

The relationship between population size and rates of language evolution

CAELA WELSH

Abstract

Languages change over time, and the evolution of languages is similar in many ways to biological evolution. But are all patterns predicted by evolutionary theory also seen in language evolution? One well-known biological pattern is that smaller populations tend to lose genetic diversity, and large populations tend to become more genetically diverse, but the effect of population size on language evolution is highly debated. Do small populations have faster rates of change through greater diffusion of new words, or slower rates of change due to strict language transmission? Do large populations gain words faster through more innovators, or lose faster due to simplification across varied speaker communities? This extract, taken from Welsh's (2015) Honours thesis, provides the background on the relationship between population size and rate of language change, discusses some of the contrasting hypotheses and methodological techniques and explains the research currently being conducted that attempts to clarify the relationship.

Language is a uniquely human phenomenon and studied by experts in many different fields. New methodological techniques are being developed to answer questions about language acquisition, origin and diversity (Levinson & Gray 2012; Berwick et al. 2013; Gong et al. 2013; Gong et al. 2014). One promising approach is the application of biological evolutionary theory and computational phylogenetic techniques to understand the patterns and processes of linguistic change. These tools have been used to investigate patterns in language evolution including the effect of population size (Nettle 1999b; Wichmann et al. 2008; Wichmann & Holman 2009; Kline & Boyd 2010). Although this effect has been widely studied in biological systems, the patterns of language change have been subject to heated debate and few hypotheses have been tested formally using empirical analyses.

The application of biological techniques to study language change is possible due to many similarities between biological and language evolution. For example, just as DNA variation naturally exists within a population of biological organisms, linguistic variation naturally exists within a speaker population, and this variation can be inherited by succeeding generations through learning and imitation (Lewontin 1970; Mesoudi et al. 2004). For both biological organisms and languages, a process of descent with modification creates a hierarchy of similarity between related lineages over time, which can be represented as a phylogenetic tree (Atkinson & Gray 2005). However, key differences exist between these two processes. Firstly, while DNA mutations occur stochastically, a linguistic change may occur intentionally. For example, two groups may strive to become linguistically distinct (a process dubbed 'schismogenesis') and change their language intentionally (Bateson 1935). Also, DNA is transmitted from parents to offspring, while language is transmitted through learning from all members of society (Cavalli-Sforza & Feldman 1981). These differences may limit the extent to which conclusions about process can be drawn from patterns obtained using the biological evolutionary method.

The effect of population size on rates of biological evolution has been widely studied and is well understood (Charlesworth 2009). Population genetic theory predicts that selection is more efficient in populations with a large effective population size because there are more individuals to generate mutations and relatively less impacts of genetic drift. This results in a faster substitution rate of advantageous mutations. In contrast, small populations should have a faster rate of gain of slightly deleterious mutations, and a higher rate of loss of advantageous mutations due to drift (Ohta 1972; Gillespie 1999; Lanfear et al. 2014). The predictions are not as clear in linguistic evolution and there are many competing theories for the mechanisms of interaction between population size and rates of linguistic change.

Firstly, just as larger populations of organisms have the potential to produce more genetic mutations, large populations of languages could gain words faster by producing more linguistic innovations (Henrich 2004; Kline & Boyd 2010; Collard et al. 2013; Baldini 2015). However, large populations may be less efficient at diffusing these innovations through the population and smaller populations may uptake these new innovations and gain words faster (Nettle 1999a). Small populations may also have a lower rate of word loss if they are resistant to change with improved linguistic transmission across generations (Trudgill 2004). In contrast, random sampling effects are stronger in small populations and could result in a higher rate of word loss through drift (Kline & Boyd 2010; Real & Griffiths 2010) or founder effects. If a small population of founders colonise a new area but do not speak the full range of linguistic variants present in the parental language, rates of word loss will be initially accelerated through this linguistic founder effect (Atkinson et al. 2008; Atkinson 2011). Alternatively, it has been proposed that languages with a large number of adult

second-language learners lose words faster through simplification and removal of complex linguistic elements (Lupyan & Dale 2010). Finally, it has been suggested that population size may have no effect on rates of language change and variation in rates is due to factors independent of population size (Bentley et al. 2007, Wichmann et al. 2008; Wichmann & Holman 2009) or the rate of language turnover is essentially identical in all languages (Swadesh 1955).

While there have been many studies investigating the effect of population size on rates of language evolution, most have been limited by a small sample size or computer simulations making it difficult to generalise the conclusions (Nettle 1999a; Wichmann et al. 2008). Many studies also fail to consider the statistical non-independence of languages; the pattern of linguistic descent from shared ancestors' results in increased similarity between closely related languages. Consequently, languages are not independent data points. This problem has long been recognised in cultural and linguistic studies as 'Galton's problem' (Tylor 1889; Mace & Pagel 1994; Bentz & Winter 2013; Roberts & Winters 2013; Ladd et al. 2015), but linguists have few methods to effectively overcome this issue. One method to circumvent statistical non-independence in comparative biology is to use sister-pair comparisons, which examines the differences between pairs of closest relatives (sisters) and compares this to other more distantly related lineages (outgroup). Differences between sisters must have accumulated since they diverged from a common ancestor and therefore are independent of descent (Felsenstein 1985; Slowinski & Guyer 1993; Vamossi 2007; Welch & Waxman 2008; Bromham 2009; Lanfear et al. 2010; Bromham 2011). This method has been used to investigate the role of different factors on rates of molecular evolution (Bromham 2009) but only recently been applied to study rates of language evolution.

There are many ways of quantifying language change as each language element (such as phonemes, vocabulary and grammatical structure) displays variation within a population and can evolve over time (Atkinson 2011; Dunn et al. 2011). In recent years, large databases of vocabulary such as the Indo-European Lexical database (ielex.mpi.nl/) and Austronesian Basic Vocabulary Database (Greenhill et al. 2008) have been made publicly accessible online. This allows researchers to investigate changes in basic vocabulary, common concepts found across languages such as walk, sun or father (Swadesh 1955). Similar word forms for a given concept are called 'cognates' and are phonetically and morphologically similar through shared descent (Nichols 1997). For example, the concept of 'water' is represented by different words in different Indo-European languages, such as *water* (English), *wasser* (German) and *vatten* (Swedish). These words have descended from a common ancestor, proto-Germanic **watōr* (ielex.mpi.nl/), so are classified as cognates. By comparing the differences in basic vocabulary between sister-pairs of languages and other members of the language family, we can identify whether words have been lost or gained since the languages diverged from a common ancestor. When this measure is combined with a date of divergence between

sister-pairs of languages, the number of differences can be converted into a rate (Hua et al. 2015), allowing rates of word loss and gain to be compared between language pairs.

Recent work (Bromham et al. 2015) adapted the common molecular evolution comparative method of sister-pair comparisons (Lanfear et al. 2010) to study the relationship between rates of word loss and gain and the language population size. Bromham et al. (2015) applied this technique to 20 Polynesian languages and found a correlation between speaker population size and rates of linguistic change. Specifically, their results conformed to the expectations of population genetics, as large populations gained words faster while small populations lost words faster. However, Bromham et al. (2015) only investigated a restricted dataset of 20 Polynesian languages, a small subgroup of the Austronesian language family (Hammarström et al. 2014). They argued that Polynesian languages were an ideal study group as they are well documented, and spoken on isolated islands that were sequentially founded as humans expanded across the Pacific Ocean (Gray et al. 2009). Their study was also limited due to its small sample size, so it is unclear if the conclusions of Bromham et al. (2015) are specific to a feature of Polynesian cultures and languages, or can be generalised to all language families.

In Welsh et al. (in press), the statistical robust comparative linguistic method developed by Bromham et al. (2015) is applied to three major language families—Austronesian, Indo-European and Bantu—to test the generality of the correlation between population size on patterns and rates of language evolution. Together, these families span four continents, have population sizes ranging from 100 million (Nauna, Austronesia) to 258 million (Hindi, Indo-European) and display different social structures and geographical distributions. Welsh et al. (in press) will outline the methods and analysis, key results and implications of their findings for the fields of evolution and linguistics. This study will inform the discussion of the underlying mechanisms of language evolution, which in turn contributes to the study of language history and diversification.

This extract was taken from the thesis 'Population size and rates of language change: a phylogenetic test', submitted in 2015 as partial fulfilment of the requirements for the Bachelor of Philosophy (Honours) in Science.

Bibliography

- Atkinson, Q.D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, 332, 346–349.
- Atkinson, Q.D. & Gray, R.D. 2005. Curious parallels and curious connections – phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, 54, 513–526.
- Atkinson, Q.D., Meade, A., Venditti, C., Greenhill, S.J. & Pagel, M. 2008. Languages evolve in punctuational bursts. *Science*, 319, 588.
- Baldini, R. 2015. Revisiting the effect of population size on cumulative cultural evolution. *Journal of Cognition and Culture*, 15, 320–336.
- Bateson, G. 1935. Culture, contact and schismogenesis. *Man*, 35, 178–183.
- Bentley, R.A., Lipo, C.P., Herzog, H.A. & Hahn, M.W. 2007. Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior*, 28, 151–158.
- Bentz, C. & Winter, B. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3, 1–27.
- Berwick, R.C., Friederici, A.D., Chomsky, N. & Bolhuis, J.J. 2013. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, 17, 89–98.
- Bromham, L. 2009. Why do species vary in their rate of molecular evolution? *Biology Letters*, 5, 401–404.
- Bromham, L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 366, 2503–2513.
- Bromham, L., Hua, X., Fitzpatrick, T.G. & Greenhill, S.J. 2015. Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences, USA*, 112, 2097–2102.
- Cavalli-Sforza, L.L. & Feldman, M.W. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, USA, Princeton University Press.
- Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature*, 461, 195–205.
- Collard, M., Ruttle, A., Buchanan, B. & O'Brien, M.J. 2013. Population size and cultural evolution in nonindustrial food-producing societies. *Plos One*, 8, e72628.

- Dunn, M., Greenhill, S.J., Levinson, S. C. & Gray, R. D. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473, 79–82.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist*, 125, 1–15.
- Gillespie, J.H. 1999. The role of population size in molecular evolution. *Theoretical Population Biology*, 55, 145–156.
- Gong, T., Shuai, L. & Comrie, B. 2014. Evolutionary linguistics: Theory of language in an interdisciplinary space. *Language Sciences*, 41, 243–253.
- Gong, T., Shuai, L. & Wu, Y.C. 2013. Multidisciplinary approaches in evolutionary linguistics. *Language Sciences*, 37, 1–13.
- Gray, R.D., Drummond, A.J. & Greenhill, S.J. 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science*, 323, 479–483.
- Greenhill, S.J., Blust, R. & Gray, R.D. 2008. The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary Bioinformatics*, 4, 271–283.
- Hammarström, H., Forkel, R., Haspelmath, M. & Nordhoff, S. 2014. *Glottolog 2.3*. [Online]. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available: www.glottolog.org, accessed 18 March 2015.
- Henrich, J. 2004. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *American Antiquity*, 69, 197–214.
- Hua, X., Cowman, P., Warren, D. & Bromham, L. 2015. Longevity is linked to mitochondrial mutation rates in rockfish: a test using Poisson regression. *Molecular biology and evolution*, 32, 2633–2645.
- Kline, M.A. & Boyd, R. 2010. Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B-Biological Sciences*, 277, 2559–2564.
- Ladd, R., Roberts, S.G. & Dediu, D. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, 1, 221–241.
- Lanfear, R., Kokko, H. & Eyre-Walker, A. 2014. Population size and the rate of evolution. *Trends in Ecology & Evolution*, 29, 33–41.
- Lanfear, R., Welch, J.J. & Bromham, L. 2010. Watching the clock: Studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, 25, 495–503.

- Levinson, S.C. & Gray, R.D. 2012. Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences*, 16, 167–173.
- Lewontin, R.C. 1970. The units of selection. *Annual Review of Ecology and Systematics*, 1, 1–18.
- Lupyan, G. & Dale, R. 2010. Language structure is partly determined by social structure. *Plos One*, 5, (1):e8559.
- Mace, R. & Pagel, M. 1994. The comparative method in anthropology. *Current Anthropology*, 35, 549–564.
- Mesoudi, A. 2015. Cultural evolution: a review of theory, findings and controversies. *Evolutionary Biology*, DOI: 10.1007/s11692-015-9320-0.
- Mesoudi, A., Whiten, A. & Laland, K.N. 2004. Perspective: is human cultural evolution Darwinian? Evidence reviewed from the perspective of *The Origin of Species*. *Evolution*, 58, 1–11.
- Nettle, D. 1999a. Is the rate of linguistic change constant? *Lingua*, 108, 119–136.
- Nettle, D. 1999b. Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proceedings of the National Academy of Sciences, USA*, 96, 3325–3329.
- Nichols, J. 1997. Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology*, 26, 359–384.
- Ohta, T. 1972. Population size and rate of evolution. *Journal of Molecular Evolution*, 1, 305–314.
- Reali, F. & Griffiths, T.L. 2010. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society B-Biological Sciences*, 277, 429–436.
- Roberts, S. & Winters, J. 2013. Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *Plos One*, 8, e70902.
- Slowinski, J.B. & Guyer, C. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. *The American Naturalist*, 142, 1019–1024.
- Swadesh, M. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21, 121–137.
- Trudgill, P. 2004. Linguistic and social typology: The Austronesian migration and phoneme inventories. *Linguistic Typology*, 8, 305–320.

- Tylor, E. 1889. On a method of investigating the development of institutions; applied to laws and marriage and descent. *Journal of Royal Anthropological Institute*, 18, 245–272.
- Vamosi, S.M. 2007. Endless tests: guidelines for analysing non-nested sister-group comparisons. *Evolutionary Ecology Research*, 9, 717.
- Welch, J.J. & Waxman, D. 2008. Calculating independent contrasts for the comparative study of substitution rates. *Journal of Theoretical Biology*, 251, 667–678.
- Welsh, C. 2015. Population size and rates of language change: a phylogenetic test. Honours thesis. The Australian National University.
- Welsh, C.F., Greenhill, S.J., Hua, X., Schneemann, H., Bromham, L. (in press). Population Size Influences the Rate of Language Evolution: A General Test across Indo-European, Austronesian and Bantu Languages. *PLoS Biology*.
- Wichmann, S. & Holman, E.W. 2009. Population size and rates of language change. *Human Biology*, 81, 259–274.
- Wichmann, S., Stauffer, D., Schulze, C. & Holman, E.W. 2008. Do language change rates depend on population size? *Advances in Complex Systems*, 11, 357–369.

This text is taken from *The ANU Undergraduate Research Journal*,
Volume Seven, 2015, edited by Daniel McKay, published 2016 by ANU eView,
The Australian National University, Canberra, Australia.