Chapter 10

Real and spurious correlations involving tonal languages

Jeremy Collins

Radboud University, Nijmegen

Why are some languages tonal? Is there a fundamental reason why some languages develop tone and others do not, and and does this have an effect on the way the rest of the language is organized? Tone is important in the context of dependencies, because there is no shortage of hypotheses about what can cause tone and what else tone can cause. For example, tonal languages are found predominantly in warm, humid climates, suggesting that they are culturally adaptive in those environments (Everett, Blasi & Roberts 2015); they are also found in places with low frequencies of two genes microcephalin and ASPM, suggesting that some populations are more likely to use tone than others because of their genetics (Dediu & Ladd 2007). One paper furthermore proposed that phoneme diversity declines with distance from Africa, and number of tones in particular, suggesting a founder effect of migrations, as well as a link with modern population size (Atkinson 2011). As for effects on the rest of the language, SVO word order (Yiu & Matthews 2013) and various other grammatical properties have been suggested to linked functionally with tone, and by Donegan & Stampe (1983) in particular for languages in the Austro-Asiatic family.

At least part of the reason for the large number of correlations proposed in the literature is the visibly skewed geographical distribution of tonal languages (Figure 1). They are predominantly found in Africa and Southeast Asia, immediately suggesting that tone will correlate with a large number of things, from humid climates and SVO languages, to serial verbs, and ancient settlement. The map of ASPM and microcephalin in particular, by Dediu and Ladd's own admission, was the inspiration for their correlation, when they saw that it was similar to the map of tonal languages. But by a similar reasoning, tone can be linked in a spurious way with other things found in those regions, such as acacia trees (Roberts & Winters 2013). It is therefore necessary to work out a way to distinguish which

of these correlations are real, and which of these are merely accidental consequences of these cultural traits being spatially auto-correlated.

A further major obstacle is the fact that tonal languages are typically not independent. Whole large language families can be tonal, such as Niger-Congo and Sino-Tibetan. The influence of these families in Africa and Southeast Asia has furthermore caused many languages in those regions to become tonal as well, if they were not already due to ancient relatedness with other tonal families (Enfield 2005). If one wants to demonstrate that tone correlates with anything, then one in principle has to use independent data points, which may prove impossible in practice.

I focus in this paper on the correlation proposed by Dediu & Ladd with ASPM and microcephalin, and briefly discuss the correlation proposed by Everett, Blasi & Roberts with humidity, based on my response to their paper (Collins 2016). My assessment of their causal claims will be primarily negative. The reasons I give will be that the evidence for their causal mechanisms are inadequate, and that the methods that they claim control for language family and geographical distance do not work. These points have broader relevance than just for tone, as these affect the way that correlations in general are studied typologically, a point also emphasized by recent work by Ladd, Roberts & Dediu (2015). I end the paper with some broader points illustrating the way that the problem of non-independence in linguistics can take some subtle forms, complicating the search for the genuine dependencies which exist in linguistic systems.

1 Tone and genes

Dediu & Ladd (2007) argue that two genes, ASPM and Microcephalin, may have an effect on the processing of tone. Speakers with particular alleles of these genes are found in regions where tonal languages are. The correlation between these genes and tone is strong (stronger than 97.3% of all gene-language correlations that they tested), and it remains significant in a partial Mantel test controlling for language relatedness and distance between languages. Since these two genes are expressed in the brain, the reason for this may be that these two genes have an effect on speakers' processing of tone, causing some languages to be less likely to develop tone than others, given that there are large differences between populations in the frequency of these genes.

This sounds impressive, until the argument is unpacked. There is nothing about the genes ASPM and Microcephalin which could lead someone to predict any effect on language, much less on a particular property of language such as

tone. In fact, the reason why the authors decide to focus on those two genes in particular is that these genes are found in the same region as tonal languages. Dediu says, both in their paper and anecdotally in a footnote in his doctoral thesis, that the idea for testing that particular hypothesis came from examining the maps of those two genes which had recently appeared in a few well-known papers (Evans et al. 2005; Mekel-Bobrov et al. 2005), as those two genes have been argued to have undergone recent natural selection because of their high frequency in Eurasian populations, and Robert Ladd suggested that they resemble the distribution of tonal languages (Dediu & Ladd 2007: 192). This particular hypothesis is just one among many resemblances between the distribution of genes and linguistic features that could have been noticed, this one distinguished only by the faint whiff of a plausible causal link – these two genes are involved in the brain, and tone is the type of property which could be affected more than most properties of languages by genetic differences. The authors perhaps did not literally search through all genes and all linguistic features in the hope of finding a meaningful-sounding correlation somewhere, but they might as well have done.
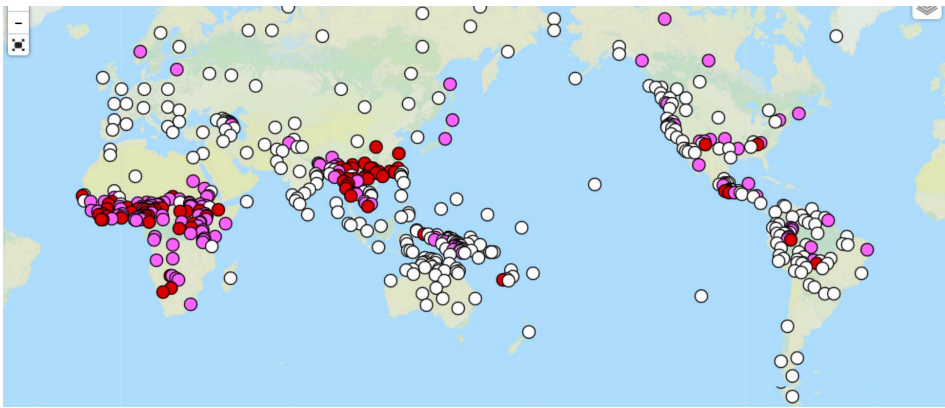

Figure 1: Map of the distribution of complex tonal languages (shown in red), simple tonal languages (pink) and non-tonal languages (white) in WALS (Maddieson 2013).

Picking two genes to focus on because they occur in the same regions as tone, itself a very spatially clustered linguistic feature, automatically makes this correlation better than most randomly selected correlations between genes and linguistic features. This makes the fact that this correlation is in the top 97.3% of gene-feature correlations unimpressive, as all this means is that Dediu and Ladd's ability to spot visual resemblance between two geographical distributions does better than 97.3% of selecting completely randomly chosen geographical distributions.

The two genes *ASPM* and *Microcephalin* had no particular reason to be tested, as their effect on cognition was unclear at the time (Dediu & Ladd 2007; Mekel-Bobrov et al. 2007). If there were an experiment that showed that people with the relevant alleles of *ASPM* or *Microcephalin* were better at tasks involving processing tone, then there would be a reason for studying it. Interestingly, since the publication of Dediu and Ladd's paper in 2007, there was a study by Wong, Chandrasekaran & Zheng (2012) that found that people with the derived allele of *ASPM* were better than those with the ancestral allele at a tone perception task. This would be an important vindication of their choice of these two genes, although a remarkable fluke given that there was nothing else that Dediu and Ladd knew about *ASPM* that could have led them to hypothesise this. However, there are two reasons why Wong, Chandrasekaran & Zheng's study cannot be taken as support for Dediu and Ladd's claim. The first is that the result of their experiment went in the opposite direction to that predicted by Dediu and Ladd; the ancestral allele is the one that is found in regions with tonal languages, not the derived allele. The second reason is that Wong, Chandrasekaran & Zheng's sample only contained thirty-two participants. This makes it quite possible that their result is a false positive.

For me, the lack of a proper justification for why they chose those genes makes much of their argument invalid, no matter how statistically well supported the correlation is, such as the comparison with other gene-feature correlations, or even the fact that it survives the controls for language family and geography. However, it is still an interesting question why the correlation is that strong, and why it continues to be after using a Mantel test.

Could correlations like that emerge because of the way that genetic variants and linguistic features cluster together due to linguistic areality? Southeast Asia in particular is one area of the world where there has been widespread bilingualism and sharing of linguistic properties such as tone across language families (Enfield 2005). To the extent that this was accompanied by gene flow between these populations, a correlation could emerge between tone and particular genetic variants beyond that predicted by language family boundaries and geographical distance. In order to answer this, I looked at mitochondrial DNA haplogroups, in order to see how often a randomly selected haplogroup would correlate with tone (either presence of tone or number of tones) in a Mantel test after controlling for language relatedness and geographical distance. Mitochondrial DNA is a good tracker of human migrations, as it is transmitted to children from their mother, and hence one can use a particular type of mitochondrial DNA (a haplotype) to trace back where ones maternal ancestors have come from. A mitochondrial

DNA haplogroup can have a historically meaningful distribution, then, which is likely in some cases to correlate with the distribution of tonal languages. There is also the intriguing possibility that if tone was carried by migration of people, such as the spread of Han people in China, that particular maternal lineages may correlate with the presence of tone and illuminate the way that it has spread.

I collected frequencies of mitochondrial DNA haplogroups from 74 populations in Africa and Eurasia, representing 26 different language families. A total of 252 mtDNA haplogrups were used, of all levels of specificity available in the literature (from tables of haplogroup frequencies rather than from the nucleotide sequence data).

Mitochondrial DNA haplogroups range in levels of specificity; for example, a person may belong to haplogroup H1, and this will mean that more generally they will belong to H, and even more generally HV, and so on, back to the haplogroup M and eventually back to L0, which all modern humans belong to. The frequencies of haplogroups were therefore calculated using a family tree of mtDNA: the frequency of H1 in a population contributes to the frequency of H, and so on upwards in the phylogeny, using the phylogeny from van Oven & Kayser (2009).

The conclusion of this analysis is that mitochondrial DNA is a good predictor of the distribution of tonal languages, and remains so after applying statistical controls for relatedness and geographical distance (although I should emphasise that these are not exactly the same as the controls that Dediu and Ladd used). Each haplogroup was tested in a partial Mantel test with number of tones that languages have, controlling for language family (here, pairs of languages are coded as 1 for being in the same family and 0 for being in different families) and geographic distance calculated using the Haversine formula (this does not take landmasses into account or geographical barriers). 26 out of 206 haplogroups correlate with number of tones in this test, meaning that there is a 12% chance that a neutral genetic marker will correlate with number of tones after apparently factoring out historical relationships between languages. This means that their result does not in fact reach conventional significance (i.e. there is a much greater than 5% chance of their result given the null hypothesis). This result is 14% if instead of number of tones languages are simply coded for presence or absence of complex tone (namely having more than two tones), as Dediu and Ladd did.

Overall difference between populations in the frequencies of haplogroups, calculated simply as Manhattan distance of these frequencies, also turns out to correlate with number of tones after controls for language relatedness and geography ($r=0.18$, $p<0.001$). The most impressive claim in their paper, therefore – that after controlling for language history their correlation holds – in fact turns out

to be explicable in terms of the way that neutral genetic markers and linguistic features cluster together, perhaps helped by migration between populations.

None of this invalidates the general hypothesis that genes can affect the structure of languages. There is probably genetic variation in linguistic abilities within populations, and when these differences can be between populations too, then one would expect that these can affect language production and what type of linguistic structures catch on in a community. The way to study these, however, is to start from a hypothesis about genes themselves, and then to test the cross-linguistic prediction. There may be variation in places such as China in ability to process tone, just as there are genetic differences in people's ability to process musical pitch; studies of speech disorders may reveal some examples. It is even possible that genes for processing pitch in language may have undergone natural selection, given the communicative importance of intonation in most languages (Cruttenden 1997) and commonalities across languages such as the use of rising intonation for questions and falling intonation for assertions (Dryer 2013).

Work currently being done by Dediu and Moisik on differences in the vocal tract between populations is one example of work on genetic influences on language which begins from a physiological mechanism and makes a cross-linguistic prediction (Dediu, this volume). This type of work may be successful, as it is quite plausible that the morphology of the vocal tract varies between populations and makes certain phonemic distinctions more likely to occur in some populations than others. But the particular case of tone, ASPM and Microcephalin is probably misguided.

2 Tone and climate

Everett, Blasi & Roberts (2015) find a correlation between humidity and complex tone, a correlation that holds up within different families and parts of the world. They suggest that dry air is known to affect the larynx and make precise phonation more difficult, precisely the kind of thing that really could (in principle) affect the way that people use a tonal language.

The number of tones that languages use correlates with humidity within five different global areas (Africa, Eurasia, South America, North America, and the Pacific), and within four different language families (Sino-Tibetan, Austro-Asiatic, Afro-Asiatic, Niger-Congo). This is better statistical support than even for word order universals, which despite having some support when sampling from different macro-areas (Dryer 1992) do not seem to hold consistently within large language families (Dunn et al. 2011).

In addition, the experimental evidence that they cite showing that dry air has an effect on the larynx raises a host of linguistic questions that are worth exploring anyway, even without this global correlation between dryness and lack of tone. Do speakers of Cantonese alter their use of tone in dryer conditions, for example? This may be a realistic expectation, if the effect of desiccated air on the larynx is as strong as it is reported in experiments. China is a natural testing ground for work of this kind, given that varieties of Chinese vary in their number of tones and in their climatic conditions.

However, as I argue in a response to their paper in the inaugural issue of the *Journal of Language Evolution* 2016, there is an important confound in the correlation between complex tone and humidity. In brief, there are a lot more languages in humid environments than in non-humid environments (the correlation is Pearson's r=0.31, p<0.001). If tone had developed purely at random in the world's languages, they would still be expected to be found in the more humid places. I show this in a series of simulations in my paper, in which tone develops in a random set of languages and then spreads by language contact to a set of neighbours. In all models, it is very likely (between 50 and 83% depending on the parameters) that tone will correlate with humidity even after using the same controls for language family and the random independent samples test that Everett, Blasi & Roberts employ.

Another finding of Everett, Blasi & Roberts's paper is that number of tones correlates within large language families, such as Sino-Tibetan (Pearson's r=0.16, p<0.01) and Niger-Congo (Pearson's r=0.3, p<0.001). However, the major confound here is once again language contact. Sino-Tibetan languages also have fewer tones when they are near to generally non-tonal Indo-European languages, and have more tones when near highly tonal Hmong-Mien languages. Niger-Congo languages similarly lose tones near non-tonal (or low-tonal) families such as Nilo-Saharan and Afro-Asiatic. This matters because speakers of non-tonal languages may be affecting the tonal systems of Niger-Congo and Sino-Tibetan languages. An example is Swahili, which has no tones despite being a Bantu language, most of which have several tones. The reason for this is probably influence from Arabic and its use by Arab traders. A phylogenetic analysis of Sino-Tibetan and Niger-Congo shows that these languages have a strong tendency to lose tones as they move towards non-tonal languages in other families, in accordance with this prediction (Collins 2016).

The causal mechanism is intriguing and worth testing in naturalistic contexts, such as in conversations in different Chinese varieties. However, language contact should be considered a serious confound in the way that it can create a posi-

tive global correlation between humidity and complex tone, including after controlling for language family, and even within families.

3 Conclusions

When trying to find dependencies between linguistic traits, one should be aware of various sources of non-independence of data points when testing a correlation. Languages are related to each other by common descent in language families, and hence there needs to be a control for relatedness. But they also influence each other by language contact, and this causes traits such as tone, which spread a long way by contact through languages such as those in Southeast Asia, to form spurious correlations with other traits found in those regions.

With regard to other dependencies discussed in this volume, there are perhaps other unexpected sources of non-independence of data points that need to be taken into account. A particularly subtle form is the historical non-independence of individual linguistic constructions. An example is the way that adpositions often derive from verbs or nouns, and hence correlate in their ordering with verb-object order or genitive-noun order (depending on their source) (Aristar 1991). Prepositions in Mandarin, for example, are thus not historically independent from verbs, as they share a common ancestor. This is a more subtle kind of Galton's problem meaning that the correlation between verb-object ordering and having prepositions in many languages cannot be deduced to be causal, if the word forms for prepositions are in fact derived from verbs. It may be taking it to an extreme to argue that word order correlations are simply due to processes such as grammaticalization, as some people have argued (Aristar 1991), and which is discussed by Cristofaro (2017, in this volume). But it is a surprisingly difficult point to convey to people, who insist that grammaticalization may just be the way that functionally motivated word order correlations can arise. They are missing the point, which is that if they wish to claim a functional motivation, they must first deal with Galton's problem – in this case, in its more subtle form of the common ancestry of constructions, which however is no less real than the common ancestry of whole languages.

This last point in particular stresses the history of individual constructions and properties of languages, for which I think the word "meme" is an appropriate cover term (Dawkins 1976). The way that individual memes can travel between languages and have interconnected ancestries of their own creates problems for inferring causal dependencies. Many of these points are echoed in this volume, as well as in recent work by Enfield such as *Natural Causes of Language* (Enfield

2014). He discusses the fact that behind the illusion of whole languages replicating and dividing into family trees, and behind the illusion of whole linguistic systems, lie instead linguistic replicators which have their own histories but nevertheless need to interact to produce functionally coherent systems. Richard Dawkins makes this point as succinctly as anyone and is arguably its originator, having argued it forcefully for the case of genes building organisms, and then in coining the word "meme" for similar processes acting in cultural evolution. To quote from his foreword to Susan Blackmore's *The Meme Machine* (Blackmore 1999):

"Every gene in a gene pool constitutes part of the environmental background against which other genes are naturally selected, so it's no wonder that natural selection favors genes that 'cooperate' in building these highly integrated and unified machines called organisms. Biologists are sharply divided between those for whom this logic is as clear as daylight, and those (even some very distinguished ones) who just do not understand it – who naively trot out the obvious cooperativeness of genes and the unitariness of organisms as though they somehow count against the 'selfish gene' view of evolution... By analogy with coadapted gene complex, memes, selected against the background of each other, 'cooperate' in mutually supportive memeplexes..."

Although Enfield does not use the word "meme" to describe these replicators, the influence of the idea on this view of language is clear. For me, the additional relevance of the "meme" idea is to think in terms of individual properties of language and the way that they can be transmitted by horizontal transfer, and themselves have interconnected ancestries, as genes can, and these need to be understood in order to where the genuine functional dependencies are.

References

Aristar, Anthony R. 1991. On diachronic sources and synchronic patterns: An investigation into the origin of linguistic universals. *Language* 67. 1–33.

Atkinson, Quentin. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332. 346–349.

Blackmore, Susan. 1999. *The meme machine*. Oxford: Oxford University Press.

Collins, Jeremy. 2016. The role of language contact in creating correlations between humidity and tone. *Journal of Language Evolution* 1(1). 46–52.

Cruttenden, Alan. 1997. *Intonation (2nd edition)*. Cambridge: Cambridge University Press.

Dawkins, Richard. 1976. *The selfish gene*. Oxford: Oxford University Press.

Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and microcephalin. *Proceedings of the National Academy of Sciences* 104(26). 10944–9.

Donegan, Patricia & David Stampe. 1983. Rhythm and the holistic organisation of language structure. In M. Marks J. F. Richardson & A. Chukerman (eds.), *Papers from the parasession on the interplay of phonology, morphology, and syntax*, 337–353. Chicago: Chicago Linguistic Society.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dryer, Matthew S. 2013. Polar questions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/116, accessed 2016-01-26.

Dunn, Michael, Simon Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word order universals. *Nature* 473. 79–82.

Enfield, N. J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.

Enfield, N. J. 2014. *Natural causes of language: Frames, biases, and cultural transmission.* Berlin: Language Science Press.

Evans, Patrick, Sandra Gilbert, Nitzan Mekel-Bobrov, Eric Vallender, Jeffrey Anderson, Leila Vaez-Azizi, Sarah Tishkoff, Richard Hudson & Bruce Lahn. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309. 1717–1720.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112(5). 1322–1327.

Ladd, D. Robert, Seán G. Roberts & Dan Dediu. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1. 221–241. DOI:10.1146/annurev-linguist-030514-124819

Maddieson, Ian. 2013. Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology 2013. http://wals.info/chapter/116, accessed 2016-01-26.

Mekel-Bobrov, Nitzan, Sandra L. Gilbert, Patrick D. Evans, Eric J. Vallender, Jeffrey R. Anderson, Richard R. Hudson, Sarah A. Tishkoff & Bruce T. Lahn. 2005.

Ongoing adaptive evolution of ASPM, a brain size determinant in homo sapiens. *Science* 309. 1720–1722.

Mekel-Bobrov, Nitzan, Danielle Posthuma, Sandra L. Gilbert, Penelope Lind, M. Florencia Gosso, Michelle Luciano, Sarah E. Harris, Timothy C. Bates, Tinca J. Polderman, Lawrence J. Whalley, Helen Fox, John M. Starr, Patrick D. Evans, Grant W. Montgomery, Croydon Fernandes, Peter Heutink, Nicholas G. Martin, Dorret I. Boomsma, Ian J. Deary, Margaret J. Wright, Eco J. de Geus & Bruce T. Lahn. 2007. The ongoing adaptive evolution of ASPM and microcephalin is not explained by increased intelligence. *Human Molecular Genetics* 16. 600–608.

Roberts, Seán G. & James Winters. 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLoS One* 8(8). DOI:10.1371/journal.pone.0070902

van Oven, Mannis & Manfred Kayser. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 30(2). E386-E394.

Wong, Patrick C. M., Bharath Chandrasekaran & Jing Zheng. 2012. The derived allele of ASPM is associated with lexical tone perception. *PLoS ONE* 7(4). e34243. DOI:10.1371/journal.pone.0034243

Yiu, Suki & Stephen Matthews. 2013. Correlations between tonality and word order type. In *The 10th biennial conference of the Association for Linguistic Typology*. Leipzig, Germany.