Language evolution needs its own journal

Dan Dediu[1],[*],[†] and Bart de Boer[2],[†]

[1]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and [2]Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium

*Corresponding author: Dan.Dediu@mpi.nl
[†]These authors contributed equally to this work.

Abstract

Interest in the origins and evolution of language has been around for as long as language has been around. However, only recently has the empirical study of language come of age. We argue that the field has sufficiently advanced that it now needs its own journal—the *Journal of Language Evolution*.

A brief history of recent language evolution research

Almost exactly 150 years ago, the *Société de Linguistique de Paris* famously banned the study of the origin of language in the second article of its charter:

La Société n'admet aucune communication concernant, soit l'origine du langage, soit la création d'une langue universelle. (The Society does not accept any contribution concerning either the origin of language or the creation of a universal language [our translation].*)*

Although it is a rare book or review article on the subject that does not at least allude to this ban and to its supposedly stifling effect on the study of language evolution (we ourselves plead guilty too), it is by now fairly well known that the story is not as simple as this. It certainly did not stop researchers from thinking about language evolution (nor about a universal language). Charles Darwin himself proposed in 1871 that language '[...] *owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals, and man's own instinctive cries. When we treat of sexual selection we shall see that primeval man [...] used his voice largely, as does one of* the gibbon-apes at the present day, in producing true musical cadences, that is in singing' (p. 59)—a proposal far from naïve, drawing on multiple sources of evidence and building upon deep evolutionary thinking. Throughout the years, other researchers have contributed to studying the evolution of language: Jespersen (1894, 1922) who advocated a musical origins theory, Negus (1938) who studied the evolution of the larynx, Hockett (1960) who defined the design features of language in comparison with other species' abilities, Lieberman and Crelin (1971) who used the first computer model to study the evolution of the vocal tract, or Pinker and Bloom (1990) who tried to unify the study of universal grammar with evolutionary theory. As far as we know, the first dedicated conference on the *origins and evolution of language and speech* was organized in 1975 by the New York Academy of Sciences (Harnad, Steklis, and Lancaster 1976).

Nevertheless, the *Société de Linguistique de Paris* did have a point: language evolution is by its very nature a subject that invites speculation, and it is a domain in which it is extremely difficult to find empirical answers. Speculation is tempting because language is such a fundamental part of what makes humans human that the question of its origins fascinates both scientists

and non-scientists alike. At the same time, hard evidence is scarce as it deals with events from the remote past, and the information is therefore usually very indirect. To add to the difficulty, a lot of this evidence is spread over many disciplines. Finally, language is a complex phenomenon in its structure, its history, and its instantiation in the brain and in human society. Small wonder that it is not easy to investigate its origins!

However, a number of factors have conspired to make language evolution more amenable to empirical investigation than ever before, and the field has sufficiently advanced that it now merits its own journal. One extremely important development is the enormous *increase in knowledge* that can be brought to bear on the questions of language origins, manifest not only in the emergence of whole new relevant fields such as brain imaging techniques or the extraction and analysis of DNA from extinct humans but also in the deeper scientific insight into animal behavior, human cognition, language acquisition, language typology, archaeological evidence, the genetic bases of language and speech, and many other areas.

We think that there are other crucial and perhaps somewhat less obvious developments. One is that *interdisciplinary exchange* has become much easier, mostly because of the Internet, which has facilitated communication between researchers, allowed better and faster access to the results and methods from different disciplines, and the powerful search engines that help locate formerly obscure information. However, we should also not forget that direct human interaction is key even to science, and here an essential role is played by The *Language Origins Society* conferences (since 1983), the EVOLANG series of conferences (held biannually since 1996), and increasingly by the Protolang conferences (now at their fourth iteration). The result of this increase in inter-disciplinary exchange is threefold: first, it is now possible to integrate the different pieces of the language origins puzzle more easily. Second, there is a growing realization that language is not an isolated phenomenon that is disconnected from other human behaviours and the environment or from the behaviour of other animals, but that there are in fact many relevant links between language and other phenomena. Third, the increased exchange of ideas has helped enrich the repertoire of methods, tools, and paradigms that allow the investigation of questions about language evolution empirically and, perhaps more importantly, it helped to make researchers realize which questions are amenable to empirical investigation (and which ones should best be left for campfire stories).

Another key factor is the availability of *computer simulations*. This has allowed us to tackle the complexity involved in the dynamics of language origins, to reconstruct history, to explore topics where it is very difficult to conduct experiments, and to do powerful analyses of complex data. Computer models allow us to go way beyond anything that could ever be achieved by pen-and-paper analysis. Although we can never be sure what happened exactly in language evolution, computer models help us to determine which scenarios are more plausible than others.

Empirical, data-driven outlook

We believe that probably the most important development in the field of language evolution of the last decades has been the full recognition that we need to base our theorizing on *actual data* (or to formulate it more negatively, that speculation unfounded in data is no longer acceptable) and the realization that we can in fact *collect* such data in a controlled and principled manner. Moreover, there has been great progress in tools to *analyze* the complex kind of data that we need to deal with.

First, while almost everybody would agree that theorizing should ultimately rest on some empirical data, it is far from clear what sort of data and how much of them are required for such theorizing. One of the debates, for example, concern the relevance of present-day languages for questions concerning language origins. Is historical linguistics with its sound changes relevant? Are typology and the patterns of linguistic diversity? Is child language acquisition? How about the neurobiological underpinnings of language processing?

We believe that a proper understanding of such present-day data is the only valid path toward a science of language evolution, and it is this belief that justifies the contents of the first issue of this journal featuring an introduction to linguistic diversity, a debate on the effects of climate on current linguistic diversity, and an introduction to cutting-edge statistical analysis of experimental results from living humans.

But what sort of data would we like to see? Our answer: any sort we can get! We believe that besides the language sciences (and we include here theoretical linguistics, historical linguistics, phonetics, typology, psycholinguistics, sign language studies, sociolinguistics, etc.), the medical sciences, animal models, genetics, (palaeo)anthropology, and psychology (among others) are highly relevant, but we will turn to this multidisciplinarity in the next section.

Given this wide-ranging outlook, collecting these data is of the highest importance, but it raises its own

questions. Obviously, such data are extremely heterogeneous, ranging from, say, gene expression profiles coming from cell lines grown in a dish, to the use of pointing gestures by non-human primates, to patterns of site occupation and use by extinct humans, to the distribution of tone contrasts in the present-day languages, to substratum effects in language shift, to brain event-related potentials during the processing of an ergative sentence, to the regularization of morphological paradigms during the repeated transmission of language across generations in the lab, and so on. Obviously, these data must be collected to the highest standards of the respective scientific disciplines, but additionally in such a way that they can be used to make cross-disciplinary inferences.

Such data, in the absence of proper analysis, are inert and useless. By analysis we mean all the appropriate techniques for checking the fit of the data to predictions (in the vein of the honored hypothesis testing approach to science), but also the newer methods of exploring data, looking for hidden patterns and for generating new hypotheses to be tested. More and more advanced and rigorous methods of data analysis have allowed surprising findings such as the influence of non-linguistic factors on present-day linguistic diversity (e.g., Everett, Blasi, and Roberts 2015) or the identical mechanical properties of modern and Neandertal hyoid bones (D'Anastasio et al. 2013), to mention just a few.

Computational modeling

Computer models have advantages over verbal arguments, making them an important tool in investigating language origins and evolution. A computer model must be *fully specified*, including its assumptions, components, processes, or parameters, allowing a much-needed transparency and clarity. They are *reproducible* and *objective*: one can simply re-run or re-implement the model and check if the reported results can be reproduced, and one can easily tweak its various aspects and investigate the effects of such interventions. Computational models can be extremely *complex*, and while too much complexity can be a hindrance, in some cases (language included) complexity cannot be avoided: to understand language origins and evolution, one must go beyond isolated biological individuals, and must instead model societies that live in, and create a cultural niche. Finally, no matter how powerful and useful computer models are, the ultimate check comes from actual data: models can be *falsified* by comparing their predictions to the data, and it is necessary that their underlying assumptions fit what we already know about the world.

There are many examples of excellent computational models to select from, and we can illustrate the width of approaches with just a few: there is of course Lieberman and Crelin's (1971) classic computational study of Neanderthal *vocal abilities*, which has engendered a literature of its own, the latest installment probably being Badin et al. (2014). A different computational tradition uses agent-based models to study how language emerges in societies of artificial agents, including the influential *iterated learning* (e.g., Smith, Kirby, and Brighton 2003) and Steels' (1995) *language game* models. There are also models of a more mathematical nature, such as those inspired by theoretical biology proposed by Nowak and his students (e.g., Nowak 2006). Yet another class is exemplified by Dunn et al.'s (2011) use of Bayesian phylogenetic models inspired from evolutionary biology to investigate language change.

Massive multi- and inter-disciplinarity

Inquiries into language origins and evolution have always drawn upon several disciplines at once, but the last few decades have seen an explosion in the breadth and depth of such collaborations between disciplines. The discovery of *FOXP2* and the continued exploration of its functions, evolution, and relevance for language and speech can serve as an example (e.g., Fisher and Vernes 2015). It involves molecular genetics, evolutionary genetics, bird song biology, language and speech pathology, cognitive neuroscience, linguistics, archaeology, paleoanthropology, and paleoecology, not to mention sophisticated computer models required for analysis.

The interaction between disciplines is not only limited to the exchange of data but also involves learning from each other's methods. For instance, computer models have helped linguists, social scientists, and cognitive scientists make their theories more explicit. Conversely, computer modelers have on the one hand been influenced by theoretical biologists, physicists, and mathematicians to make their models simpler (more analyzable), while on the other hand, they have been influenced by empirical scientists to better link their models to empirical data. Moreover, the field of language evolution has now developed a set of experimental paradigms of its own to test language evolution in a laboratory setting (e.g., Galantucci 2009; Kirby, Griffiths, and Smith 2014).

Mastering such a diversity of paradigms, methods, and results far exceeds anyone's capacity (although we have the impression that some young students in the field try valiantly). Therefore, it becomes almost inevitable that progress in the field of language evolution will require interdisciplinary efforts, most likely produced by teams of authors from different disciplines. Of course it will remain possible to make individual contributions,

but even individual efforts may make use of multiple disciplines and will be addressed at an interdisciplinary audience.

The combination of the above developments means that, on the one hand, there is more high-quality work relevant to language origins and evolution than ever but, on the other, that it is harder for such work to find a 'natural place' given its use of increasingly specialized methods and datasets while being multi-disciplinary. This has resulted in a fragmentation of the literature and a lack of cross-talk between the relevant disciplines.

Work pertaining to the evolution of language has been published in journals such as *Science*, *Nature* (or *Nature Genetics*, etc.), *PNAS*, *Proceedings of the Royal Society B*, *PLoS One* (or *PLoS Genetics*, etc.), *Frontiers In* (Language, Psychology, ...), *Journal of Phonetics*, *Journal of Human Evolution*, *BBS* (and so on) and in edited books and conference proceedings (such as the Evolang series): as a researcher, it is impossible to keep track of all these venues!

This fragmentation also results in unequal ways of reporting. Various disciplines have different requirements on what is acceptable, including how to conduct an experiment, the value of observational data, the appropriate statistical techniques, the use of exploratory data analysis, or the inclusion of a discovery and a replication sample. Also, different journals and books have their own requirements, starting from the word limits and bibliography style, to more substantive issues concerning the review process (blind or not), the number of reviewers, and the role of the editors, to what is an appropriate contribution (e.g., how relevant is the potential impact). Finally, less specialized journals, such as most of the ones mentioned above, tend not to accept solid negative results, even though these may be important for preventing researchers exploring the same dead ends and for meta-analyses.

Aims and mission of the journal

The *Journal of Language Evolution* (*JoLE*) will answer the needs for a unified forum of researchers investigating language origins and evolution. *JoLE* aims to be the *venue of choice* for language evolution research, attracting the very best research concerning these topics and being the first option for publishing such research. This goal is extremely ambitious given the strong competition from both generalist and specialized journals that already publish such research but also the widely different criteria relevant for the various disciplines.

JoLE is by design *highly inter- and multi-disciplinary*, welcoming technical contributions from various disciplines as well as across disciplines as long as their relevance for the broader questions of language origins and evolution are clarified. *JoLE* will encourage the real cross-disciplinary understanding and sharing of core paradigms, methods, and results by publishing *Introductions*, *How-to*'s, and *Reviews* targeted at a broad scientific audience, but we will steer away from any unnecessary oversimplifications and 'dumbing-down'. *JoLE* will aim to unify standards across disciplines, having a single model of peer-review (double blinded: the authors' and reviewers' identities are hidden from each other) and submission process, a single core set of values (highest scientific quality, preference for empirical research and methodological soundness, meeting the standards of the relevant disciplines while still relevant to the overarching questions), and a unified format.

JoLE welcomes contributions using a variety of *methodological approaches* including theoretical work (with an empirical component or making empirically testable predictions), computational investigations (testing predictions or exploring complex systems), database-driven (using modern statistical and quantitative methods), and experimental work, as long as the methods and their application are sound, their assumptions are met and the interpretation of their results is realistic. *JoLE* accepts work emerging from a variety of *disciplines* including linguistics, cognitive (neuro-)sciences, psychology, anthropology, biology, evolutionary theory, computer sciences, philosophy, and many others, as long as the contributions meet the highest standards of quality in the disciplines they draw upon (judged by specialist reviewers) but are also relevant to the wider questions and the many fields involved (as judged by reviewers from different disciplines and reinforced by the handling editor).

These aims are reflected by the members of the *Editorial Board* and the *Associate Editors*, covering a variety of scientific fields while being interested in the 'big questions' and capable and willing to communicate across disciplinary borders. In inviting the *JoLE* editors, the accent was on scientific excellence, a wide coverage of the relevant scientific fields, a fair gender and age balance, and the representation of the major schools of thought around the world.

Contributions to the journal

The major types of contributions are listed below:

Research articles and **Short reports** should focus on the actual data, methods, and results of a scientific investigation relevant to language origins and evolution. They are not judged by their potential impact; instead

the scientific quality and empirical foundations are the main criteria for selection. Research articles are longer (up to 8,000 words) than Short reports (up to 3,000 words), the latter being much more concise and focused, benefitting, in exchange, from a much faster review process. In our view, good science means first and foremost replicable science, and we encourage both *negative results* and *replications* (both successful and failed) as long as the design is sound, faithfully addresses the original research to be replicated, and they are accompanied by a sound statistical power analysis, so that failures due to insufficient power or inadequate design are ruled out.

Methodology articles are medium-size contributions (up to 5,000 words) that address the need for methodological innovation and transfer that are one of the main directions of growth in language evolution research. They are focused on introducing novel methods (genuinely new or new to language evolution) even in the absence of stand-alone publishable results but having passed benchmarks that ensure their validity and reliability. These methods can come from any relevant discipline and concern data analysis, data acquisition, interpretation, and presentation.

Reviews, **Debates**, and **Target articles** are very important tools for advancing the field and promoting cross-disciplinary communication. Reviews should summarize, in a comprehensive, objective, and nonpartisan manner, recent developments relevant to language origins and evolution in a specific field or across several fields. They should not propose new major theories or further individual agendas or opinions, these being expressed through Debates and Target articles. For the first, two opposing points of view are presented in two separate but mutually engaging articles, while for the second, a controversial, new, or unexpected proposal is presented, followed by short comments from peers and a response from the original authors. These are large contributions (up to 10,000 words) and must offer a comprehensive literature review supporting their claims, must be wide-ranging, cross-disciplinary, and relevant to a wide variety of scientists interested in language origins and evolution.

Introductions and **How-tos** are answering the problems generated by the massive inter- and multidisciplinarity of the journal and the breath-taking speed at which science advances. These two types of contributions can reach 5,000 words and are aimed at presenting very clearly and accessibly foundational paradigms, concepts, methods, and results that must form the common ground for language origins and evolution research. While the first can be quite theoretical in nature, the latter are very practical, showing exactly how to do things,

what buttons to press, what commands to type, how long to centrifugate, or how to elicit smell terms from your informants.

JoLE also accepts **book reviews** and **news** and **announcements**, and for a full and updated list, please consult the journal's website at http://jole.oxfordjournals.org/.

Conclusions

Language origins and evolution research need a dedicated forum, and we hope that the *JoLE*, reflecting the current trends and composition of the relevant scientific fields, will fulfill that role. Our view is that language origins and evolution is now a mature scientific topic, that there are methods and datasets that can answer its questions and that, more importantly, there is a set of shared paradigms and values that make effective collaboration across the disciplines possible and fertile. Our belief is that the questions concerning the way language came to be the way it is are relevant to everybody as we each try to continuously define ourselves and that *JoLE* will be a major force in helping those answers take shape.

Acknowledgements

Funding

References

Badin, P. et al. (2014) 'Keep the Lips to Free the Larynx: Comments on de Boer's Articulatory Model (2010)', *Journal of Phonetics*, 46: 161–7.

D'Anastasio, R. et al. (2013) 'Micro-Biomechanics of the Kebara 2 Hyoid and Its Implications for Speech in Neanderthals', *PLoS One*, 8: e82261.

Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Dunn, M. et al. (2011) 'Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals', *Nature*, 473: 79–82.

Everett, C., Blasi, D. E., and Roberts, S. G. (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: 1322–7.

Fisher, S. E., and Vernes, S. C. (2015) 'Genetics and the Language Sciences', *Annual Review of Linguistics*, 1: 289–310.

Galantucci, B. (2009) 'Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action', *Topics in Cognitive Science*, 1: 393–410.

Harnad, S. R., Steklis, H. D., and Lancaster, J. (1976) *Origins and Evolution of Language and Speech*. New York: New York Academy of Sciences.

Hockett, C. (1960) 'The Origin of Speech', *Scientific American*, 203: 88–111.

Jespersen, O. (1894) *Progress in Language: With Special Reference to English*. London: Swan Sonnenschein & Co.

——. (1922) *Language, Its Nature, Development and Origin*. London: reprinted Allen and Unwin, 1968.

Kirby, S., Griffiths, T. L., and Smith, K. (2014) 'Iterated Learning and the Evolution of Language', *Current Opinion in Neurobiology*, 28: 108–14.

Lieberman, P. H., and Crelin, E. S. (1971) 'On the Speech of Neanderthal Man', *Linguistic Inquiry*, 2: 203–22.

Negus, V. E. (1938) 'Evolution of the Speech Organs of Man', *Archives of Otolaryngology*, 28: 313–28.

Nowak, M. A. (2006) *Evolutionary Dynamics*. Cambridge, MA: Harvard University Press.

Pinker, S., and Bloom, P. (1990) 'Natural Language and Natural Selection', *Behavioral and Brain Sciences*, 13: 707–84.

Smith, K., Kirby, S., and Brighton, H. (2003) 'Iterated Learning: A Framework for the Emergence of Language', *Artificial Life*, 9: 371–86.

Steels, L. (1995) 'A Self-Organizing Spatial Vocabulary', *Artificial Life*, 2: 319–32.