Building the tower of babble

Natalia Komarova and Marc Hauser

The Language Learning and Evolution Workshop was held at the Institute for Advanced Study, Princeton, NJ, USA, 17–19 May 2001.

Language is an apparent miracle. Children master it with exceptional ease, whilst at the same time struggling to walk, hold a fork, and recognize that others have thoughts and emotions that differ from their own. They perform, with near perfection, mental computations that are virtually impossible for state-of-the-art computers. They grasp the tree-like phrase structure of language even though their parents have never taught them, and most probably couldn't even if they wanted to. And they babble on about the present, past and future, creating imaginary worlds that no one but themselves can see.

'Children (...) perform mental computations that are virtually impossible for state-of-the-art computers.'

Children eventually grow up, and they keep on babbling. Some of these grown-up children descended on the Institute for Advanced Study in Princeton to use their language to better understand its miraculous appearance in the child, as well as its evolution in the species. A key question in all of the approaches presented at this conference was: what are the core computations underlying language evolution and language learning?

Ray Jackendoff kicked off the meeting with a broad sweeping perspective on language evolution, arguing that it evolved by means of an accumulation of gradual innovations[1]. A first link in this chain is a primitive, one-word proto-language, a system that is to some extent shared with other animals. Towards the end of the chain is a final link that provides the capacity for a contemporary, syntactically structured language; this link is clearly missing in animals. A crucial question is therefore: which links in the chain do we share with animals and which evolved uniquely within our species?

Marc Hauser approached the problem by using what is known as the Chomsky hierarchy, a formal approach to language that explores different computational capacities[2]. Based on experiments with captive primates, Hauser showed that one point of evolutionary departure between humans and other animals lies in our capacity for recursion. Dorothy Cheney showed how animals use their vocalizations and the extent to which their calls are word-like[3]. Although non-human primates appear to have a large vocal repertoire, the system for generating sounds is neither generative nor open-ended, and their signals are restricted to objects and events in the present.

For many, language is built out of various computational capacities, and especially, mechanisms specialized for learning. Leslie Valiant has been developing a formal system that aims to relate the problems of logical reasoning to those of statistical learning. Daniel Osherson used formal logic and learning theory to study human judgment and reasoning when evaluating competing information[4].

A more applied approach was taken by Deb Roy, who reported on his progress in building machines that learn to communicate in human-like ways. Roy has designed robots that, based on input from audio–video recording of mothers communicating with their infants, can successfully pick out the words from the mother's speech. Understanding how lexical items are perceived is closely linked with questions of phonology and its physical constraints. Louis Goldstein suggested that utterances emerge as organized patterns of basic units of phonological contrast, or gestures.

The final step in mastering a language is learning its grammar. Kenneth Wexler presented a formal analysis of grammatical errors by children, placing these results in the context of the principles and parameters theory[5]. Results showed that language acquisition unfolds as a consequence of positive examples or input, as opposed to feedback through instruction or correction. Alan Prince and Bruce Tesar argued for a different approach to grammar acquisition, namely the framework of optimality theory[6]. This framework is based on a suite of constraints that are functionally ranked in terms of their importance, and provides a high degree of predictive power with respect to accounting for patterns of language use. Paul Smolensky suggested that the optimality framework could be effectively linked to the theory of evolution by natural selection and to neural network modeling.

Partha Niyogi has developed a mathematical approach to grammar learning which is based on Chomsky's theory of generative linguistics, the theory of learning and dynamical systems. Niyogi outlined a research program for studying historical linguistics in the context of population learning. Charles Yang used Niyogi's framework to explain why children make certain mistakes during learning.

In his study of pidgins and creoles, Salikoko Mufwene has found that language evolution is much like the process of speciation: languages can experience selective advantages or disadvantages, and can compete and adapt. David Lightfoot emphasized that there are different modes of language evolution[7]. Besides the classic gradual Darwinian evolution, languages sometimes change abruptly, and then settle into a period of stasis, in a parallel with the paleontological notion of 'punctuated equilibrium'.

Martin Nowak and Natalia Komarova have integrated ideas from evolutionary biology and current learning theory into a series of mathematical models of language acquisition and evolution[8]. They identified the basic cognitive requirements to develop and maintain a coherent language. Scott Weinstein took this framework one step further by showing how language and learning acquisition devices might co-evolve.

The overall aim of the meeting was to discuss to what extent evolutionary biology and learning theory can shed light

http://tics.trends.com 1364-6613/01/$ – see front matter PII: S1364-6613(00)01768-X

on the nature of human language and its origins. It has become clear that a new field is emerging that makes use of theories of formal linguistics, machine learning and evolutionary biology. As is often the case, mathematics is the 'language' of the new field, and the phenomenological work is its inspiration. But we must be careful, as Stephen Anderson warned, to avoid the pitfalls of unconstrained speculation, which is so deeply tempting whenever the grand questions of 'origins', 'nature' or 'evolution' are considered.

The miracle of language, and its deeply perplexing origins, was eloquently brought home by William Wang, who recounted the tale of Chairman Mao and his farewell greeting to a group of western guests. Perhaps with a bit of poetic license, the guests heard Mao to say: 'I am a lone monk walking through the world with a leaky umbrella…'. Unfortunately, what Mao intended was something quite different: 'I am above the law, defying the powers of heaven'. The two phrases are homonymous in Chinese. Clearly, then, language has not been perfectly designed for communication. But that is part of its mystery, and a reason for our infinite curiosity concerning its origins and development.

References

1 Jackendoff, R.S. (1997) *The Architecture of the Language Faculty*, MIT Press
2 Hauser, M.D. *et al.* (2001) Segmenting a continuous acoustic speech stream: serial learning in cotton-top tamarin monkeys, *Cognition* 78, B53–B64
3 Cheney, D.L. and Seyfarth, R.M. (1992) *How Monkeys See the World: Inside the Mind of Another Species*, University of Chicago Press
4 Osherson, D. *et al.* (1986) *Systems That Learn*, MIT Press
5 Wexler, K. and Culicover, P. (1980) *Formal Principles of Language Acquisition*, MIT Press
6 Tesar, B. and Smolensky, P. (2000) *Learnability in Optimality Theory*, MIT Press
7 Lightfoot, D. (1999) *The Development of Language: Acquisition, Changes and Evolution*, Blackwell
8 Nowak, M.A. *et al.* (2001) Evolution of universal grammar. *Science* 291, 114–118; Nowak, M.A. and Komarova, N.L. (2001) Towards an evolutionary theory of language. *Trends Cognit. Sci.* 5, 288–295

Natalia Komarova
Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA; and Dept of Applied Mathematics, University of Leeds, Leeds, UK LS2 9JT.
e-mail: natalia@ias.edu

Marc Hauser
Dept of Psychology, Harvard University, Cambridge, MA 02138, USA.