Language Sciences 40 (2013) 12–23

Contents lists available at SciVerse ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

# Computer simulation as a scientific approach in evolutionary linguistics

Tao Gong [a,*], Lan Shuai [b]

[a] *Department of Linguistics, University of Hong Kong, Pokfulam Road, Hong Kong*
[b] *Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2699, USA*

ARTICLE INFO

*Article history:*
Received 20 December 2012
Received in revised form 7 April 2013
Accepted 15 April 2013

*Keywords:*
Evolutionary linguistics
Computer simulation
Complex Adaptive System
Self-organization

ABSTRACT

Computer simulation has recently been adopted into the research of language evolution. As a general review to linguists, we discuss in this paper: why we need computer simulation in evolutionary linguistics; what are the primary advantages and inevitable limitations of this approach; and how to transform the abstract theories or hypotheses of language evolution into the physical mechanisms or scenarios in language evolution models. In the end, we point out the necessity of multidisciplinary collaboration between the modelers and linguists for the future development of computer simulation research in evolutionary linguistics.



## 1. Introduction

*Evolutionary linguistics* (MacWhinney, 1999) aims to identify when, where, and how human language (henceforth simply 'language') emerges, changes, and dies out (Ke and Holland, 2006). As a linguistics subfield, it examines not only *language change* (the process whereby phonetic, semantic, lexical or syntactic features of particular language(s) are modified) and *language acquisition* (the process whereby an infant acquires particular language(s) from the environment) (Clark, 2003), but also *language origin* (the process whereby Homo sapiens made the transition from a prelinguistic communication system to a communication system with the sort of languages we use today) (MacWhinney, 1999). Regarding language origin, several questions remain obscure: (a) what constitutes *the language potential* (the capacities by which to master and use any natural language) (Hauser et al., 2002; Jackendoff, 2002; van der Hulst, 2010); (b) whether this potential is specified in the human genome (Chomsky, 1972; Pinker and Bloom, 1990), or it derived from the competences initially not language-specific (Elman et al., 1996); and (c) how this potential triggers *the language universals* (the features or principles of language structure and use that hold across most, but not necessarily all, languages) (Christiansen and Kirby, 2003a).

Unlike the other human products, linguistic behaviors are hard to preserve in fossils (Lieberman, 2006). This fact does not entail that the studies of language evolution are unscientific, instead, it reveals *an interdisciplinary nature* of evolutionary linguistics; in order to better assess language and its evolution, we have to incorporate the knowledge, findings, and approaches from disciplines other than linguistics. Apart from many disciplines (e.g. archaeology and anthropology, animal behaviors, genetics and neuroscience, and many others) that have shed light on language evolution, computer simulation has now become a new direction in evolutionary linguistics, after James Hurford (1989) designed a model showing that a coordinated signaling system could be formed via iterated communications. The past two decades have witnessed the appearance of many computer models on language evolution (e.g. Hurford et al., 1998; Knight et al., 2000; Briscoe, 2002; Cangelosi and

* Corresponding author. Fax: +852 25464943.
*E-mail addresses:* gtojty@gmail.com (T. Gong), susan.shuai@gmail.com (L. Shuai).

0388-0001/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.langsci.2013.04.002

Parisi, 2002; Christiansen and Kirby, 2003b; Bickerton and Szathmáry, 2009), and many interested linguists have started to accept and try this approach in their research.

In evolutionary linguistics, *computer simulation* can be viewed as the 'operational' hypotheses or theories expressed in terms of computer programs (Parisi and Mirolli, 2007), and the simulation results obtained by running these programs become the empirical predictions derived from those incorporated hypotheses or theories (Cangelosi and Parisi, 2002). Computer models can verify the available theories, suggest some new perspectives, and ask more focused questions (Jaeger et al., 2009). By tracing the dynamics of language evolution or simulating the individual behaviors during linguistic interactions, computer models can evaluate whether a particular, minimal set of assumptions can parsimoniously generate the essential aspects of language. They can also inspire some reconsideration of the importance of certain factors in language evolution (Lyon et al., 2007). On both aspects, computer simulation helps transform linguistics from a descriptive science into an explanatory one (Mareschal and Thomas, 2006).

As many disciplines have undergone significant development in terms of theories, methods, and terminology, it has been increasingly hard for the modern academics to remain sufficiently knowledgeable about various fields and to conduct some interdisciplinary studies, yet both of these aspects are essential for scientific research. In order to bridge the disciplinary gap, systematic reviews of computer simulation to evolutionary as well as general linguists have now become very necessary. Noting that most available reviews targeted primarily on scholars in artificial intelligence or evolutionary computation (e.g. Perfors, 2002; Wagner et al., 2003; Lyon et al., 2007), in this paper, we offer linguists a general picture of computer simulation in evolutionary linguistics. Since the primary goal of simulation lies in transforming the abstract theories into the physical mechanisms and evaluating these theories based on the simulation results, we focus on describing how such transformation is achieved at different stages of designing and implementing a computer model of language evolution.

## 2. Why we need computer simulation in evolutionary linguistics?

### 2.1. Overcome the time limit

Many quantitative methods from information science and bioinformatics have helped the historical and comparative linguists to evaluate the universality and diversity of language (e.g. Evans and Levinson, 2009) and to reconstruct the evolution of linguistic components (lexical forms or word order, etc.) and the origins of some language families (e.g. Atkinson, 2011; Dunn et al., 2011; Bouckaert et al., 2012; Levinson and Gray, 2012). Nonetheless, due to lacking the direct records tracing language development in its prehistoric states, the current linguistic approaches cannot go further back than the earliest records of written languages, which emerged much later than protolanguage. Therefore, most available linguistic explorations of language origin face a time limit, beyond which little information is available (Trask, 1996).

Researchers have proposed several ways in order to overcome this time limit. For example, some linguists (e.g. Bickerton, 1990) advocate using *the ontogeny recapitulating phylogeny analogy* (a term borrowed from biology, viewing how a child learns a language and how this learning led from no language to the kind of languages we use today as twin questions) to study language origin via language acquisition. However, the ontogeny (acquisition) and the phylogeny (origin) of language are distinct: language acquisition focuses on the individual learning mechanisms (Steels, 2005), and this process occurs primarily in a microhistory timescale (Wang, 1991), during which the linguistic input to a child comes from a presumably full-fledged language; however, language origin concerns both the individual learning and the population or social dynamics (Steels, 2005), and this process takes place mainly in a macrohistory timescale (Wang, 1991), during which both the language components and language environments keep changing. In addition, some neuroscience studies have shown that the changes in brain size and structural or functional connectivity are correlated across species, which makes it increasingly difficult to evaluate whether the ontogenetic maturation actually repeats the phylogeny particularly in humans (Fedor et al., 2009). Therefore, applying the ontogeny recapitulating phylogeny analogy in evolutionary linguistics has to be cautious.

Another way to overcome the time limit is to compare learned, culturally varied behaviors of humans with those of nonhuman animals, especially primates (Hauser et al., 2002; Oller and Griebel, 2004; Washburn, 2007; Fitch, 2010). Such comparison gives us a sense of the behavioral or cognitive options that early hominins would have had (Stanford, 2006; Herrmann et al., 2007; Hurford, 2007; Hilliard and White, 2009). However, no matter how subtle these precursors might be in other species, the communication systems of nonhuman species never exhibit the complexity, creativity, and richness as in language (Greenspan and Shanker, 2004; Oller and Griebel, 2008). Therefore, apart from the comparative evidence, we need other testing ground to evaluate whether and how those precursors eventually led to the corresponding language processing abilities.

A third way to overcome the time limit is to compare the normal with the language-impaired individuals in particular tasks that highlight various types of linguistic functions. Based on the genetic and neuroimaging techniques, we have obtained many findings about the neural and genetic bases of language-related abilities, such as the FOXP2 gene identified from the affected members in the KE family, who had a mutated FOXP2 gene compared with the normal people and showed many linguistic deficits (Fisher et al., 1998). The follow-up study on the selective sweep of this gene in the human lineage has revealed a rough time span for language origin (Enard et al., 2002). In addition, this approach has also led to many insightful understandings of the critical cortical areas for certain linguistic functions in the human brain and the close correlations be-