

The cultural origins of language

Bas Cornelissen

August 9, 2017

Human language is one of the most intricately structured communication system in the natural world. Over the last decades researchers in various fields have developed the idea that languages are primarily shaped by processes of cultural evolution, and that these processes can account for the structure of language. Computational models play an important role in their arguments. This thesis asks what those models can learn us about cultural language evolution. To that end, the first part of this thesis connects the two main branches of agent-based models, naming games and iterated learning, in a proposed *Bayesian language game*. The game gives a unified view on the field and suggests a characterisation of the behaviour exhibited by the main agent-based models of language evolution. It moreover addresses shortcomings of earlier models. We find lineage-specific languages reflecting the innate biases of the learners. The second part of this thesis aims to compare that behaviour with the evolution of actual language. Numeral systems are argued to be an ideal empirical test case for models of cultural language evolution. We revisit pioneering work on the modelling of the emergence of numeral systems, and discuss some further results.

Communication abounds in the natural world, but few, if any, communication systems are so intricately structured as human language. We can understand the meaning of a sentence that has never been uttered before, because we understand the words it consists of and the way in which they are combined — that is to say that language is by and large *compositional*. The words, morphemes, themselves also have an internal structure: they consist of phonemes, combined in a accordance with a combinatorial, phonological system specific to the language. The semantic and phonological combinatorics together give language what Hockett (1960) dubbed a ‘duality of patterning’ — a design feature not commonly found in communication systems of other species. Most animal communication systems are *holistic*: every vocalisation has one specific function, and no further internal structure (Zuidema 2013). There are, it seems, few interdependencies between vocalisations. But language is quite different: just about any two sentences will have many interdependencies in their phonology, morphology, hierarchical structure, and so on. It is in this sense that language has a distinct *systematicity* (Kirby 2017).

Now, the Big Question is this: where does it come from? And, indeed, why only us?

Well, we find ourselves in good company. Robert Berwick and Noam Chomsky (2017) have been thinking about the same question, and leave us no doubt how we should *not* try to account for the structure of language: “by means of a cultural-biological interaction”. No, “this latter effort fails in all respects”. Some of that work is “trivially inadequate”, or otherwise the “Kirby-type models”, “the Kirby work” and “the Kirby line of research” do of course “not say anything about the initial origin of compositionality”.

Good, good! More than enough reason to write a thesis about this “Kirby type work” — and in passing perhaps even cite more than two of his papers. But let’s leave the polemics there. After all, what greater good do they really serve?

Berwick and Chomsky (2017) raise some fair concerns, for example regarding the testability and empirical validity of the Kirby-type theories. Or regarding their explanatory power: if your ‘agents’ already know, say, context-free grammars, and they ‘evolve’ a compositional language, does your model explain anything? Are you then, if anything, addressing the evolution of universal grammar, or just modelling language change? These are all valid concerns — or so I think, since I address some of them in this thesis, too. But reading Berwick and Chomsky’s paper — I haven’t read the book yet — one starts to understand what it means if two scientific traditions have lost their common ground and have started talking about radically different things, only superficially using the same words.

My apologies if I overwhelmed the reader with this, well, lack of introduction, but I want to get started sooner rather than later. One should know that the Kirby-type theories, in the broad sense, are really the work of a large group of linguists, cognitive scientists, biologists, anthropologists, computer scientists, statisticians, and even physicists, who over the last 30 years or so have tried to turn the question of language evolution upside down. The ‘traditional’ debates centred around the question whether language was an adaptation or not. Does language perhaps add to our reproductive fitness, or was it a key mutation event that catapulted language in the world? In either case, the origin of language was thought to be a fundamentally biological concern. We only had to find out how those brains evolved to accommodate language. Well, the Kirby-type work turned that around: how do languages have to change, in order to fit in our brains (Christiansen and Chater 2016a)? Rather than evolving brains, let’s think about evolving languages.

For this to even make sense, one has to drastically change the, let’s say, Chomskyan conception of language. Language now becomes a complex adaptive system in its own right, that is subject to all kinds of pressures, at multiple different levels and timescales simultaneously (Kirby and Hurford 2002; Christiansen and Chater 2016a; Smith 2014; Kirby 2017; Steels 2016). On a biological level and evolutionary timescale the innate mechanisms that underly human cognition and language need to develop. At an individual level and much shorter timescale, humans acquire language, constrained by what their biological makeup can accommodate. These learning biases influence, on a cultural level and historic timescale, the cultural dynamics of language. Universals that emerge through processes of cultural evolution in turn shape the fitness landscape and indeed, interactions exist within in and across all levels. Rather than looking solely at the biology of language, the question is how it can interact with acquisition and use. The idea developed is that “language has been adapted through cultural transmission

over generations of language users to fit the cognitive biases inherent in the mechanisms used for processing and acquisition” (Christiansen and Chater 2016b, p. 12).

Returning to Berwick and Chomsky (2017), I think they are quite right to point out that this work “does not really tackle questions about the evolution UG, but rather questions about how particular languages change over time, once the faculty of language itself is in place”. I doubt Kirby would disagree. Indeed, this is the entire point. Since if it turns out that particular languages tend to change over time in a somewhat systematic way that *explains* their structure, then the explanatory burden is lifted from the faculty of language. And this is precisely what the Kirby-type theories argue for: that biology does not have to carry the entire explanatory load, but that a fair share of linguistic structure can be explained by a process of *cultural evolution*. As Kirby (2017) concludes, “we expect the language faculty to contain strong constraints only if they are domain general (e.g. arising from general principles of simplicity) and that any domain-specific constraints will be weak.” If anything, the Kirby-type work is in the business of explaining away UG, rather than explaining it.

Now, the evolution of language is a notoriously hard problem — I believe this is where one is supposed to mention the *Société* — but how does one go about it if language is moreover a complex adaptive system? Using mathematical and computational models is one solution. Modelling makes all assumptions absolutely transparent, allows one to verify their coherence and consistency and generate new hypotheses, or even predictions (Jaeger et al. 2009). Accordingly, models have figured prominently in the literature on cultural language evolution. And that is where the habitat of this thesis is to be found. *What can these models of language evolution actually learn us about language evolution?* That is the heart of this thesis, but as such, the question is too open ended. I therefore break up the question in two parts.

1. *What kind of behaviour can we expect from agent-based models of language evolution?* To answer this question, I aim to formulate a model that captures a substantial share of the agent-based modelling tradition, and try to characterise its behaviour. The ‘substantial share’ is easily identified, since the field falls apart in two (strictly separated) traditions. **Chapter ??** introduces the ‘vertical’ tradition around *iterated learning*. This is the Kirby-type work that addresses how vertical transmission between generations can shape language. **Chapter ??** introduces the ‘horizontal’ tradition around *naming games*, which focusses on the self-organising power resulting from local interactions within a generation. Although the traditions are strictly separated, I will argue in **chapter ??** that their models are very similar. To do so, I take inspiration from Bayesian models of iterated learning and propose the *Bayesian language game*. By further changing the population model I can interpolate between both traditions and thus analyse their behaviour in a single unified framework. I moreover argue that it addresses some of the problems left open by Bayesian models of iterated learning.
2. *How does that behaviour relate to actual evolved language?* Once it is clear what kind of behaviour these agent-based models exhibit, one should ask what this learns us about language evolution. But how can one start answering such a question without an empirical test case? In **chapter ??** I therefore argue that numeral systems are a good test case, and explain in some detail what the structures

are one should try to explain. In **chapter ??** I make a first start with simulating the emergence of numeral systems, which largely amounts to revisiting and extending the pioneering work of James Hurford.

The reader might want to note that the summaries at the start of every chapter further flesh out this outline.

The problem of language evolution challenges many disciplinary boundaries. This thesis alone borders at least several branches of linguistics, statistical physics, biology, probability theory and Bayesian (cognitive) modelling. When starting with the current work, I was new to pretty much all of these fields (except perhaps some courses in the latter two fields) and many important results might very well have escaped my attention. But some things I deliberately left out, or only touch on in passing. These include (1) models of biological evolution, (2) evolutionary game theory, (3) genetic algorithms, (4) evolution of UG, (5) empirical studies of transmission chains and (6) experimental semiotics. Neither will I further discuss broader debates, of which the Berwick and Chomsky paper is part. Instead, I try to address issues in the field of agent-based modelling of language evolution on its own terms. Indeed, the reader will notice that focus in the majority of this thesis is decidedly on the models, more so than on possible interpretations thereof. This partly reflects personal interest, but more importantly, I believe interpretations of models should, insofar as possible, be built on a sound understanding of the model themselves. That is what I hope this thesis, if anything, contributes to.

SOURCE CODE AND DATA I have made the source code of all experiments publicly available via bascornelissen.nl/msc-thesis (including all figures and LaTeX files). A reference to the ‘raw’ data from all experiments can also be found there. The captions of many figures contain a code like **BNG06** or **FIG03**. This identifies the experiment or figure in the repository.

NOTATION Throughout the thesis I use several notational conventions. Vectors are written in boldface, as in $\mathbf{x} = (x_1, \dots, x_K)$, and indexed by k if they have length K . Sometimes it will be convenient to abbreviate $x_0 := \sum_{k=1}^K x_k$. We often need to decorate variables with time-indications. The most consistent unambiguous solution, $\mathbf{x}^{(t)}$, $x_k^{(t)}$ etc., often clutters the notation. Therefore, vectors (boldface) simply get their indication in the subscript $(x_1, x_2, \dots, x_t, \dots)$ and I only use the superscript (t) when confusion can arise, as in $x_k^{(t)}$. If X is a random variable with distribution Dist parametrized by λ , then by $X \sim \text{Dist}(\lambda)$ we mean that $p_X(x) = \text{Dist}(x \mid \lambda)$ where the latter is the density (mass) function. We mostly drop the random variables and write $p(x)$ and $p(x \mid z)$ rather than $p_X(x)$ and $p_{X|Z}(x \mid z)$. Normalizing constants are often irrelevant and we write $p(x) \propto f(x)$ to indicate proportionality, i.e. that $p(x) = 1/C \cdot f(x)$ where C does not depend on x . With regard to sets, \mathbb{N} , \mathbb{Z} , \mathbb{R} denote the natural numbers, integers and reals. If A and B are sets, their Cartesian product is $A \times B$ and B^A denotes the set of all functions $f : A \rightarrow B$. ‘Spaces’ typically have a calligraphic character, so x lies in \mathcal{X} and parameters tend to be Greek. Finally, x is nearly always an observable variable (e.g. an utterance), z and unobservable variables (e.g. internal representation of a language), m a meaning and s a signal.

Bibliography

- Berwick, Robert C and Noam Chomsky (2017). “Why only us: Recent questions and answers”. In: *Journal of Neurolinguistics* 43.B, pp. 166–177. DOI: 10 . 1016 / j . jneuroling . 2016 . 12 . 002.
- Christiansen, Morten H. and Nick Chater (2016a). *Creating language. Integrating evolution, acquisition, and processing*. Cambridge, Massachusetts: MIT Press.
- Christiansen, Morten H and Nick Chater (2016b). “The Now-or-Never bottleneck: A fundamental constraint on language”. In: *The Behavioral and brain sciences* 39. DOI: 10 . 1017/S0140525X1500031X.
- Hockett, Charles F. (1960). “The Origin of Speech”. In: *Scientific American* 203.3, pp. 88–96. DOI: 10.1038/scientificamerican0960-88.
- Jaeger, Herbert et al. (2009). “What Can Mathematical, Computational, and Robotic Models Tell Us about the Origins of Syntax?” In: *Biological Foundations and Origin of Syntax*. Ed. by Derek Bickerton and Eörs Szathmáry. The MIT Press, pp. 385–410. DOI: 10.7551/mitpress/9780262013567.003.0018.
- Kirby, Simon (2017). “Culture and biology in the origins of linguistic structure”. In: *Psychonomic Bulletin & Review* 24.1, pp. 118–137. DOI: 10 . 3758 / s13423 – 016 – 1166 – 7.
- Kirby, Simon and James R Hurford (2002). “The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model”. In: *Simulating the Evolution of Language*. London: Springer London, pp. 121–147. DOI: 10 . 1007 / 978 – 1 – 4471 – 0663 – 0_6.
- Smith, Andrew D.M. (2014). “Models of language evolution and change”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 5.3, pp. 281–293. DOI: 10 . 1002 / wcs . 1285.
- Steels, Luc (2016). “Agent-based models for the emergence and evolution of grammar”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1701, p. 20150447. DOI: 10 . 1098 / rstb . 2015 . 0447.
- Tomasello, Michael (1999). *The cultural origins of human cognition*, p. 248.
- Zuidema, Willem (2013). “Language in Nature: On the Evolutionary Roots of a Cultural Phenomenon”. In: *The Language Phenomenon: Human Communication from Milliseconds to Millennia*. Ed. by P.-M. Binder and Kenny Smith. Heidelberg: Springer. Chap. 8, pp. 163–189. DOI: 10 . 1007/978-3-642-36086-2_8.