Journal of Phonetics 43 (2014) 57–68

Contents lists available at ScienceDirect

Journal of Phonetics

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

Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals

Tessa Verhoef[a,*], Simon Kirby[b], Bart de Boer[c]

[a] ACLC, University of Amsterdam, Spuistraat 210, Amsterdam, 1012 VT, The Netherlands
[b] Language Evolution and Computation Research Unit, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, UK
[c] Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 12 December 2012
Received in revised form
31 January 2014
Accepted 9 February 2014
Available online 5 March 2014

ABSTRACT

Human speech has combinatorial structure, but it is still unclear how this type of organization emerged in the course of language evolution. There are two positions in the debate about the evolution of combinatorial structure: one stresses the importance of distinctiveness, while the other stresses economy and efficient reuse of building blocks. Different sources of evidence can be used to investigate the origins of combinatorial structure, such as emerging sign languages, animal communication systems, analysis of modern language and computer simulations but each source has its problems. In this article it is demonstrated that a novel empirical method from the field of language evolution can help to gain insight into the emergence of phonological combinatorial organization. This method, experimental iterated learning, allows investigating cultural evolution and the development of structure over time with human participants. We present data from an experiment in which combinatorial structure emerges in artificial whistled languages. We show that our experiment can give insight into the role of distinctiveness and reuse of building blocks and how they interact. We argue that experimental iterated learning offers a valuable new tool for investigating questions on evolutionary phonology and phonetics.

1. Introduction

A salient difference between human language and vocalizations of other apes is that language has combinatorial structure: a small set of building blocks is combined into an unlimited set of utterances (Hockett, 1960). There is a debate about whether combinatorial structure and any other property of human language is mainly due to cognitive adaptations – perhaps uniquely human – (e.g. Baker, 2003; Chomsky, 2007), or whether it is the result of cultural evolution under functional pressures related to learning and communication (Christiansen & Chater, 2008; Evans & Levinson, 2009).

This seems to be related to the increased interest in the evolution of language, for two reasons. The first reason is that understanding biological evolution of language requires identifying which cognitive changes took place in human evolution since the latest common ancestor of apes and humans (Fitch, 2010; Hauser, Chomsky, & Fitch, 2002). Did new, uniquely human language-specific cognitive mechanisms evolve, or is language based on pre-existing cognitive mechanisms? The second reason is that the study of language evolution has shown that there is not necessarily a direct mapping between properties of languages and human cognitive mechanisms. Such a direct mapping was often assumed in early structuralist (e.g. Jakobson & Halle, 1956) and generative work (Chomsky & Halle, 1968), and in early work on language universals (Greenberg, 1963). However, modern work in language evolution, which looks at the interaction between individual learning, cultural evolution and biological (genetic) evolution has shown that this mapping need not be direct. This first became clear through the use of computer models that simulate the interaction between individual learning behavior and cultural processes in the emergence of structure in language (Christiansen & Chater, 2008; Kirby, 2002; Steels, 1997). More recently, mathematical proofs have established more precisely under which circumstances cultural evolution reveals learning biases directly (Griffiths and Kalish, 2007) and under which circumstances there is a less direct mapping (Kirby, Dowman, & Griffiths, 2007).

It is thus very difficult to establish whether properties of language are mainly due to (language-specific) cognitive mechanisms or to cultural processes. Different researchers have very different positions in this debate, ranging from those who feel there are strong language-specific mechanisms (e.g. Baker, 2003; Chomsky, 2007) that can have biologically evolved (Pinker & Bloom, 1990) to those who feel that there are no strong language-specific mechanisms (Evans & Levinson, 2009; Tomasello, 2003, 2009) and that such mechanisms are unlikely to evolve at all (Christiansen

* Corresponding author. *Present address*: Center for Research in Language, University of California, San Diego, 9500 Gilman dr., La Jolla, CA 92037, USA. Tel.: +1 858 534 3625.
E-mail addresses: tverhoef@ucsd.edu (T. Verhoef), simon@ling.ed.ac.uk (S. Kirby), bart.de.boer@ai.vub.ac.be (B. de Boer).

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2014.02.005

& Chater, 2008; Thompson, Smith, & Kirby, 2012). The seemingly radically opposite positions can be explained by differing interpretations of what counts as language-specific mechanisms and by the complexity of the interaction between learning, cultural evolution and biological evolution.

Recently, experimental researchers have started to test cultural transmission and language evolution in laboratory conditions (Scott-Phillips & Kirby, 2010). One of these new empirical approaches, *experimental iterated learning* (Griffiths, Kalish, & Lewandowsky, 2008; Kirby, Cornish, & Smith, 2008; Reali & Griffiths, 2009; Smith & Wonnacott, 2010, and reviewed in more detail in Section 3), was the inspiration for the method that was used here to investigate the emergence of phonological combinatorial structure. In the organization of human speech, a small set of building blocks is combined into an unlimited set of utterances (Hockett, 1960). We show that the complex interactions between individual learning mechanisms and cultural processes involved in the formation of this structure can be investigated with the experimental iterated learning paradigm. This paradigm has already been used to investigate emergence of structure in language, but so far, has mostly focused on either aspects of language that can be represented in a discrete, symbolic way (morphology and compositional syntax, for instance Kirby et al., 2008; Smith & Wonnacott, 2010) or on the emergence of graphical symbols (del Giudice, Kirby, & Padden, 2010). In this paper we present results of an experiment (Verhoef, 2012; Verhoef, Kirby, & Padden, 2011) as an example of how this method can be applied to continuous, acoustic stimuli and argue that this creates a platform that is valuable for studies into the evolution of phonetics and phonology. In order to exclude pre-existing linguistic biases as much as possible, we have used non-speech signals. These signals, although less complex than real speech, are continuous and acoustic.

As will be discussed in more detail in Section 2, two perhaps partly conflicting accounts of the emergence of combinatorial structure in speech exist. One stresses the importance of acoustic distinctiveness (de Boer, 2000; Zuidema & de Boer, 2009; Hockett, 1960; Liljencrants & Lindblom, 1972; Nowak, Krakauer, & Dress, 1999) and relies on cultural mechanisms to explain structure. The other stresses the importance of economy (Clements, 2003; Martinet, 1949; Ohala, 1980) and relies more on cognitive mechanisms. In Sections 4–6 we show that experimental iterated learning can give insight into the role these processes play and how they interact when the evolution of structure in continuous signals is simulated with human participants.

2. Emergence of combinatorial structure

While discussing one of the design features of language, ‘Duality of Patterning’, Hockett (1960) proposed a possible way in which the combinatorial structure of speech could have emerged. According to him, a growing vocabulary increased the need for combinatorial structure and drove its emergence. The signal space limits the number of holistic signals that can be distinguished. When the number of meaningful elements that need to be expressed increases, signals become closer to their neighbors in that space and discriminability decreases. This problem can be solved by combining a smaller number of elements into a larger repertoire of signals. Hockett's account therefore suggests that (phonological) structure emerged out of pressures for expressivity and discriminability.

Optimization for distinctiveness has been studied with the use of computer models. Liljencrants and Lindblom (1972) were the first to show that optimization for acoustic discriminability results in realistic (small) vowel systems. More recently, agent-based simulations have been used to demonstrate that optimally dispersed discrete and combinatorial systems can emerge without explicit optimization. The optimization is the result of self-organization under pressures of good communication and learnability (de Boer, 2000; Oudeyer, 2006). Building on these results, Zuidema and de Boer (2009) have shown that optimization for distinctiveness may also result in combinatorial structure, while de Boer and Zuidema (2010) have shown that, as in the case of vowel systems, optimization can result from self-organization in a population. These results appear to conform to Hockett's (1960) proposal in which he explains the emergence of combinatorial structure on the basis of pressures from signal distinctiveness and vocabulary expansion.

However, it has been suggested that a focus on optimization for distinctiveness alone may not be enough. Liljencrants and Lindblom (1972) already observed that larger vowel systems are less well explained through dispersion. Looking at consonant inventories, Ohala (1980) suggested that the organization in speech sounds seems to follow a principle of ‘Maximal use of Available distinctive features’. This was based on the observation that features used in the inventories are efficiently recombined and maximally reused, which does not always result in more dispersion. Berrah and Laboissière (1997), using computer models, have shown that applying this idea to vowel systems leads to improved prediction of larger systems, for example those using length, nasalization or front rounding as extra features. Clements (2003), when referring to the theory of *feature economy*, expressed similar ideas about the importance of re-using features: ‘languages tend to maximize the combinatory possibilities of features across the inventory of speech sounds: features used once in a system tend to be used again’. Both Ohala's and Clements' principles focus on the efficient reuse of distinct features to make up a system of sounds. A related proposal was made involving speech gestures by Maddieson (1995) who described structure in speech in terms of articulatory gestures and efficient reuse of places of articulation.

These theories that are based on economy principles may differ in the assumptions about whether the basic elements for reuse are abstract features or physical gestures, but what these theories have in common is that they all propose a rather different approach compared to the dispersal models mentioned before. A general cognitive tendency towards efficient representation of information appears to be assumed. This implies a more direct involvement of language learning and cognitive biases in explaining combinatorial structure.

A possible source of evidence that can shed light on the question of whether combinatorial structure was the result of pressures for discriminability when vocabularies expanded, could be the study of a newly emerging sign language. Established sign languages have phonological structure that uses discreteness and recombination just as spoken languages do (Corina & Sandler, 1993; Sandler, 1994a, 1994b). Al-Sayyid Bedouin Sign Language (ABSL) is a sign language that is only a few generations old and in which the emergence of phonological structure is currently being observed (Israel & Sandler, 2011; Sandler, Aronoff, Meir, & Padden, 2011). Even though it is a fully functional and expressive sign language with a large vocabulary and a rich, open-ended meaning space, it appears that its combinatorial structure is less discrete than those of established sign languages (Sandler et al., 2011). This example shows that a growing vocabulary can be maintained without combinatorial structure.

A different source of evidence that weakens the assumption of dependence between combinatorial structure and complex semantics is the study of song systems of for instance birds and whales (Doupe & Kuhl, 1999; Payne & McVay, 1971). Here we find systems of predictable patterns similar to combinatorial structure in human language, with absence of complex semantics and mostly relatively small repertoires of signals. This shows that combinatorial structure can exist without apparent pressure from a growing vocabulary. Combined with the case of ABSL, the connection between combinatorial structure and growing repertoires of meanings is weakened in both directions: large repertoires of meanings exist without combinatorial structure and combinatorial structure exists without large repertoires of meanings.

In addition, the existence of pseudo words in human language suggests independence as well (Fitch, 2010). There are many more possible words that are well formed in a language than are actually used in the vocabulary, which is puzzling if one assumes that vocabulary drove the expansion of possible words. In summary, many sources can be used to answer questions about the emergence of combinatorial structure, but the results so far are inconclusive and it could therefore be helpful to investigate this question experimentally.