EVALUATING THE ROLE OF QUANTITATIVE MODELLING IN LANGUAGE EVOLUTION

MAURICIO MARTINS[1,3]
ARCHISHMAN RAJU[2]
ANDREA RAVIGNANI[1]

[1]*Department of Cognitive Biology, University of Vienna, Althanstrasse 14, Wien, 1090, Austria*

[2]*St. Stephen's College, Delhi University, University Enclave, New Delhi, 110007, India*

[3]*Language Research Laboratory, Lisbon Faculty of Medicine, Av. Professor Egas Moniz, Lisboa, 1649-028, Portugal*

Models are a flourishing and indispensable area of research in language evolution. Here we highlight critical issues in using and interpreting models, and suggest viable approaches. First, contrasting models can explain the same data and similar modelling techniques can lead to diverging conclusions. This should act as a reminder to use the extreme malleability of modelling parsimoniously when interpreting results. Second, quantitative techniques similar to those used in modelling language evolution have proven themselves inadequate in other disciplines. Cross-disciplinary fertilization is crucial to avoid mistakes which have previously occurred in other areas. Finally, experimental validation is necessary both to sharpen models' hypotheses, and to support their conclusions. Our belief is that models should be interpreted as quantitative demonstrations of logical possibilities, rather than as direct sources of evidence. Only an integration of theoretical principles, quantitative proofs and empirical validation can allow research in the evolution of language to progress.

1 Introduction

It is clear that modelling has gained in importance in studies of language evolution. For example, at the latest International Conference on the Evolution of Language (Evolang IX), from a total of 83 talks, 19 (23%) reported results from quantitative models. These models make hypotheses about events in the past that are no longer observable, and explore whether these premises could lead to conclusions compatible with the current state of matters. They are, therefore, often able to give insights into the potential stages and mechanisms involved in the emergence of language. For instance, computer models have played an important role in demonstrating that cultural transmission, in the absence of strong biological constraints, can be seen as a plausible mechanism leading to language universals (Kirby, 2001). Quantitative models also generally require greater "clarity and rigor" (Hurford, 2002) than non-quantitative theorizing and "enforce explicitness in the formulation of an explanation" (Christiansen & Kirby, 2003).

Despite the potential benefits of using quantitative tools to model language evolution, there are important pitfalls that can arise from taking models as direct evidence to support or refute specific theoretical frameworks. This practice has generated an unwarranted confidence that modelling can provide data similar in quality to experimental work (Cangelosi & Parisi, 2001). In this paper, we argue that this is a dangerous conclusion. With examples from recent literature, we show how different scenarios rendering different conclusions can be coherently modelled, without an obvious procedure to assess which one better approximates the reality it is built to simulate. We think models should only be used as an intermediate step between theory and experiments. Given this, we examine the role of empirical data in justifying these models.

2 The Problem of Diversity

One of the potential problems of treating models as equivalent to empirical tools is that too far a range of situations can be plausibly modelled. For instance, consider the debate regarding the relative weight of biological and cultural evolution in shaping language structure. Some models have suggested that the distribution of different languages in a population reflects the (innate) priors of individual agents (Burkett & Griffiths, 2010; Griffiths & Kalish, 2007). Other models propose instead that weak innate biases are sufficient for languages to become established in populations, since mechanisms of cultural transmission have the power to amplify weak biases (Kirby, Dowman, & Griffiths, 2007; Smith, 2009). One important assumption of these models is how, given a set of different hypotheses\languages, agents select which hypothesis\language better explains the utterances they are exposed to. The role of innate biases seems to be emphasized when agents are assumed to choose a language proportionally to its probability in the posterior hypothesis space; and the effects of innate biases are minimized when agents are assumed to choose disproportionally more often the language with the highest probability in the hypothesis space. As Smith (2009) points out, "the true nature of human hypothesis selection strategy is ultimately an empirical question". This should be kept in mind when interpreting the results of different models (Burkett & Griffiths, 2010; Griffiths & Kalish, 2007; Kirby et al., 2007; Smith, 2009) which have different assumptions and support contrasting conclusions about language evolution.

The broader problem is that the same mathematics can be used to argue from opposite theoretical views, such as the idea of an innate universal grammar and the notion of a pure cultural evolution of language (Griffiths, Kalish, & Lewandowsky, 2008; Nowak, Komarova, & Niyogi, 2001). Furthermore, models with completely different mechanistic principles can generate very similar outputs. For instance, both the principles and parameters and the iterated learning frameworks can be used to agree with the same data in historical linguistics (Niyogi & Berwick, 1997; Reali & Griffiths, 2009).

Further examples of model flexibility abounded during Evolang IX. While Baronchelli et al. (2012) argue that cultural change is too fast for biological evolution to play a role in the structure of languages, McCrohon (2012) demonstrates that some aspects of language resistant to change may be targets of biological evolution. This diversity of scenarios would not be problematic, if the models were taken as demonstrations of logical possibilities instead of direct sources of evidence. Unfortunately, statements from the conference proceedings suggest more than the mere demonstration of logical possibilities: "Cultural transmission renders the biological evolution of strong domain-specific innate constraints unlikely" (Thompson, Smith, & Kirby, 2012); "Only adaptations for flexible learning combined with continual cultural evolution can reconcile the diversity of languages, with the biological uniformity of their speakers" (Baronchelli, Chater, Pastor-Satorras, & Christiansen, 2012); "duality of pattern emerges as a consequence of pure cultural dynamics" (Tria, Galantucci, & Loreto, 2012).

We are aware that modelling research in all disciplines usually draws conclusions under the tacit assumption that these are only valid within the theoretical scope of a specific modelling framework. However, given the particularly multidisciplinary readership of language evolution research, the connection between model results and empirical reality should always be stated. Given that several alternative scenarios can be all coherently modelled, there is a fundamental problem in assessing which model among the many provides a better description of the mechanisms involved in language evolution. This assessment is impossible without external instruments of validation.

3 Model Validation

There are two general ways to assess the validity of quantitative models. A first, obvious method is the comparison of the model outcomes with the reality it exemplifies. Whenever a model is inconsistent with observable facts, then a reformulation of the model is necessary. For instance, if a certain model attempts to simulate the establishment of conventionalized communication systems, then the failure to do so generally argues against its validity. This method of validation is not, however without pitfalls. As mentioned above, both the principles and parameters and the iterated learning frameworks can be used to fit the same historical data.

Moreover, the problem of using only historical data is further demonstrated by other disciplines such as social sciences and economics, where models built to fit past data failed to predict future outcomes. Such failures render these models poor approximations of reality and reflect the disregard of important variables, the reliance on false assumptions or the impossibility to adequately describe a complex system with few minimalistic equations.

A second method to assess the validity of models is to check whether assumptions and selected parameters hold against empirical inquiry. For example, in classical (normative) microeconomics, agents (humans) are assumed to be rational decision-makers. Research in psychology has proven this assumption wrong (Kahneman & Tversky, 1979). Recent models on iterated decision-making (Camerer, Ho, & Chong, 2004) do incorporate these assumptions of irrationality and cognitive limitations, fitting well both *empirical evidence on processes' outcomes* and, crucially, *mechanistic constraints leading to such outcomes*.

3.1 Examples from Social Sciences

There is a basic, widespread family of models that can be found, explicitly or tacitly used, both in the language evolution literature and in the social sciences. Both areas engage themselves in the hard task of modelling and predicting states of complex systems characterized by a number of unobservable variables. Markov models, in all their simplicity and flavours, are a building block of applied mathematics. Simplicity and elegance are a tacit, golden rule of modelling. However, these come at a cost: few parameters can, if any, only capture the most basic forces in a complex system. It is therefore important for modellers to always keep in mind this essentiality vs. depth of scope trade-off when interpreting model results. Moreover, the failure of a model to predict an event in another discipline should instigate caution in drawing inferences on the evolution of language using a closely related modelling technique.

An interesting example of this is a modelling study by Schrodt (2000) investigating conflict patterns in the former Yugoslavia war. The panel data consist of weekly reports about the magnitude of the conflict (low vs. high) over a period of 8 years. Two hidden Markov models are created, corresponding to low and high conflict magnitudes. Independent variables include several possible predictors of conflict (ranging, for instance, from demonstrations to ethnic expulsions). The goodness of the model is then evaluated as the ability of a time pattern of events to be attributed to - therefore explained by the model of - high or low conflict periods. The results suffer from an issue common to such models: accuracy (how well an outcome is predicted), sensitivity (how uncertainty propagates from assumptions to results) and precision (variability in replicated forecasts) of the model have to be traded for one another. Here, for instance, 80% accuracy corresponds to 25% sensitivity and 60% precision, while a model with 92% short-run (near future) accuracy has only 30% precision. The forecasting horizon emerges as a fourth parameter in assessing the goodness of Schrodt's model. This is extremely common in any model containing a time component: the closer in time the prediction, the higher its reliability. Results on forecasting horizon from other disciplines appear particularly relevant to language evolution research: Language models deal with time spans which are several orders of magnitude greater that the critical time horizons in the social sciences.

As another parallel, it is interesting to mention what Cowles (1932) reports on the ability of economic models to forecast future market developments, notably at a time when mathematical economics was starting to develop. Cowles, an advocate of mathematical accuracy in economics, reviews 45 different attempts at predicting stock markets, made by just as many prestigious agencies and financial publications. He compares these forecasts to null statistics achievable by pure chance and finds little difference between these two ways of predicting unknown events. "The most successful records are little, if any, better than what might be expected to result from pure chance. There is some evidence, on the other hand, to indicate that the least successful records are worse than what could reasonably be attributed to chance." (Cowles, 1932)

Twelve years later, as modelling techniques advance, Cowles (1944) finds evidence of some predictability in stock prices. He claims that "A simple application of the "inertia" principle, such as buying at turning points in the market after prices for a month averaged higher, and selling after they average lower, than for the previous month, would have resulted in substantial gains for the period under consideration". Paraphrasing, a Markov process with three states and one month lag was better than chance in suggesting fruitful investments. These simple models have since then been replaced by more complex ones, which became the standard college-level tuition in finance and whose basic assumptions were in turn criticized down to the bone (Hudson & Mandelbrot, 2004). What happened to stock market forecasting is already history.

Our point is that if a specific technique used in modelling language has demonstrated lack of robustness or predictive power when used in other disciplines, researchers should be cautious in interpreting the result of their model. To make matters worse, language evolution models are designed to predict only past, rather than future events. Hence, if different evolution models fit the historical data equally well, it may be impossible to refute these models based on their outcome. This renders the selection and validation of a priori assumptions even more fundamental. Within the field of language evolution, this empirical validation comes from historical linguistics, psychology, biology and demographics. In the next section we will discuss a paradigmatic example of this interaction between models and empirical data: the iterated learning model.

3.2 Iterated Learning Model

The "iterated learning model" (Kirby, 2001) involves a meaning space, a signal space, a learner and a teacher. At each step of the model, the teacher agent produces a string. The learner tries to construct the most parsimonious mapping between meaning and symbols. It has been observed that there is a gradual regularization of this mapping over many steps of the model. The iterated learning model has been linked to two distinct sets of empirical tests.

The first source of empirical evidence supporting the iterated learning model comes from psychological experiments, in which people are asked to make signal-object associations (Kirby, Cornish, & Smith, 2008). The experiments show evidence of compositionality emerging spontaneously in an artificial language after a few iterations of cultural transmission, an outcome similar to previous simulation results. These experiments are important to demonstrate that language-specific features can emerge from cultural mechanisms (without the need of language-specific biological adaptations). However, they do not constitute a strong proof that evolution of language occurred due to these mechanisms alone. One reason to be cautious about these inferential jumps is raised by Littauer (2012): "Early language communities may have had different pressures on linguistic evolution and morphological complexity than modern languages, including differences in the amount of shared information. (…) Agreement is an integral part of language evolution, and the origin of agreement in protolanguage may not have followed the same paths as modern agreement formation processes". In fact, the psychological experiments conducted to validate the model of iterated learning were performed on modern humans, while pre-linguistic hominids may have had different cognitive skills and biases. With this setup, one can investigate mechanisms of cultural transmission under the cognitive biases that are characteristic of modern humans, but one cannot assess whether the evolution of these biases in the hominin lineage might have changed the compositional features of our communication signals.

Still within experimental psychology, we can find other reasons warning against straightforward conclusions: When horizontal transmission (intra-generation) is included along with vertical transmission (inter-generation), the scenario of cultural evolution becomes more complex (Berdicevskis, 2012; Tamariz, Cornish, Smith, Roberts, & Kirby, 2012). For example, the need to negotiate meaning with conspecifics in horizontal transmission leads to low fidelity in vertical transmission (Berdicevskis, 2012), even in a scenario of a highly structured language (*contra* (Kirby et al., 2008).

The second source of evidence comes from historical linguistics. Recently, Reali and Griffiths (2009) showed that three distinct language features can be explained by a neutral iterated learning model. These include the characteristics of verb regularization, Zipf's law and the character of innovations in language. The authors propose that neutral models should be used as null models for language dynamics. Their opinion is that if neutral models can be used to explain a particular characteristic of languages, then there is no need to appeal to selective forces. Recently, Blythe (2011) also justified that neutral models qualify as good null models for language dynamics. While this agreement of theory and data from historical linguistics might be both interesting and useful, the data is unfortunately over a time scale much smaller than that of language evolution. Since the evolution of the faculty of language refers to a period of time that goes back at least up to the last common ancestor between humans and chimpanzees, its (cognitive, ecological, cultural, etc.) context differs from that of the evolution of languages (object of concern for historical linguistics). Naturally, the theoretical model cannot be extrapolated to a different time scale and a different set of conditions.

4 Better Evaluation Techniques and Better Models

We think models can be useful if there are better common practices in their evaluation and validation. There is, however, the danger that modelling research will go towards increasingly complex models that rely on an increasing number of plausible hypotheses which rely on the intuition of the modeller. It is important to instead focus on identifying and limiting the hypotheses within these models and look for ways of testing them. As pointed out, a crucial method to evaluate models is to check whether both their *assumptions* and *mechanistic constraints* are supported by independent experimental research. This applies to a variety of assumptions usually only scrutinized by plausibility analyses. For instance, we should test how humans select hypotheses before assuming a particular strategy (Smith, 2009); and we should test whether humans display irrational biases or cognitive limitations on certain tasks before assuming they will behave as rational agents in processes involving those tasks (Camerer et al., 2004; Kahneman & Tversky, 1979). If it turns out that human behaviour departs from rationality, realistic models should incorporate these constraints.

An important point for modelling language evolution is the question of what biological/cognitive evolution means, and how it can be encoded in models. The evolution of language is the product of both biological and cultural evolution. The weight of each of these processes in shaping language structure is currently unknown. However, modelling discrete stages of biological evolution requires the incorporation of discrete cognitive constraints present in the populations evolving. For instance, while humans can encode hierarchical information in vocal utterances, it is unlikely that cats do so, even though both species can communicate vocally. It is unlikely that successive generations of cats communicating vocally will develop a hierarchical system of communication, even if this communication strategy would be the most efficient. A realistic model of evolution of vocal communication, from the last common ancestor between humans and cats up to the modern human, would need to somehow encode these cognitive constraints, and the evolution of these constraints across successive generations of agents (Jones & Love, 2011). Cues about the processes underlying the evolution of these cognitive constraints can be obtained using a comparative approach, in which different species are experimentally tested in similar cognitive tasks, with the goal of identifying the cognitive constraints of each. Though it might be impractical to include such fine-grained information, perhaps some general law of 'cognitive evolution' will emerge from the comparative approach. This law could be incorporated in future models of language evolution.

5 Conclusion

In this paper we have suggested how a large variety of scenarios can be coherently modelled. Since different scenarios are achieved by selecting different parameters and assumptions, this selection process should be rigorously scrutinized. We have shown that leaving fundamental parameters out of models may bias the conclusions in order to meet previous theoretical assumptions. If computer simulations are taken as direct evidence to support particular theories over others, then there is a danger of the scientific process becoming circular. The role of experimental work is fundamental to validate both models' results and assumptions. Finally, we note that it is not enough to check models against empirical data, but also to examine the relevance of this agreement to language evolution. Data relevant to the evolution of languages is not necessarily relevant to the evolution of the faculty of language. A wider range of validation techniques will be required.

We see quantitative modelling as an intermediate step between theory and experiment. It helps improve on theory by clarifying assumptions, adding insights and showing the plausibility or processes. It can help analyze available data and lead us to new sources of empirical evidence. It is then a worthwhile endeavour as long as it continues to interact with experiments and data. If the results of these models are carefully interpreted, they could play an important role in our understanding of language evolution.

Acknowledgments

This work was supported by the FCT grant SFRH/BD/64206/2009 to M. Martins, by the ERC Advanced Grant SOMACCA, project number 230604 (awarded to W.T. Fitch), to A. Ravignani and by a KVPY fellowship to A. Raju.

References

Baronchelli, A., Chater, N., Pastor-Satorras, R., & Christiansen, M.H. . (2012). *Reconciling the diversity of languages with the biological uniformity of their* *speakers*. Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

Berdicevskis, A. (2012). *Introducing pressure for expressivity into language evolution experiments*. Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

Burkett, D., & Griffiths, T.L. (2010). *Iterated learning of multiple languages from multiple teachers*. Paper presented at the The Evolution of Language - Proceedings of the 8th international conference, Utrecht.

Camerer, C., Ho, T.H., & Chong, J.K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics, 119*(3), 861-898.

Cangelosi, A., & Parisi, D. (2001). Computer simulation: A new scientific approach to the study of language evolution. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 3-21). London: Springer.

Christiansen, M.H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300-307. doi: 10.1016/s1364-6613(03)00136-0

Cowles, A. (1932). Can stock market forecasters forecast? *Econometrica, 1*, 309-324.

Cowles, A. (1944). Stock market forecast. *Econometrica, 12*, 206-214.

Griffiths, T.L., & Kalish, M.L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science, 31*(3), 441-480.

Griffiths, T.L., Kalish, M.L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3503-3514. doi: 10.1098/rstb.2008.0146

Hudson, R., & Mandelbrot, B. (2004). *The (mis)behavior of markets: A fractal view of risk, ruin, and reward*. New York: Basic Books.

Hurford, J. (2002). Expression/induction models of language evolution: Dimensions and issues. In E.J. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 301-344). Cambridge: Cambridge University Press.

Jones, M., & Love, B.C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences, 34*(04), 169-188. doi: 10.1017/s0140525x10003134

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Kirby, S. (2001). Spontaneous evolution of linguistic structure— an iterated learning model of the emergence of regularity and irregularity. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 5*(2), 102-110.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686. doi: 10.1073/pnas.0707835105

Kirby, S., Dowman, M., & Griffiths, T.L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, 104*(12), 5241-5245. doi: 10.1073/pnas.0608222104

Littauer, R. (2012). *The evolution of morphological agreement.* Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

McCrohon, L. (2012). *Simulating the effects of cultural coadaptation on rates of endogenous and exogenous language change.* Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

Niyogi, P. , & Berwick, R.C. (1997). A dynamical systems model for language change. *Complex Systems, 11*, 161-204.

Nowak, M., Komarova, N., & Niyogi, P. (2001). Evolution of universal grammar. *Science, 291*(5501), 114-118. doi: [DOI:10.1126/science.291.5501.114]

Reali, F., & Griffiths, T.L. (2009). Words as alleles: Connecting language evolution with bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences, 277*(1680), 429-436. doi: 10.1098/rspb.2009.1513

Schrodt, P.A. (2000). *Forecasting conflict in the balkans using hidden markov models*. Paper presented at the American Political Science Association meetings, Washington, DC.

Smith, K. (2009). *Iterated learning in populations of bayesian agents.* Paper presented at the Proceedings of the 31st annual conference of the cognitive science society.

Tamariz, M., Cornish, H., Smith, K., Roberts, S., & Kirby, S. (2012). *The effects of generation turnover and interlocutor negotiation on linguistic structure.* Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

Thompson, B., Smith, K., & Kirby, S. (2012). *Cultural evolution renders linguistic nativism implausible.* Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.

Tria, F., Galantucci, B., & Loreto, V. (2012). *Naming a structured world: A cultural route to duality of patterning.* Paper presented at the The Evolution of Language - Proceedings of the 9th International conference Kyoto.