the corner

hypothesis and the difference in medians is not even compatible with the corner hypothesis, since it points in the opposite direction, i.e. that non-complex tone languages tend to have higher median humidity, and may even be statistically significant (I say ‘may’ rather than ‘is’ because I suspect the association will vanish under control for areality).

3. Testing the corner hypothesis

Now EBRPP (5) claim there is a ‘robust statistical association between ambient desiccation and the absence of lexical tone [*sic*—complex tone is intended]’ (EBRPP: 5). So what does EBRT then claim is the association? EBRT (1325), as in the above, sample one language with complex tone per family and then take an equal number of non-genealogically related non-complex tone languages. They then look at the 15th, 25th, 50th (median), and 75th percentiles (EBRT: 1325), i.e. with sixty-six languages with complex tone from different families, the 10th lowest, 16th lowest, 33rd lowest, and 50th lowest value, not the mean. They find (EBRT: 1325) that the complex tone languages have a higher humidity for the 10th lowest (89 per cent of the samples) and 16th lowest (88 per cent of the samples), but more often than not for the (median) 33rd lowest (43%) and 50th lowest (49%). But not even the 89 per cent and 88 per cent values are actually significant at conventional levels of significance (where they would have to be at least 95 per cent to reach 0.05 significance). Finally, the measure that is supposed to be significant is stated: ‘In Fig. 3 ... the difference distributions have location parameters outside the 95 per cent confidence interval for the null hypothesis’ (EBRT: 1325). What is meant is that the difference (not merely which is bigger) in humidity for the 10th lowest (15th percentile) and the 16th lowest (25th percentile) value between the complex tone languages and the non-complex tone languages is significant with respect to the null hypothesis. But what is the null hypothesis? That the difference in humidity for the 15th/25th percentile between complex tone and non-complex tone languages is larger than the difference between complex tone languages and random languages? That the difference is larger than that between a random set of languages and non complex tone languages? That the difference is larger than that exhibited by two sets of random languages? The interest for the corner hypothesis is the nonrandom behaviour of the complex tone languages, so we will follow this line of inquiry. I could verify that the difference in humidity between complex tone and non-complex tone languages was larger than the corresponding difference between a random set of languages

and non-complex tone languages, but only for the 15th percentile (at $P \approx 0.037$)—no other positive percentile difference attained significance at conventional levels. According to EBRPP (5), this ‘demonstrated that the association was clear and not the result of confounds such as language or areal relatedness between particular data points’. But this result was never controlled for areal influence (either among the complex tone or non-complex tone languages) nor for multiple testing, where it would perish immediately by, e.g. Bonferroni correction. When it could not be demonstrated to significance that complex tone languages were avoided in dry areas more often than non-complex tone languages, the statistical association that the authors resort to, i.e. the size humidity difference at the 15th percentile, does more harm than good to the underlying theory (which predicted rarity, not magnitude difference, of complex tone in arid regions). While it was already challenging to understand how the desiccated larynx knows that it is only supposed to suppress complex tone, it is even more difficult to see how it manages to do it in the 15th percentile specifically and to only then do it by a larger margin.

EBRT (1325) continues regarding isolates⁵ reporting that there are seven isolates with complex tone from Amazonia, Africa, and New Guinea, and 108 isolates without complex tone from all over the world. The claim is that ‘The average humidity for isolates with complex tone is 0.017, whereas the average for other isolates is 0.013. This cross group disparity is significant ($p = 0.02$, Mann–Whitney).’ (EBRT: 1325). But the Mann–Whitney test does not test whether the means are different but (the more general case) whether the populations differ as to which population more often has a larger value. There is indeed some difference among the two populations, but not with respect to their mean (or median) humidity, but probably related to their different sizes. If one samples the same number (as the number of complex tone isolates) of isolates from the total population of isolates 1,000 times, we find that the observed set of complex tone isolates do not have significantly higher mean humidity than random subsets ($p \approx 0.16$).

The intra-family tests (EBRT: 1325; EBRT: SI1) are equally unsatisfactory from a hypothesis testing perspective. Only four families are checked for internal correlations between tonality and humidity: Afro-Asiatic, Sino-Tibetan, Nilo-Saharan, and Niger-Congo, which are said to be the ‘families with complex tone that

straddle extremely diverse ecological zones’ (EBRT: 1325). These are not the largest families with complex tone, which include, e.g. Austronesian, Indo-European, and (Nuclear) Trans New Guinea. How exactly the four families ‘straddle extremely diverse ecological zones’ (EBRT: 1325) is never explained. They are not the ones with the largest span in humidity (which is what is actually being tested against), cf. Indo-European having the largest span. The selection of the four families, then, is difficult to explain as anything other than cherry-picking, since the four families, but not the others succeed in the test the authors apply (EBRT: SI1). In this case, the authors actually directly test tonality versus humidity in line with the trade-off hypothesis—as I argue, this should have been the general test—and find that internally in the four mentioned families, humidity and tonality correlate positively at conventional levels of significance (EBRT: SI1). But this test does not come out in favour of the authors’ hypothesis for the other large families, e.g. the largest family Austronesian exhibits a significant correlation in the opposite direction. But the test is deficient in any case, as it treats every language as independent, without discounting for relative genealogical distance—a simple phylogenetic least-squares test would have been an improvement (Symonds and Blomberg 2014). In addition, the areality of tone would have been especially relevant as a rival explanation for tonality in the four mentioned families (Clements and Rialland 2008; Brunelle and Kirby 2015).

Finally, for the only tests that address areality (EBRT:1325–6; EBRT: SI1), the authors test for tonality and humidity separately for the areas of North America, South America, Africa, and Eurasia given another conveniently vaguely formulated delineation (‘four major landmasses include numerous frigid and dry regions, as well as many tonal and non tonal languages’, EBRT: SI1) which excludes the entire Pacific. The authors did not test the corner hypothesis for areal confounds but instead, in line with the trade-off hypothesis, performed direct linear regressions on humidity and tonality in each area separately. They found that tonality-humidity correlated significantly at conventional levels of significance for Africa and Eurasia, but not for North America or South America (EBRT: SI1). But these intra-continent correlations were not controlled for language family, and the two continents that were significant are no longer so when one language per family is sampled.

Now, to summarize EBRPP, the present commentator argues that the theoretical motivation is too sweeping and that the empirical part referred to is problematic from start to finish. The statistical treatment of the empirical data on tonality and humidity fails to rule out the

5 Even though as far as I can tell from the description, they were (correctly) included in the family-level tests as one-member families (EBRT:1325).

classical areal and genealogical confounds as well as statistical orthodoxy. The tests applied controlling for genealogy and areality⁶ can be challenged, but, in fact, no test applied by the authors even aims to control for genealogy and areality *at the same time* which is required for anyone who wants to claim to have ‘demonstrated that the association was clear and not the result of confounds such as language or areal relatedness between particular data points.’ (EBRPP: 5). Different statistical tests are used for otherwise congruent predictions: percentile difference distributions between complex versus non-complex tone languages (in the case of genealogical control), a Mann–Whitney test (in the case of isolates), and linear regression between humidity and gradient tonality (in the case of intra-family and intra-continent tests). The corner hypothesis requires one line of testing and the trade-off hypothesis another—one cannot cherry-pick between an array of different tests and at the same time insist that the patterns are ‘robust’ (EBRPP: 5, 14). As I argue, it is the (empirically false) trade-off hypothesis that follows from the theoretical background and the ingenuity of the other strategies, along with a lack of concern for multiple testing, reflects poorly on the authors, reviewers, and editors who saw it through.

Funding

This research was made possible thanks to the financial support of the Language and Cognition Department at the Max Planck Institute for Psycholinguistics, Max-Planck Gesellschaft, and a European Research Council’s Advanced Grant (269484 “INTERACT”) awarded to Stephen C. Levinson.

Acknowledgements

The author wishes to thank Seán Roberts for sharing the data and script used for the original EBRT paper and for useful feedback on an earlier version of this commentary. The usual disclaimers apply.

References

- Brunelle, M., and Kirby, J. (2015). ‘Re-assessing Tonal Diversity and Geographical Convergence in Mainland Southeast Asia’, in: Enfield, N. J. and Comrie, B. (eds.) *Languages of Mainland Southeast Asia: The State of the Art*, 80–108. Canberra: Pacific Linguistics.
- Clements, N., and Rialland, A. (2008). ‘Africa as a Phonological Area’, in: Heine, B. and Nurse, D. (eds.) *A Linguistic Geography of Africa*, 36–87. Cambridge: CUP.
- Divale, W. T., and Zipin, C. (1977). ‘Hunting and the Development of Sign Language: A Cross-Cultural Test’, *Journal of Anthropological Research*, 33: 185–201.
- Everett, C., et al. (2015). ‘Climate, Vocal Folds, and Tonal Languages: Connecting the Physiological and Geographic Dots’. *Proceedings of the National Academy of Sciences of the United States of America*, 112: 1322–27.
- , et al. (2016). ‘Language Evolution and Climate: The Case of Desiccation and Tone’. *Journal of Language Evolution*, this issue.
- Hammarström, H., et al. (2015). ‘Glottolog 2.6’ (Jena, Max Planck Institute for the Science of Human History; pubd online 2015. <<http://glottolog.org>> accessed 6 Jan 2016.
- Meyer, J. (2005). ‘Description typologique et intelligibilité des langues sifflées, approche linguistique et bioacoustique’, Doctoral Dissertation, Lyon: Université Lumière Lyon 2.
- Symonds, M. R. E., and Blomberg, S. P. (2014). ‘A Primer on Phylogenetic Generalised Least Squares’, in: Garamszegi, L. Z. (ed.) *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, 105–30. Berlin: Springer.

6 We may also note that the results of these tests are not even reported on correctly in EBRPP (13) as the unqualified ‘In Everett et al. (2015) we did observe that the predicted patterns held within large language families and on a continent-by-continent basis’ fails to mention the counter-examples, untested cases and non-significant cases.