

Linguistic change and biological evolution

Unni Leino,¹ Kaj Syrjänen,¹ Outi Vesakoski²

¹ Tampere University, Faculty of Information Technology and Communication Sciences

² University of Turku, Department of Biology

Abstract

Early conceptualisations of evolution, e.g. by Charles Darwin, saw similarities in the evolution of languages and of biological species. This similarity has been disputed at times, but in recent decades linguists have again started looking at how biology studies evolution.

Some linguistic processes appear similar to processes in biological evolution. In particular, the variation in linguistic and biological phenomena can be coded in similar ways, and to some extent the same research tools can be applied to both. This is not a new approach – phylogenetic trees of language families have been around for a century and a half, as have isogloss-based studies of dialectal variation – but in the past, systematic co-operation between the two fields has been relatively rare.

Discussion on the parallels between linguistic change and biological evolution has suffered from simplifications, as for most of the past century historical linguists have lacked a state-of-the-art understanding of evolutionary biology and vice versa. It is, however, possible to go beyond simple analogies between linguistic and biological evolution. One possible path is to continue towards meta-theories that incorporate biological and cultural evolution into a single theoretical framework. Here, one attempts to see fundamental similarities at a structural level instead of individual, simplified analogies in individual processes.

On the other hand, it is not strictly necessary to have a full-fledged epistemological understanding of the connections between biological and linguistic evolution, especially as few people have the required understanding of both fields. A large fraction of present computational methods are based simply in finding spatio-temporal patterns in linguistic variation and, in some cases, similarities in the behaviour of linguistic and genetic or cultural variation and change. To use these tools, one needs to understand both the tool and the field it is applied to, regardless of whether a more fundamental connection exists between biological and linguistic change.

Background

The similarity between the differentiation of languages and biological species is not a new idea but rather something that was observed already at the initial stages of evolutionary biology:

The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same. [...] We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation

(Darwin 1871, I:59–60)

This notion of similarity can also be seen in historical terminology of language research, as the field of diachronic linguistics was called ‘evolutionary linguistics’ in the nineteenth century; diachronic linguistics became popularized later alongside Saussure’s dichotomy between synchrony and diachrony (Aronoff 2017). Indeed, early historical linguists considered this similarity as yet another aspect of how historical linguistics should be considered a science and use methods that are scientifically rigorous:

Dass bei den Sprachforschern die naturwissenschaftliche Methode mehr und mehr Eingang finde, ist ebenfalls einer meiner lebhaftesten Wünsche. Vielleicht vermögen die folgenden Zeilen einen oder den andern angehenden Sprachforscher dazu in Betreff der Methode bei tüchtigen Botanikern und Zoologen in die Schule zu gehen.

(Schleicher 1863, 5–6)

It should also be noted that the phylogenetic tree itself was not invented by Darwin, even though he was instrumental in spreading the concept into biology. Similar trees were used to represent other types of historical lineages in linguistics, in the form of family trees, as well as manuscript lineage trees even before the publication of *The Origin of Species* (Atkinson and Gray 2005; O’Hara 1996). Even earlier, Leibniz (1710) had used linguistic and onomastic evidence in his study on the history and origins of different peoples. Of course, this was not strictly speaking an analogy between linguistic and biological evolution, as the evolution of species had not yet been proposed. Still, on the level of human populations Leibniz noted how the origins and divergence of different peoples was coupled to that of the languages, and in this his work can be seen as a precursor to modern evolutionary theories (e.g. Cavalli-Sforza 2000).

In the early 20th century, the importance of historical linguistics decreased markedly while the emphasis shifted towards structuralist approaches, focussing on describing present-day languages. As a consequence, work on language change as a type of evolution was mostly discontinued, and similarities between language change and evolution generally downplayed. In the shift from ‘evolutionary’ to ‘diachronic’ linguistics – as in other aspects of the early development of general linguistics – the influence of Saussure (1916) appears to have been notable. The rise of generative grammar further shifted focus away from linguistic evolution, and in fact Chomsky (2017) states clearly his view that languages ‘do not evolve, in the technical sense of the term’. In his view, a full-fledged concept of linguistic evolution would appear to be incompatible with a universal grammar and language capacity as an autonomous system in human cognition. On the other hand, the ‘evolution’ that historical linguistics studies is different from what Chomsky argues against – historical linguists are not talking about the evolution of the cognitive mechanisms involved in human language but rather in the evolution-like changes and divergence seen in languages, in a similar way as evolution-like change is observed in the history of other cultural objects.

In the last decades of the century, sociolinguistics started to look at the mechanics of

language change and its relationship with variation. The way this relationship is seen (e.g. Labov 1994) is not all that different from the way biologists see the relationship between variation within a population and the divergence of populations into species. Both are also open towards computational methods and a data-driven approach to research.

Since the late 20th century, evolutionary approaches to the study of language have become more visible. While the concept of linguistic evolution is not new and the recent developments are best seen as a matter of a renaissance, the trend raises not only methodological but also deeper theoretical questions. General meta-theories have been proposed, most notably by Dawkins (1976 onwards) and Hull (1988) to cover evolutionary change in biology and the domain of human culture. These are by their very nature very general, and while they appear to be in line with what is known of language change the mapping of the concepts used in the general meta-theory, biology, and linguistics is not always simple.

From the point of view of linguists trying to make sense of the development and divergence of languages, the key question is whether traditional, biology-based views of evolution are sufficiently compatible with the phenomena in linguistic 'evolution', and thus whether these linguistic processes can be called evolution in a meaningful fashion and studied from this perspective. In order to do so one must be confident of two things: first, that the tools used in evolutionary biology are adaptable for research of linguistic variation and change, and second, that the underlying concepts have sufficient similarities that this kind of approach is honest scholarship. To do the first, one needs a team with expertise in computational methods and linguistics; to do the second, a team with expertise in linguistics and biology.

Computational approaches to language change

As computers were developed, their applicability to linguistic analysis was seen almost immediately. As Weaver (1949) proposed, the principal goal was to develop tools for automatic translation, and among the methods suggested were linguistic and semantic analysis, utilising both language universals and contextual information. Later on, this developed into the more full-fledged modelling and analysis of language structure that became the primary meaning for computational linguistics.

In studying language variation, applying computational methods to dialectology led to the development of dialectometry in the 1970's (Goebel 1982), even as traditional dialectology itself was giving way to sociolinguistics. The starting point of dialectometry was the realisation that a dialect atlas was the compiler's interpretation of a massive amount of data, and that this data could be analysed more thoroughly. More recently, research has started to expand, for instance to generalising the relationship of dialectal and geographic distance over languages (Wieling and Nerbonne 2011) or looking at explanatory factors other than simple geographic distance (Honkola et al. 2018).

Meanwhile, sociolinguistics has moved the focus of studying linguistic variation from

geographically-defined dialects to include such variational axes as social class or gender, and further to seeing contextual and stylistic variation as different aspects of the same phenomenon as the variation seen between dialects (e.g. Eckert 2012). Some sociolinguistic corpora (such as Helpuhe 2014) are useful for studying ongoing language change, partly using computational data analysis similar to that used in population biology (e.g. Kuparinen et al. 2019).

Around the end of 20th century, computational linguistics also started to get interested in language divergence in ways that resemble earlier research in language evolution. A major feature in the current renaissance is the use of computational methods (e.g. McMahan and McMahan 2006). While some of the earlier quantitative linguistic methods, such as tree-building tools, bore striking resemblance with their contemporary biological counterparts (see e.g. Embleton 1986; Felsenstein 2004; Atkinson and Gray 2005), the tools in biology became much more popular than their linguistic counterparts and developed more rapidly. In this endeavour arguably linguistics is catching up to decades of theoretical and methodological advances within biology.

As was apparent already in the mid-19th century, the phylogenetic tree is a conceptual abstraction that appears to work in cases where we are summarizing historical processes characterized by branching events, which include, among others, both linguistics and biology. Such trees have been in use continuously since their introduction, and they have proved useful in not only these fields but also in a wide range of other contexts. Nevertheless, the question remains whether a tree showing the relationships between biological species is similar enough to one showing relationships between languages that essentially the same tools can be used for studying both.

It's important to keep in mind that what we are looking at is not just one single, clear-cut claim that the 'evolution' apparent in the divergence of languages is similar to that apparent in the divergence of species. The similarity can be seen from different perspectives and its existence is not a yes/no question. At one end, one can go deep into meta-theoretical studies of evolution (Dawkins 1976; Hull 1988); at the other end, one can take a more pragmatic approach and see whether methods used in one field can be adapted to work in the other.

While the pragmatic approach sounds naive it is supported by cognitive theories of how this kind of analogy is at the centre of human abstract thought (Fauconnier and Turner 2003; Lakoff and Johnson 1980). Analogies between two different domains are never perfect or complete, but they can still be useful, and in the case of applying an evolutionary view to language divergence, there is already a long history of some of the tools being used in both biology and linguistics. Because of this, it doesn't seem too far-fetched to look into the rest of the toolbox.

Ontological versus methodological similarity

When one talks about an evolutionary view of language one can mean a wide range of

different things. For instance Croft (2006) lists three main approaches. First, one can take an approach similar to evolutionary psychology that attempts to study the human capacity for language. However interesting such research may be, Chomsky (2017) is correct in considering this literal approach to language evolution to be outside the scope of linguistics.

Second, one can look at the correlations and interactions between the diversification of language, culture, and human populations. In this approach biological evolution and language change are seen as partially intertwined historical processes that also resemble one another in how they operate. This approach has been actively pursued by scholars in one form or another since Leibniz (1710), and as data sets are becoming available in such fields as linguistics, genetics, and archaeology, research along these lines is increasing in popularity (see Pakendorf 2014 for review of studies focusing on parallel diversification of genetic and linguistic populations).

Third, language change itself can be studied as a process that has similarities with biological evolution – and as a consequence, research methods and tools can be adapted from one of these fields to the other. Croft (Croft 2006) points out that using simple analogies is limited and limiting, and that a genuinely evolutionary theory of language change is only possible if it is derived from a systematic evolutionary framework or meta-theory. His choice is to turn towards Hull (Hull 1988), and this choice appears sound.

Croft ends up requiring a generalised theory of evolution, of which both the biological and the linguistic are specific cases. His reasoning (Croft 2008, 220) is that such a framework is needed to determine which parallels can be legitimately drawn and what is needed for the analogy to make sense. We feel that his position is at the same time too lax and too strict.

On the one hand, a generalised theory will by necessity be quite general. If one wants to use a method developed for biological evolution to analyse language change, it is usually not enough to just point out that both these changes are evolutionary in the sense Croft or Hull use the term, but rather one needs to go much deeper into analysing how the assumptions of the method relate to the realities of the phenomenon it is used to study.

On the other hand, using a method first developed for biology for linguistic analysis does not require a deeper general relationship between the two fields. It is sufficient that the concepts used in the computational tool can be properly mapped to the concepts in the linguistic theory it is used to support; that the tool was originally developed for something else are at this point irrelevant.

Biological and linguistic evolution are similar enough that tools developed for one can be adapted for the other. This does not mean that the two phenomena are similar in all respects – for instance, the replication of genes behaves rather differently from that of linguistic features. All in all, as Andersen (2006) shows, the mechanics of language change are very different from those of biological evolution. However, it is not necessary that the underlying mechanisms are completely analogous; in order to use a computational method, it is enough that the assumptions of the specific method that is being applied hold in the new target field.

As an example, Syrjänen et al. (2016) studied variation within Finnish using the population genetic clustering tool STRUCTURE to analyse how dialectal variation could be clustered into dialects. STRUCTURE – similarly to many of the population genetic computational methods – assumes that the study object could follow the mathematical theorem of Hardy-Weinberg equilibrium, which states that allele frequencies of the sample remain constant over generations and no geographical variation exists. This assumption is a null model assuming no mutation, no natural selection and infinite population size (Pritchard, Stephens, and Donnelly 2000).

Adopting such a model to dialect data requires a clear understanding of what the Hardy-Weinberg equilibrium means in terms of linguistic variation. In terms of population biology, it describes an ideal state where the allele frequencies do not change over generations; however, it is not immediately obvious how this translates into linguistic variation. Moreover, this assumption does not hold for typical biological populations but the method still yields results that are useful within limitations. As it turns out, this is what happens with linguistic data as well: the lack of Hardy-Weinberg equilibrium does not mean the method cannot be used, although it reduces the strength of the results. However, this level of detail cannot be derived from a reasonable generalised theory, but instead one needs to look rather closely into the specific fields and study the behaviour of the data under the restrictions of the model.

Not all computational models adoptable from biology to linguistics have deep-level evolutionary assumptions built within them – many of them aim to model a phenomenon as simply as possible, using reasonably simple mathematical assumptions. An example of a fairly simple method that does not rely on strong evolutionary underpinnings is the TIGER algorithm, a metric originally designed for fine-tuning phylogenetic analyses (Cummins and McInerney 2011), somewhat similarly to e.g. how phylogenetic models can assume either a constant rate of change for the data, or a relaxed rate of change (see Maurits et al. 2019). The TIGER algorithm itself is free of deep-level evolutionary assumptions, as it is based on measuring how consistently the characters represent the data as subsets. In linguistics TIGER has been tested for studying cognate character data, which is typically used with phylogenetic tools to study the evolutionary patterns of language families. Syrjänen et al. (submitted ms.) analysed basic vocabulary cognate data with the TIGER algorithm. Intriguingly, unlike in biological studies which use TIGER rates to optimize tree-building, they focused on exploring the nature of the metric itself and noted that the TIGER rates can also be used as a measure of historical heterogeneity, or treelikeness, for a data set – i.e. which linguistic features in the data carry more treelike patterns and which less treelike. This shows yet another interesting factor regarding the adoption of evolutionary methods: they do not need to be applied in exactly the same way as they are used in biology, but rather adapted in a way that makes sense in the linguistic domain.

The application of evolutionary approaches may become challenging in situations where a conflict exists between how cultural history and biological evolution operate. Horizontal transmission – i.e. the transmission of material between generations in a non-treelike pattern

– is one such case where incompatibility is often seen between biology and culture, and horizontal transmission is indeed extensive in cultural evolution, including linguistic evolution (e.g. Greenhill, Currie, and Gray 2009; Wichmann et al. 2011). This not only undermines the results of tree-based analysis techniques, such as Bayesian phylogenetic inference, but also calls for methods and even frameworks that would be better suited for describing and studying phenomena such as language and culture. Identifying a phenomenon is the first step in ensuring it is taken into account, and the aforementioned TIGER rate, for instance, provides a new tool for separating vertically non-treelike and treelike transmission of linguistic material from one another.

In general, discussion in the linguistic domain about the evolutionary parallels between biology and linguistics has tended to lack state-of-the-art knowledge of evolutionary biology. For example, also within the biological domain the tree model of evolution is regarded as insufficient: modern genetic analyses reveal that lateral transfer of genetic material between species through hybridization and introgression is in fact more common than traditionally assumed. Extensive horizontal transfer is present not only in viruses, bacteria and fungi (Boto 2009), but also in eukaryotes (e.g. Neelapu et al. 2019; Taylor and Larson 2019). Hybridization of species applies also to humans as ancient DNA analyses have revealed that the evolution of *Homo sapiens* includes admixture with Neanderthals and Denisovans (Sankararaman et al. 2016). One current challenge in evolutionary biology is indeed to build techniques for addressing horizontal transfer or material alongside vertical evolution (Liu et al. 2019).

With this in mind, method development in problematic cases such as this, rather than a complete abandonment of an entire range of evolutionary tools, is called for also in linguistics. Rather than abandoning the endeavour altogether because of this, a more productive approach is to develop the techniques to address these issues. For the case of vertical vs. horizontal transfer of linguistic material, various solutions have been proposed to address horizontal transmission, including the use of metrics of treelikeness (Gray, Bryant, and Greenhill 2010; Wichmann et al. 2011; Syrjänen et al. submitted ms.), techniques for visualizing non-treelike patterns, (see e.g. Bryant and Moulton 2003; McMahon and McMahon 2006; Nelson-Sathi et al. 2011), as well as techniques for exploring non-treelike patterns with the help of multiple tree-based analyses (Verkerk 2019).

Conclusions

The parallels between biological evolution and language diversification have been known since the mid-19th century. Over time, these parallels have been used to develop both fields, and they have also been left unused for long stretches of time. In recent years, their importance has again become more pronounced. This is partly due to improvement in computational capacity, the availability of large data sets in digital humanities, and general renewed interest in interdisciplinary research.

General evolutionary meta-theories give support for finding and using similarities between biology and linguistics. Such general theories are, however, too general for the task of adapting specific tools from one field to another, so that such methodological work requires analysing both fields to see whether the specific case has requirements that cannot be analogically projected from the originating field to the other one. Here, it is useful to look not only into evolutionary theories but also to other, less clearly evolutionary sub-fields where similar parallels exist. Between biology and linguistics, the network of functional similarities extends from evolutionary biology and historical linguistics to population biology and dialectology / sociolinguistics.

Combining linguistic and biological evolution does not necessarily mean forcing the entirety of linguistic change into an evolutionary frame, but it can mean simply exchanging and adapting computational approaches and modelling solutions from natural sciences to humanities, building interdisciplinary co-operation between linguists and computational scientists. Seen this way, the goal is to build computational methods to specifically study linguistic change from the ever-growing selection of data sets for historical linguistics, dialectology, corpus linguistics, sociolinguistics and so on.

As an overall summary, there is clearly enough common ground between biology and linguistics that analytical tools and to some extent even research methodologies can be adapted from one field to the other. However, whether or not this common ground is due to the fundamental similarity of linguistic and biological evolution is irrelevant: in practice, the real requirement is to understand the method and its relationship to the target field. Even if the fundamental similarity exists one needs a thorough methodological understanding in the context of the field it is applied to, and with the support of such understanding one can confidently use the method even in the absence of a more fundamental similarity between the two fields.

References

- Andersen, Henning. 2006. 'Synchrony, Diachrony, and Evolution'. In *Competing Models of Linguistic Change: Evolution and Beyond*, edited by Ole Nedergaard Thomsen, 59–90. *Amsterdam Studies in the Theory and History of Linguistic Science. Series IV, Current Issues in Linguistic Theory* 279. John Benjamins.
- Aronoff, Mark. 2017. 'Darwinism Tested By The Science Of Language'. In *On Looking into Words (and Beyond)*, edited by Claire Bowern, Laurence Horn, and Raffaella Zanuttini, 443–56. Berlin: Language Science Press.
- Atkinson, Quentin D., and Russell D. Gray. 2005. 'Curious Parallels and Curious Connections – Phylogenetic Thinking in Biology and Historical Linguistics'. *Systematic Biology* 54 (4): 513–526. <https://doi.org/10.1080/10635150590950317>.
- Boto, Luis. 2009. 'Horizontal Gene Transfer in Evolution: Facts and Challenges'. *Proceedings of the Royal Society B: Biological Sciences* 277: 819–827. <https://doi.org/10.1098/rspb.2009.1679>.

- Bryant, David, and Vincent Moulton. 2003. 'Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks'. *Molecular Biology and Evolution* 21 (2): 255–65. <https://doi.org/10.1093/molbev/msh018>.
- Cavalli-Sforza, Luigi Luca. 2000. *Genes, Peoples, and Languages*. New York: North Point Press.
- Chomsky, Noam. 2017. 'The Language Capacity: Architecture and Evolution'. *Psychonomic Bulletin & Review* 24: 200–203. <https://doi.org/10.3758/s13423-016-1078-6>.
- Croft, William. 2006. 'The Relevance of an Evolutionary Model to Historical Linguistics'. In *Competing Models of Linguistic Change: Evolution and Beyond*, edited by Ole Nedergaard Thomsen. *Amsterdam Studies in the Theory and History of Linguistic Science. Series IV – Current Issues in Linguistic Theory* 279. Amsterdam/Philadelphia: John Benjamins.
- . 2008. 'Evolutionary Linguistics'. *Annual Review of Anthropology* 37: 219–234. <https://doi.org/10.1146/annurev.anthro.37.081407.085156>.
- Cummins, Carla A., and James O. McInerney. 2011. 'A Method for Inferring the Rate of Evolution of Homologous Characters That Can Potentially Improve Phylogenetic Inference, Resolve Deep Divergence and Correct Systematic Biases'. *Systematic Biology* 60 (6): 833–44. <https://doi.org/10.1093/sysbio/syr064>.
- Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. Vol. I. London: John Murray. http://darwin-online.org.uk/converted/pdf/1871_Descent_F937.1.pdf.
- Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Eckert, Penelope. 2012. 'Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation'. *Annual Review of Anthropology* 41: 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>.
- Embleton, Sheila M. 1986. *Statistics in Historical Linguistics*. Bochum: Brockmeyer.
- Fauconnier, Gilles, and Mark Turner. 2003. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, Massachusetts: Sinauer Associates Inc.
- Goebel, Hans. 1982. *Dialektometrie: Prinzipien Und Methoden Des Einsatzes Der Numerischen Taxonomie Im Bereich Der Dialektgeographie*. Wien: Österreichischen Akademie der Wissenschaften.
- Gray, Russell D., David Bryant, and Simon J. Greenhill. 2010. 'On the Shape and Fabric of Human History'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 3923–33. <https://doi.org/10.1098/rstb.2010.0162>.
- Greenhill, Simon J., Thomas E. Currie, and Russell D. Gray. 2009. 'Does Horizontal Transmission Invalidate Cultural Phylogenies?' *Proceedings of the Royal Society B: Biological Sciences* 276: 2299–2306. <https://doi.org/10.1098/rspb.2008.1944>.
- Helpuhe. 2014. 'The Longitudinal Corpus of Finnish Spoken in Helsinki (1970, 1990, 2010)'. <http://urn.fi/urn:nbn:fi:lb-2014073041>.
- Honkola, Terhi, Kalle Ruokolainen, Kaj Syrjänen, Unni-Päivä Leino, Ilpo Tammi, Niklas Wahlberg, and Outi Vesakoski. 2018. 'Evolution within a Language: Environmental Differences

- Contribute to Divergence of Dialect Groups'. *BMC Evolutionary Biology* 18 (1): article 132. <https://doi.org/10.1186/s12862-018-1238-6>.
- Hull, David L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.
- Kuparinen, Olli, Liisa Mustanoja, Jaakko Peltonen, Jenni Santaharju, and Unni Leino. 2019. 'Muutosmallit Helsingin Puhekielessä'. *Sananjalka* 61: 30–56. <https://doi.org/10.30673/sja.80056>.
- Labov, William. 1994. *Principles of Linguistic Change. Vol 1. Internal Factors*. Oxford: Blackwell.
- Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Leibniz, Gottfried Wilhelm, Freiherr von. 1710. 'Brevis Designatio Meditationum de Originibus Gentium, Ductis Potissimum Ex Indicio Linguarum'. In *Miscellanea Berolinensia Ad Incrementum Scientiarum, Ex Scriptis Societati Regiae Scientiarum Exhibitis Edita, I:1–16*. Berolinum: Johan. Christ. Papeus. urn:nbn:de:kobv:b4360-1004858.
- Liu, Liang, Christian Anderson, Dennis Pearl, and Scott V. Edwards. 2019. 'Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model'. In *Evolutionary Genomics*, edited by Maria Anisimova. *Methods in Molecular Biology* 1910. New York: Humana press.
- Maurits, Luke, Mervi de Heer, Terhi Honkola, Michael Dunn, and Outi Vesakoski. 2019. 'Best Practices in Justifying Calibrations for Dating Language Families'. *Journal of Language Evolution* 5 (1): 17–38. <https://doi.org/10.1093/jole/lzz009>.
- McMahon, April, and Robert McMahon. 2006. *Language Classification by Numbers*. Oxford Linguistics. Oxford: Oxford University Press.
- Neelapu, Nageswara Rao Reddy, Malay Ranjan Mishra, Titash Dutta, and Surekha Challa. 2019. 'Role of Horizontal Gene Transfer in Evolution of the Plant Genome'. In *Horizontal Gene Transfer: Breaking Borders Between Living Kingdoms*, edited by Tomás Gonzalez Villa and Miguel Viñas, 291–314. Cham: Springer.
- Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russell D. Gray, William Martin, and Tal Dagan. 2011. 'Networks Uncover Hidden Lexical Borrowing in Indo-European Language Evolution'. *Proceedings of the Royal Society B: Biological Sciences* 278: 1794–1803. <https://doi.org/10.1098/rspb.2010.1917>.
- O'Hara, Robert. 1996. 'Trees of History in Systematics and Philology'. *Memorie Della Società Italiana Di Scienze Naturali e Del Museo Civico Di Storia Naturale Di Milano* 27: 81–88.
- Pakendorf, Brigitte. 2014. 'Coevolution of Languages and Genes'. *Current Opinion in Genetics & Development* 29: 39–44. <https://doi.org/10.1016/j.gde.2014.07.006>.
- Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. 'Inference of Population Structure Using Multilocus Genotype Data'. *Genetics* 155: 945–959.
- Sankararaman, Siriam, Swapan Mallick, Nick Patterson, and David Reich. 2016. 'The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans'. *Current Biology* 26 (9). <http://dx.doi.org/10.1016/j.cub.2016.03.037>.
- Saussure, Ferdinand de. 1916. *Cours de Linguistique Générale*. Edited by Charles Bally, Albert

Sechehaye, and Albert Riedlinger. Lausanne – Paris: Payot.

Schleicher, August. 1863. Die Darwinsche Theorie Und Die Sprachwissenschaft. Weimar: Hermann Böhlau.

Syrjänen, Kaj, Terhi Honkola, Jyri Lehtinen, Antti Leino, and Outi Vesakoski. 2016. 'Applying Population Genetic Approaches within Languages: Finnish Dialects as Linguistic Populations'. Language Dynamics and Change 6: 235–283. <https://doi.org/10.1163/22105832-00602002>.

Syrjänen, Kaj, Luke Maurits, Unni Leino, Terhi Honkola, Jadranka Rota, and Outi Vesakoski. submitted ms. 'Crouching TIGER, Hidden Structure: Exploring the Nature of Linguistic Data Using TIGER Rates'. Journal of Language Evolution.

Taylor, Scott A., and Erica L. Larson. 2019. 'Insights from Genomes into the Evolutionary Importance and Prevalence of Hybridization in Nature'. Nature Ecology & Evolution 3: 170–177. <https://doi.org/10.1038/s41559-018-0777-y>.

Verkerk, Annemarie. 2019. 'Detecting Non-Tree-like Signal Using Multiple Tree Topologies'. Journal of Historical Linguistics 9 (1): 9–69. <https://doi.org/10.1075/jhl.17009.ver>.

Weaver, Warren. 1949. 'Translation'. Memorandum. New York: The Rockefeller Foundation. <http://www.mt-archive.info/Weaver-1949.pdf>.

Wichmann, Søren, Eric W. Holman, Taraka Rama, and Robert Walker. 2011. 'Correlates of Reticulation in Linguistic Phylogenies'. Language Dynamics and Change 1 (2): 205–40. <https://doi.org/10.1163/221058212X648072>.

Wieling, Martijn, and John Nerbonne. 2011. 'Measuring Linguistic Variation Commensurably'. Dialectologia Special issue II: 141–162.