MODELLING AND LANGUAGE EVOLUTION: BEYOND FACT-FREE SCIENCE

BART DE BOER

Artificial Intelligence Laboratory, Vrije Universiteit Brussel
1050 Brussels, Belgium

This paper investigates two pitfalls for computer models of language evolution: fact-free science (models looking for a question to answer) and cargo cult science (models that do not tell us anything that we didn't think already). Three successful lines of research related to computational modeling are investigated: mathematical modeling, experimental iterated learning and reconstruction of language history. It is analyzed why this research has had more impact than computer modeling. It is proposed that attention to what questions are relevant and to methodological rigor are important factors to increase impact.

1. Introduction

A year before the first EVOLANG conference, John Maynard Smith (1995) introduced the term "fact free science" in his review of Depew and Weber's (1995) book *Darwinism evolving*:

> "But first I must explain why I have a general feeling of unease when contemplating complex systems dynamics. Its devotees are practicing fact-free science. A fact for them is, at best, the output of a computer simulation: it is rarely a fact about the world."

Although Maynard Smith's unease was about the complex systems approach to the evolution of organisms, his observation is equally applicable to a complex systems approach to the evolution of language.

It must be noted that Maynard Smith was not against fact-free science per se. Certain fact-free science can be very useful, mathematics and computer science being prime examples. Maynard Smith's unease was about whether complex systems science could make a real contribution to biology:

> "My difficulty, then, is that I do not know what observations complex systems dynamics is trying to explain. It is a theory looking for a question to answer."

This is a criticism that is often (implicitly) raised against computer models of language evolution. The first of two issues this paper raises is how computer models of language evolution move beyond fact-free science.

The second issue has to do with what Richard Feynman (1974) in his commencement address at Caltech – incidentally, a year before the *Origins and Evolution of Language and Speech* conference of the New York Academy of Sciences, the proceedings of which have been published as (Harnad *et al.*, 1976) – has called *cargo cult science*:

> "I think the educational and psychological studies I mentioned are examples of what I would like to call Cargo Cult Science. In the South Seas there is a Cargo Cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas – he's the controller – and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things Cargo Cult Science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land."

Feynman identifies a number of issues that may cause a scientific endeavor to become a cargo cult. Most of these issues have to do with lack of methodological rigor that allows scientists to fool themselves. While early modeling of language evolution may have suffered from lack of methodological rigor (as did many other early complex systems modeling efforts) more recent work is much more methodologically rigorous. Papers illustrating models with a single run and containing sloppy descriptions are no longer considered acceptable. Moreover, highly abstract models are nowadays considered less interesting than models based on linguistic facts. Yet, recent computer modeling of language evolution may still suffer from a more subtle problem that Feynman (1974) identified:

> "There is also a more subtle problem. When you have put a lot of ideas together to make an elaborate theory, you want to make sure, when explaining what it fits, that those things it fits are not just the things that gave you the idea for the theory; but that the finished theory makes something else come out right, in addition."

This issue is again one that crops up in computer models of complex systems: if the number of parameters of the model is equal to or larger than the number of degrees of freedom of the data that the model attempts to explain, it is always possible to find a good match through parameter tuning. If furthermore there is some leeway in interpreting the results of the model (as there usually is in the relatively high-level models that tend to be used in the field) the risk of the researchers fooling themselves and purveying cargo cult science becomes even bigger. The second issue that this paper raises is therefore how to avoid cargo cult science in models of language evolution.

These issues will be addressed by analyzing the efforts that have been made in previous EVOLANG-conferences to advance the interaction between computer modeling and experiments, and by analyzing successful papers and lines of research that have combined models and the evolution of language.

2. What makes a good modeling paper?

A number of efforts have been made at different EVOLANG-conferences to address the issue of the (lack of) interaction between modeling on the one hand and experimental and observational work on the other. There has been a discussion on this topic at the 2002 conference in Harvard, a tutorial at the 2004 conference in Leipzig and a workshop at the 2010 conference in Utrecht. Although these events may have helped improve methodological quality of modeling papers[1] (as can be observed by comparing early papers on the topic with more recent ones) and increase awareness of the issues, it does not appear that they have resulted in major papers on modeling language evolution.

Another issue that appears to have had little influence on the quality and impact of papers is the sharing of software and the availability of simulation platforms such as Babel (McIntyre, 1998; Loetzsch *et al.*, 2008) and THSim (Vogt, 2003), even though this issue received ample discussion in all the above-mentioned events. What does appear to have been influential however, is the on-line availability of large corpora of data and the availability of a set of tools that have their origins in the study of genetics. Another influential approach is a tendency for computer modelers to do their own experimental work.

It is instructive to look in some depth at three different lines of research that have resulted in high-impact publications, and to investigate how this work has avoided the problems of fact-free or cargo cult science (or not, of course). These lines are mathematical work about language dynamics, work in the area of ex-

[1] On the other hand: improvement of methodological quality appears to be a trend of the general field of computer modeling of life-like phenomena, probably related to the maturation of the field.

perimental iterated learning and work on reconstructing language history using tools from biology. Of course, other lines of work have been successful as well, but these examples have been selected for their high impact.

2.1. *Mathematical modeling of language dynamics*

Agent-based models exploded on the evolution of language scene starting from the mid 1990's (Christiansen, 1994; Steels, 1995; Oliphant, 1996; Batali, 1998; Kirby, 1998) although comparable work had been going on for some years (Hurford, 1989, 1991; Werner & Dyer, 1991). Much of this work was limited to very specialized journals and conferences. In contrast, a few years later Nowak and colleagues published a number of papers about *mathematical* modeling of language evolution in high profile general journals (Nowak & Krakauer, 1999; Nowak *et al.*, 1999; Nowak *et al.*, 2001). Why was it that the mathematical modeling work had a much higher profile than the agent-based work?

Part of the reason was of course that these papers were co-authored by Nowak, who already had a good reputation and experience with publishing in these journals. However, this begs the question somewhat, because why would a scientist with his reputation choose to focus on mathematical modeling rather than agent-based computational modeling?

Perhaps the higher profile can be understood by looking at the level of fact-free-ness and cargo-cult-ness of the research. One could argue that Nowak's research program was equally fact-free as the agent-based modeling efforts, because it used highly abstract mathematical models. In a sense, it was more fact-free than most of the early agent-based models, as these made many attempts at realistically implementing aspects of human cognition and behavior. However, Nowak's work focused on very specific research questions, whereas the agent-based models was mostly done in the spirit – common to much early artificial life work – of "Let's see if we can recreate this kind of behavior in a computer model." Thus Nowak's work avoided Maynard Smith's objection that complex systems work is a theory looking for a question to answer.

Interestingly, the higher complexity of agent-based models may have led to a more reluctant reception. Many of the design decisions made in building an agent-based model of language are relatively arbitrary; after all, even linguists and cognitive scientists do not agree about how language works and how it is implemented in the brain. As the higher complexity does result in less transparency and an increased number of parameters and design decisions, such models become vulnerable to Feynman's subtle problem of cargo cult science. Does the behavior of the model really lead to an increased understanding of the phenomenon, or would a different, possibly much simpler model give the same results? What exactly is the reason the model behaves as it does? And does the model really tell us something new, or does it only conform the biases that were (unconsciously) put in? As mathematical models are simpler, more transparent and, most importantly, can be analyzed uncontroversially with mathematical tools, they are much less vulnerable to Feynman's subtle problem.

Mathematical analysis also makes it easier to do a sensitivity study. This is a study to identify how the model's behavior changes when certain (combinations of) parameters are manipulated. Of course, different parameter values are generally investigated for agent-based models, but this is almost never done with the same rigor as is possible with mathematical models.

A final problem with agent-based models is that because of their complexity, it is often hard to identify the exact difference between models, or to compare their results directly. Even though nowadays models are usually described in a reimplementable way, there is no strong tradition of making only small variations on models and investigating their effect. This lack of systematicity results in young researchers falling into the same pitfalls again and again[2].

Of course, this is not to say that everybody should stop making agent-based models and start making mathematical models. There are certain things that are very difficult if not impossible to model and analyze with standard mathematical tools. However, we should always keep in mind the above-mentioned reasons why mathematical models avoid being either fact-free or cargo cult science. It is therefore important for agent-based modelers to have familiarity with existing mathematical models.

2.2. *Experimental Iterated Learning*

Another line of work that has sprung from agent based modeling and that has resulted in more high-profile publications is experimental iterated learning (Galantucci, 2005; Fay *et al.*, 2008; Kirby *et al.*, 2008; Smith *et al.*, 2008; Scott-Phillips & Kirby, 2010). In this experimental paradigm, cultural learning is studied in a laboratory setting using protocols that have been adapted from earlier computational models (Smith *et al.*, 2003). This experimental adaptation of a computational model has led to a rapidly expanding body of work.

Its popularity is understandable. First of all, it is a new paradigm in which many questions are still open and unexplored. In addition experimental iterated learning work is relatively easy to do. Although it requires large numbers of

[2] Thanks to Willem Zuidema for drawing my attention to these points.

participants, the participants are not from a specific population (as would be the case if in studying e. g. bilingualism, newly emerging sign languages, or specific types of aphasia) and no complex equipment is needed. Finally, experimental science is the opposite of fact-free science and therefore more acceptable than computer models to the general scientific community.

Does experimental iterated learning avoid being cargo cult science, however? Does it teach us something about real human behavior, or only about highly artificial behavior in an experimental setting? Methodologically, experimental iterated learning is in full development, and there is much to learn about how to correctly investigate cultural evolution experimentally. Nevertheless researchers do take care to avoid the methodological pitfalls identified by Feynman, often through cooperation between evolution of language researchers and cognitive scientists (e. g. Smith & Wonnacott, 2010). Therefore, experimental iterated learning is probably not more or less cargo cult science than ordinary cognitive science, and therefore acceptable to a large scientific community.

Again, this does not mean that we should all switch from modeling work to experimental work. After all, it is still very difficult if not impossible to deduce from the experimental results what the exact underlying cognitive mechanisms are. Therefore an interaction between experiment and modeling is probably the best way to achieve insight. Given that more than a few modelers have made the transition and that the cognitive science community appears to be increasingly interested in iterated learning, this may be a promising avenue for modelers.

Of course, interaction between models and experimental and observational work can be successful in other areas. An example is the evolution of anatomy, in which I have made some contributions (de Boer, 2009, 2010, 2011, 2012). However, an emerging consensus appears to be that the crucial innovations in the evolution of language were cognitive. Moreover, there is a lively debate (Evans & Levinson, 2009) about the nature of these innovation and about the roles of culture and cognition. The interaction between culture and cognition is therefore perhaps the more exciting topic.

2.3. *Biological models of language change*

The final type of high profile research that will be discussed here is the reconstruction of language history using techniques from biology (Dunn *et al.*, 2005; Lieberman *et al.*, 2007; Pagel *et al.*, 2007; Atkinson *et al.*, 2008; Atkinson, 2011; Dunn *et al.*, 2011). In this type of research it is attempted to reconstruct the history of groups of languages using large corpora of data and computer models that have been adapted from biology and genetics. Usually the reconstructions are used to draw some conclusions about the nature of the historical processes involved.

Again, an important factor in all these papers is that they have been co-authored by established researchers – often biologists – with strong reputations, but as in the case of mathematical models, agent-based computer modelers need to ask themselves why these researchers invest in this line of work, rather than in agent based modeling. One reason is that this research is definitely not fact-free science: it makes use of large corpora of data. A second reason is that they all address questions that are central to (historical) linguistics. Another reason is that the computational tools used in this research have an established record of success in biology and genetics. It is clear that these models, when used in the context of biological evolution do not lead to cargo cult science.

However, because the models are so well-established, they are often described very incompletely in the papers that use them. Therefore, it is not clear at all how applicable the assumptions made in these models are to cultural evolution of language and there is a risk that these papers are cargo cult science. Atkinson (2011) is a case in point: a more detailed look at the data on which his paper is based reveals his result may be spurious (Ian Maddieson, *personal communication*).

The weakness of these models is that they make certain assumptions about population behavior. Their strength is that given these assumptions, they result in very clear statistics about possible historical scenarios/evolutionary mechanisms. This is in contrast with most agent-based modeling, where in the best cases only a range of values for different random initializations is given. It would be ideal if the strength of agent-based models – the emergence of population behavior from individual behavior – could be combined with the strength of these biological techniques – clear statistics over the likelihood of different scenarios or mechanisms given observational data.

3. Conclusion

The twin pitfalls of fact-free science and cargo cult science are great risks for computational modeling of language evolution. It is easy to construct a model that does not answer a question and is therefore fact-free science. However, just incorporating facts about language and behavior may lead to cargo cult science because the model may have so many degrees of freedom that it is no longer clear whether it really answers the questions that it was designed to answer.

This paper has investigated three lines of research that show that modeling of language evolution can lead to high profile and high quality work. Fact-free modeling is possible, but it needs to be mathematically rigorous in order to be general – this has been called external validation at an earlier EVOLANG workshop on computer modeling (Zuidema & de Boer, 2010). It is also possible for computer modelers to test and investigate their models by reformulating them in terms of real experiments – it really is not that hard to do experiments – or to apply models to large corpora of existing data. This has been called external validation by Zuidema and de Boer (2010).

In all cases it is important to address questions that are relevant to other researchers (thus avoiding fact-free science) and to be as rigorous as possible (thus avoiding cargo cult science). Concerning rigor it may be useful to take inspiration from the biological models that have been used to reconstruct linguistic history. Finally, it is important to keep in mind that we are trying to answer questions about the world, not about our models. Although modelers are often more at home in the abstract world of models and computational abstractions, this is not the world we want to investigate. If we keep in mind these simple points (although it is far from trivial to apply them in practice) computational modeling work will continue to have impact for the field of language evolution.

Acknowledgements

This work was funded by the NWO vidi grant "Modeling the evolution of speech", grant number 276-75-007.

References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*, 346-349.

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Languages evolve in punctuational bursts. *Science, 319*, 588.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 405–426). Cambridge: Cambridge University Press.

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure.* University of Edinburgh, Edinburgh.

de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *Journal of the Acoustical Society of America, 126*(6), 3329–3343.

de Boer, B. (2010). Investigating the acoustic effect of the descended larynx with articulatory models. *Journal of Phonetics, 38*(4), 679-686.

de Boer, B. (2011). First formant difference for /i/ and /u/: A cross-linguistic study and an explanation. *Journal of Phonetics, 39*(1), 110-114.

de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of Human Evolution, 62*(1), 1-6.

Depew, D. J., & Weber, B. H. (1995). *Darwinism evolving*. Cambridge (MA): MIT Press.

Dunn, M., Greenhill, S. J., Levinsion, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature, 473*(7345), 79-82.

Dunn, M., Terrill, A., Reesink, G., Foley, R. A., & Levinsion, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science, 309*, 2072–2075.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429-492.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society of London B, 363*, 3553–3561.

Feynman, R. P. (1974). Cargo cult science. *Engineering and Science, 37*(7), 10-13.

Galantucci, B. (2005). An experimental study of the emergence of human communication. *Cognitive Science, 29*, 737–767.

Harnad, S. R., Steklis, H. D., & Lancaster, J. (Eds.). (1976). *Origins and evolution of language and speech* (Vol. 280). New York (NY): New York Academy of Sciences.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77*(2), 187–222.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition, 40*(3), 159–201.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 359-383). Cambridge: Cambridge University Press.

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.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature, 449*, 713–716.

Loetzsch, M., Wellens, P., De Beule, J., Bleys, J., & Van Trijp, R. (2008). *The Babel2 manual, AI-memo 01-08*. Brussels: Vrije Universiteit Brussel.

Maynard Smith, J. (1995). Life at the edge of chaos? *The New York Review of Books, 42*(2), 28-30.

McIntyre, A. (1998). Babel: A testbed for research in origins of language. In *COLING '98, proceedings of the 17th international conference on computational linguistics - volume 2*. Stroudsburg, PA: Association for Computational Linguistics.

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

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.

Oliphant, M. (1996). The dilemma of Saussurean communication. *Biosystems, 37*(1–2), 31–38.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout indo-european history. *Nature, 449*, 717-721.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences, 14*, 411-417.

Smith, K., Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2008). Introduction. Cultural transmission and the evolution of human behaviour. *Philosophical Transactions of the Royal Society of London, 363*(1509), 3469–3476.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life, 9*(4), 371–386.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*, 444-449.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life, 2*(3), 319–332.

Vogt, P. (2003). THSim v3.2: The talking heads simulation tool. In W. Banzaf, J. Ziegler, T. Christaller, P. Dittrich & J. T. Kim (Eds.), *Advances in artificial life* (Vol. 2801, pp. 535-544). Heidelberg: Springer Verlag.

Werner, G., & Dyer, M. (1991). Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, J. Framer & S. Rasmussen (Eds.), *Artificial life ii* (pp. 659–687). Redwood City, CA: Addison-Wesley.

Zuidema, W., & de Boer, B. (2010). *Models of language evolution: Does the math add up?* Paper presented at the Models of Language Evolution Workshop, EVOLANG 8, Utrecht, the Netherlands.