Models of Language Evolution:
Does the Math Add Up?

Bart de Boer
Willem Zuidema
{b.g.deboer, w.h.zuidema}@uva.nl

1 — Introduction

The last two decades have seen an explosion of interest in mathematical and computational models of language evolution. Formal modelling is seen by increasingly many in the field as an approach that ensures internal consistency of evolutionary scenarios. However, there has been little attention to the question of how well the many different models fit together. Are they consistent with and complementary to each other? Is there a series of models that really covers the evolutionary emergence of modern language from a prelinguistic, ancestral state? Are the assumptions that go into a particular model, if not (yet) supported by empirical findings, made plausible by results from other models?

In this paper, we argue that these problems deserve much more attention than they currently receive. For sustaining the success of modelling approaches in language evolution research, it is crucial that models start living up to their promise: modellers must make explicit how their models fit in with other work in complete scenarios on the origins of language(s), and how their modelling results affect judgments of plausibility of one scenario against another. Moreover, they must do so based on careful consideration of other work, without overstating their results and misusing the prestige that comes with mathematical approaches.

Our arguments are based on a particular view on the role of modelling in scientific research in general, and in "historical" research fields with a paucity of direct evidence in particular. In section 2, we will therefore start with some considerations about the methodology of modelling in our field. To ground the discussion, however, we will quickly move to concrete examples. In section 3 we will discuss the contributions and shortcomings of models of the evolution of speech. In section 4 we will then draw some general lessons from this case study, and sketch an agenda for future research in the language evolution modelling field at large.

2 — Modelling methodology

From the great many distinctions one can make between different model studies, there are three particularly useful ones that also allow us to establish some common terminology and formulate the goals of this paper. The first is a distinction based on function, between predictive models and explanatory models (Gilbert & Troitzsch, 2005). Predictive models try to model a system as accurately as possible, and to make accurate predictions about the real system's behaviour, as in weather forecasts for example. Predictive models can also be used to reconstruct behaviour in the past, and could for example be used in reconstructing the spread of language families or of particular instances of language change (e.g., Landsbergen, 2009). Explanatory models, in contrast, aim to increase insight in a phenomenon. Explanatory models are generally much more abstract and further removed from reality than predictive models. The phenomenon under study is not modelled in all its detail, but instead only its essentials are modelled. Crucially, what counts as 'essential' very much depends on the research question, and simplifications that are appropriate for one question can be totally indefensible for another. Good explanatory models, moreover, explain the phenomenon of interest in terms of lower-level phenomena that can, at least in principle, be independently motivated (models that simply reproduce the phenomenon of interest without providing such an explanation are sometimes called *phenomenological models*).

The second important distinction is one based on form, between mathematical and computational models. The distinction is not always strict, but mathematical models tend to be the most abstract and to strip down phenomena to their barest essentials. Typically (but not exclusively), mathematical modelling papers provide both a formalization of a phenomenon (e.g., using matrix algebra, logic, differential equations) and proofs about properties of the formal system. Such proofs are, by definition, universally valid and allow inferences about specific cases (deduction), although the simplifications necessary to arrive at a proof often greatly limit the applicability.

Computational models tend to be much more concrete and complex. Phenomena are formalized in a programming language, and the resulting programs studied experimentally. From different runs with different parameter settings, the modeller tries to infer general properties of the formal system (induction). The programs can be very complex, allowing for models with fewer abstractions but often barring analytic proofs. In some cases, computational models are used to investigate versions of a mathematical model that are too complicated to study analytically (including *numerical models*, that are defined algebraically but studied using numerical methods on the computer).

A third major distinction concerns the validation of models: we distinguish between internal validation and external validation. Internal validation is about demonstrating that the phenomenon of interest indeed follows from the stated assumptions, and mathematical proof provides its most powerful form. This is much harder to achieve with computer models, although extensive testing and systematic exploration of the parameter space of a computational model can lead to a great degree of confidence. External validation is about checking whether the stated and unstated assumptions are supported by empirical evidence, or by the outcome of other, independent models, and whether the model's predictions are confirmed in the real world. As computational models are often formulated in more concrete terms, it tends to be easier to achieve external validation.

In this paper, we are mainly concerned with the external validation of explanatory models, which in all cases requires an interpretative step: explanatory models have, by definition, abstracted away many details of the phenomenon of interests, making it a matter of judgment whether abstractly formulated assumptions and prediction are supported by concrete evidence. In historical research fields such as the origins of language (and likewise, the origins of life or the universe), external validation is further complicated by the fact that there is little direct evidence about which assumptions and predictions are valid. External validation is thus only achievable by *model sequencing*: assumptions and prediction of any particular model are validated mainly by results from other models, and only at various points in a string of models do empirical results come into play.

Moreover, because this field deals with complicated phenomenon for which the appropriate simplifications haven't been established yet, modelling research should employ *model parallelisation*: for any particular phenomenon, researchers should develop multiple formalisations, compare results and relate observed differences to explicit and implicit assumptions embodied in these alternative models. Hence, modellers in language origins research must – much more than is currently practiced – work out relations between different models, whether they stand in sequence or in parallel to each other.

These observations may seem straightforward to many readers, but we find that much modelling work suffers from lack of clarity on these conceptual foundations. In the following section we will further explore our methodological points in the context of models of the evolution of speech, which, whilst home to some great controversies, provides a good case study as it has an abundance of easily interpretable data and a strong modelling tradition.

3 — The evolution of speech and repertoires of phonemes

When studying the evolution of speech in relation to the evolution of language, the focus is usually on the differences between modern humans and the hypothetical latest common ancestor (henceforth, LCA) of humans, chimpanzees and bonobos. The vocal abilities of the LCA are inferred from the abilities that humans, chimpanzees and other apes share or do not share. From such comparisons, it can be derived that the LCA had a repertoire of calls for communicative purposes, and therefore a limited ability to modulate the vocal tract. However, it most likely had a vocal anatomy more comparable to that of chimpanzees and vocal cords comparable to those of chimpanzees and gorillas. The LCA did not, it seems, have modern human's descended larynx, it had less voluntary control over breathing (MacLarnon & Hewitt, 1999) and probably did have supralaryngeal air sacs. As all modern apes only have limited voluntary control over their vocalizations, only learn their vocalizations to a very limited extent and lack internal (combinatorial) structure in their calls, it can be assumed that this was also the case for the LCA.

Modern humans, on the other hand, do have a descended larynx, have voluntary control over speech (but much less so over emotional utterances), and have a large learned repertoire of linguistic utterances. Moreover, those utterances have complex internal structure that is used productively, and there are regularities in the repertoires of speech sounds that humans use (the phonological universals). The challenge for research of the evolution of speech is to give an account of how the modern phenotype evolved from the LCA's phenotype: i.e., how did the descended larynx, voluntary control, vocal learning, combinatorial phonology and phonological universals evolve? A key issue here is to what extent the evolutionary changes should be considered adaptations for language, or to what extent they evolved for other reasons.

Computer models (and to some extent mathematical models) have been used for a long time to investigate such issues – but in the existing literature (as reviewed in de Boer, 2005; de Boer & Fitch, *in press*) there are some striking gaps in the range of topics considered and some disturbing confusions about the role of various models. The most studied topics are the evolution of the vocal tract (Lieberman & Crelin, 1971; Boë *et al.*, 2002; de Boer, 2009) and the emergence of phonological universals (de Boer, 2000b; Oudeyer, 2005; Zuidema & de Boer, 2009); the evolution of voluntary control, vocal learning and combinatoriality have received much less attention in the modelling literature, and the issue of how models of these different aspects fit together has been almost completely ignored.

Starting point for many models of how speech evolved are models of how speech perception and production works in human adults. Surveying the literature, we quickly find that many models that have been developed for the study of human speech are not necessarily directly usable in the study of the evolution of speech. Illustrative examples from modelling the acoustic production of speech are the 3-parameter model (Stevens & House, 1955; Fant, 1960), the couple mass-spring model (Dudgeon, 1970; Ishizaka & Flanagan, 1972) of the vocal cords and the source-filter model of speech production (Fant, 1960). These are simplified, explanatory models of the human vocal tract, the human vocal cords and the (lack of) interaction between the human vocal cords and the vocal tract, respectively.

These models are well established in phonetics, and provide valuable insights in the process of speech production. However, researchers in the evolution of speech cannot simply reuse these models to represent properties of vocal tracts of our evolutionary ancestors or of other species (see the discussion about Riede *et al.*, 2005 in Lieberman, 2006); doing so is misunderstanding the explanatory nature of the existing models, that involved simplifications which were very helpful for understanding speech production but are specific to human adult vocal tracts. It is, in fact, unlikely that ape-like vocal tracts can make the deformations of the vocal tract that are assumed by the 4-tube model, and it is clear that the acoustic effects of supralaryngeal air sacs are not captured by it. It is further unknown whether chimpanzee-like vocal cords work in the same way as human vocal cords, and whether in chimpanzee-like vocalizations the vocal cords can really be considered acoustically independent of the vocal tract. Simplifications made in building these models must thus be re-evaluated in the light of what is known about ape and fossil vocal anatomy.

A second problem with existing models of the evolution of speech anatomy concerns its relation to models of the biological and cultural evolution of communication, i.e., with external validation through model sequencing. Even if we could establish a sequence of vocal tracts, leading from ape-like to human-like shapes in gradual steps, that in itself, although an important step, would not provide an evolutionary explanation. As we and others argued elsewhere (e.g., Parker & Maynard Smith, 1990; Zuidema & de Boer, 2003, 2009), evolutionary explanations must provide a 'path of ever increasing fitness', where every new variant provides a fitness advantage in a population where the previous variant is still common. In the case of vocal tract evolution, it is unclear what the appropriate fitness function is. Existing models tend to assume that it is a simple function of the size of the acoustic space allowed by a particular vocal tract configuration. But fitness due to speech must be a function of how well an individual communicates with others in a population, which in turn depends on the communication system the population uses. However, the relation between the repertoire of speech sounds that emerges in a population and the anatomical and neurocognitive features of individuals is far from trivial.

Models that study the emergence of such repertoires have focused on vowel inventories, and on a role for self-organization in shaping them (Glotin, 1995; Berrah & Laboissière, 1999; de Boer, 2000a; Oudeyer, 2005), given constraints on the vowel space formalised by existing models of vowel perception and production. This group of models is a good example of model parallelization: different models all show the emergence of similar phenomena. They are not a good example of model sequencing, however: although these models have yielded a beautiful connection between empirical data on vowel systems and biophysical constraints, it is clear that they only scratch the surface of the full set of phonological universals: they have, for instance, little to say about consonants, syllable-structure or supra-segmental speech patterns.

Ultimately, the connection between phonology and anatomical and neurocognitive features needs to become clear to allow us to evaluate particular scenarios of the evolution of speech. However, despite the progress in modelling vocal tract evolution and vowel universals, we're still quite far from a model-based understanding of the evolution of speech. In the required sequence of explanatory models we still observe, for a variety of reasons, many gaps.

One reason is that, when addressing these more complex issues, the limits of what is at present possible with computer models are reached quickly. It is then tempting to use high-level abstractions (such as distinctive features, constraints and rule-based phonological explanations). However, making use of such abstractions, which have after all been derived for description of modern human language, and are in general not based on direct observation of neurocognitive mechanisms, incurs the risk of implicitly including the phenomena to be explained in the model – and thus resorting to phenomenological rather than explanatory modelling. For example, from typological studies it is known which consonants are unusual (for example uvular plosive [q]) and which are common (for example velar plosive [k]), but there is no language-independent biophysical and neurocognitive model that reliably predicts which articulations are more difficult to produce than others. Thus research into more complex aspects of speech is not only hampered by the computational complexity of such models, but also by our lack of knowledge about the underlying phenomena.

Likewise, we have no models of the evolution of the vocal cords. Although there are many models for human vocal cords (Dudgeon, 1970; Ishizaka & Flanagan, 1972; Titze, 1973, 1974, 2008) and some models of the interaction between the vocal cords and the vocal tract (Flanagan & Meinhart, 1964; Titze, 2002, 2008) as far as we are aware, no models exist of either chimpanzee vocal cords or of hypothetical ancestral vocal cords. This has undoubtedly to do with the lack of anatomical data (although some has recently been presented Demolin & Delvaux, 2006) but also with the fact that vocal cords (and their interaction with the vocal tract) are much more difficult to model than the acoustics of the vocal tract itself.

Another reason is that in spite of much parallel modelling effort, in some domains no consensus is reached. There is, for example strong controversy in the study of the articulatory abilities of Neanderthals and the role of modern human vocal anatomy (with its descended larynx). In this debate, Lieberman (Lieberman & Crelin, 1971) and Carré et al. (Carré *et al.*, 1995) propose that vocal anatomy has evolved for speech, while Boë et al.

(2002) propose that it has not evolved for speech, because (neural) control is more important. They reach opposite conclusions, even though they use very similar modelling techniques. The debate has lead to a rather heated exchange (Boë *et al.*, 2007; Lieberman, 2007).

Finally, some topics seem to be simply overlooked. For instance, important innovations in the cognitive adaptations for using speech that occurred between the LCA and modern humans have not been addressed by modelling. These include the ability to productively use combinatorial structure of speech and the (related) ability to learn large sets of complex utterances. Such models would be quite complex computationally, but their results might be transferable to other aspects of language, most notably syntax. After all, it has been proposed that the sequential processing and learning that are necessary for using syntax are based on adaptations for the sequential processing and learning mechanisms that are necessary for using combinatorial utterances (Carstairs-McCarthy, 1999).

Given these gaps in our understanding of the evolution of speech, the possibilities for external validation are at present limited and we should guard against over-interpreting modelling results. A case in point is the reception of Nowak *et al.* (1999), who presented an information-theoretic model and a mathematical proof of the conditions for combinatorial coding to have a fitness advantage. This proof is an elegant example of internal validation. The model fits into a larger research program in which a number of proofs of mathematical models related to the evolution of language have been presented in high-profile publications (Nowak & Krakauer, 1999; Nowak *et al.*, 2001, 2002). These models have been interpreted by other researchers as having "...demonstrated the evolvability of the most striking features of language..." (Pinker, 2000). However this confuses internal validation (the models are internally consistent) with external validation (the models correspond to reality). The latter is unfortunately far from established, given the many simplifying assumptions in Nowak et al.'s (Nowak et al., 1999) model, as we have pointed out elsewhere (Zuidema & de Boer, 2009).

4 — Conclusions

There are a number of lessons we would like to be drawn from our analysis of the state-of-the-art of language evolution modelling. First of all, it seems modellers should pay more attention to how their models relate to other models, and how they fit into particular scenarios. Although most papers on modelling the evolution of language do a good job at internal validation and at crediting other researchers' work, authors do not often make explicit which scenario they feel their model fits in and in what way their model provides external validation for other models or how other models provide it for theirs.

Second, we note that there is no lack of models and no lack of data, but there is a rather uneven distribution of modelling effort over relevant questions. It is perhaps not surprising that (as in other fields of scientific inquiry) the majority of papers are concentrated around the easiest questions. Understandable as this is, we have now reached a stage where we should also attempt to tackle the more difficult questions, and consider carefully whether a collection of models together constitute a convincing scenario.

Papers presenting 'verbal', complete scenarios can be very useful in structuring such a research program — even if we agree that one should be careful with papers that present scenarios of complex historical processes such as the evolution of language (it is all too easy to resort to speculation and wishful thinking). Jackendoff (2002) is one of the few authors who provides a rather detailed scenario that may provide a useful framework; the research field would benefit if more authors would provide a sketch of such a scenario with their work and describe how they feel their work and previous work fits in the scenario. This may help to identify the areas of language evolution that are relatively well-understood and well-studied and areas that are still *terra incognita*.

In Jackendoff's theory, a number of major transitions occurred in the evolution of language that correspond to major design features of natural language. He views, among others, an open learned vocabulary, a combinatorial phonological system, a compositional semantics, hierarchical phrase-structure and a system of syntactic categories to convey semantic relations as crucial innovations in the evolution of human language from an ancestral primate communication system. Of those five, only combinatorial phonology and compositional semantics have been addressed by multiple, parallel modelling studies.

Jackendoff's scenario is not the final word on language evolution, of course. It lacks attention for the communicative setting in which language evolved, for the possible selective advantages it offered, and to whom, and for the question why languages are so diverse and continue to change. Evolutionary biology offers many models of the evolution of altruistic traits and communication, but the relation between such models and models of the evolution of other aspects of natural language have not received much attention. Likewise, sociolinguistics and (formal) pragmatics offer many ideas about the function of language and language variation, and about the question who benefits. Also their relation to models of the evolution of language's design features remain underexplored.

From these considerations and those presented earlier in this paper, we can compose a list of key challenges for language evolution modelling that we hope will be addressed in the next few years. We present such a list in table 1; if these challenges — on the evolution of particular traits studied in the traditional subfields of linguistics and the relation to the human phenotype more broadly — are taken up by the field, we should have in a few years several models for each issue *in parallel*, as well as a set of models that *in sequence* really speak to the plausibility of a particular scenario. Only then are we approaching *external validation* of *explanatory models* of language evolution, and is the modelling approach really proving its worth to the language evolution field at large.

Table 1: Key open challenges in language evolution modelling

Phonetics & phonology:

1. Modelling the evolution of the human vocal cords;
2. Modelling the evolution of human-like (combinatorial) phonology: consonants, syllable structure, pitch/formant relation, intonation contours;

Semantics & pragmatics:

3. Modelling the transition from a closed to an open, learned repertoire of signs;
4. Modelling the evolution of duality of patterning: combinatorial phonology with compositional semantics in a unified model;
5. Modelling the evolution of human-like (compositional) semantics: quantifiers, numerals, functional/contentive split, categoricity/vagueness relation, negation;
6. Modelling dialog: how can structured, repeated communicative interactions evolve (as opposed to isolated signals);

Morphosyntax:

7. Modelling the evolution from "flat" utterances of hierarchical phrase-structure, ;
8. Modelling the evolution of word order/rich morphology trade-off;
9. Modelling the evolution of syntactic categories over and above semantic categories;

Language change & sociolinguistics:

10. Modelling the evolution of ongoing linguistic change – why are there no 'sinks' in language change?;

Relation to non-linguistic issues:

11. Language as a green beard – connection between evolution of language and altruism;
12. Language as a mental tool – connection between language and other uniquely human cognitive traits (music, consciousness, reasoning).

References

Berrah, A.-R., & Laboissière, R. (1999). Species: An evolutionary model for the emergence of phonetic structures in an artificial society of speech agents. In D. Floreano, J.-D. Nicoud & F. Mondada (Eds.), *Advances in artificial life, lecture notes in artificial intelligence* (Vol. 1674, pp. 674–678). Berlin: Springer, 1999.

Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics, 30*(3), 465–484.

Boë, L.-J., Heim, J.-L., Honda, K., Maeda, S., Badin, P., & Abry, C. (2007). The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a). *Journal of Phonetics, 35*(4), 564–581.

Carré, R., Lindblom, B., & MacNeilage, P. F. (1995). Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences, Série II, 320*(série IIb), 471–476.

Carstairs-McCarthy, A. (1999). *The origins of complex language: An inquiry in the evolutionary beginnings of sentences, syllables, and truth*. Oxford: Oxford University Press.

de Boer, B. (2000a). Emergence of vowel systems through self-organisation. *AI Communications, 13*, 27–39.

de Boer, B. (2000b). Self organization in vowel systems. *Journal of Phonetics, 28*(4), 441–465.

de Boer, B. (2005). Evolution of speech and its acquisition. *Adaptive Behavior, 13*(4), 281–292.

de Boer, B. (2009). Why women speak better than men (and its significance for evolution). In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 255–265). Oxford: Oxford University Press.

de Boer, B., & Fitch, W. T. (in press). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*.

Demolin, D., & Delvaux, V. (2006). A comparison of the articulatory parameters involved in the production of sounds of bonobos and modern humans. In A. Cangelosi, A. D. M. Smith & K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference (evolang6)* (pp. 67–74). New Jersey: World Scientific.

Dudgeon, D. E. (1970). Two-mass model of the vocal cords. *Journal of the Acoustical Society of America, 48*(1A), 118.

Fant, G. (1960). *Acoustic theory of speech production*. 'sGravenhage: Mouton.

Flanagan, J. L., & Meinhart, D. I. S. (1964). Source-system interaction in the vocal tract. *Journal of the Acoustical Society of America, 36*(10), 2001-2002.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist, second edition*. Maidenhead (UK): Open University Press.

Glotin, H. (1995). *La vie artificielle d'une société de robots parlants: Émergence et changement du code phonétique*. Grenoble: DEA sciences cognitives-Institut National Polytechnique de Grenoble.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell system technical journal, 51*(6), 1233-1268.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Landsbergen, F. (2009). *Cultural evolutionary modeling of patterns in language change: Excercises in evolutionary linguistics*. Utrecht: LOT.

Lieberman, P. H. (2006). Limits on tongue deformation - diana monkey formants and the impossible vocal tract shapes proposed by riede et a. (2005). *Journal of Human Evolution, 50*(2), 219-221.

Lieberman, P. H. (2007). Current views on Neanderthal speech capabilities: A reply to Boë et al. (2002). *Journal of Phonetics, 35*(4), 552–563.

Lieberman, P. H., & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry, 2*, 203–222.

MacLarnon, A., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology, 109*(3), 341–343.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science, 291*(5501), 114–118.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*(6889), 611-617.

Nowak, M. A., & Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences, 96*, 8028–8033.

Nowak, M. A., Krakauer, D., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London, 266*, 2131–2136.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology, 233*(3), 435–449.

Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature, 348*, 27–33.

Pinker, S. (2000). Survival of the clearest. *Nature, 404*, 441–442.

Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production in a non-human primate: Morphological data and a model. *Journal of Human Evolution, 48*(1), 85-96.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America, 27*(3), 484–493.

Titze, I. R. (1973). The human vocal cords: A mathematical model part i. *Phonetica, 28*(3), 129–170.

Titze, I. R. (1974). The human vocal cords: A mathematical model part ii. *Phonetica, 29*(1), 1-21.

Titze, I. R. (2002). Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model. *Journal of the Acoustical Society of America, 111*(1 Pt 1), 367-376.

Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. *Journal of the Acoustical Society of America, 123*(5), 2733-2749.

Zuidema, W., & de Boer, B. (2003). How did we get from there to here in the evolution of language? *Behavioral and Brain Sciences, 26*(6), 694–695.

Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics, 37*(2), 125–144.