Efficient coding in speech sounds: Cultural evolution and the emergence of structure in artificial languages

Verhoef, T.

Citation for published version (APA):
Verhoef, T. (2013). Efficient coding in speech sounds: Cultural evolution and the emergence of structure in artificial languages
Introduction

No species on this planet other than mankind uses a system for communication as intricate as human language. How did we get from the chirps, howls and calls of monkeys and apes to the complex and sophisticated signal of human speech? What is the origin of this unique form of communication? This is a question that has fascinated researchers since long ago and the work presented in this thesis belongs to the scientific field in which it is studied. The specific area addressed here is cultural evolution and the emergence of structure in sound systems used for speech. This first chapter sketches the context for this thesis and provides an overview of what can be expected to be found in the following chapters.

1.1 Evolution of language

Language is one of the most important features that separate us humans from the rest of the animal kingdom. Yet, we do not have a clear picture of how it arose and what it is exactly that gives humans the ability to use it. Until relatively recently it was hard to approach questions on language evolution without resorting to speculation because there is not much tangible evidence to be found in this area (Müller, 1861). Speech is a rapidly fading signal and we do not have recordings of human’s first utterances. Written language is a relatively recent phenomenon, so the history of writing systems will not help us to study the origins of spoken language. Fossil records may reveal data about the evolution of the human vocal tract and biological adaptations such as the descended larynx and the loss of air sacs can be shown to aid the production of speech (de Boer, 2012; Fitch, 2000), but there are other functions that could have driven the evolution of these adaptations as well (de Boer, 2009; Fitch, 2000). We can therefore not be sure they evolved especially for speech. So for a long time the data that could be used for developing theories about the evolution of language was limited. The results of early surmises received nicknames such as ‘the bow-wow theory’ for the idea that the first words were imitations of sounds such as animal vocalisations or other sounds from the environment and ‘the pooh-pooh theory’ for ideas assuming that the first words were the sounds people make when expressing emotions such as fear or joy (Müller, 1861).
1. Introduction

It was at first assumed that there had to be a special innate language module unique to humans (Chomsky, 1976; Piattelli-Palmarini, 1989; Pinker and Bloom, 1990). How else could we explain why children acquire their language so easily and reliably, while other species did not seem to have these abilities? It was assumed that language “belongs more to the study of human biology than human culture” (Pinker and Bloom, 1990). Whether such a specialised language module evolved as a biological adaptation through natural selection (Pinker and Bloom, 1990) or by accident (Piattelli-Palmarini, 1989) is a matter of debate between proponents of this view, but they share the idea that humans are born with a special language faculty and that language should be studied as if it is a biological organ like any other in the human body (Chomsky, 1976). Until now, researchers have not been able to identify such an organ or module unique to humans that may account for our linguistic abilities. Studies involving the human brain (Deacon, 2009; Fisher and Marcus, 2006) as well as investigations into molecular genetics (Fisher and Marcus, 2006) suggest that language most likely arose in response to the reorganisation of many different systems that humans share with their ancestors and evidence for the existence of a single special module is therefore lacking.

In 1976, a conference was organised by Stevan Harnad and others (Harnad et al., 1976), in which researchers from many different disciplines were brought together to discuss issues on ‘the origins and evolution of language and speech’. It was recognised that the speculative nature of research into this topic could only be overcome by taking a multi-disciplinary approach. This meeting involved sessions on a variety of topics including perception and cognition in humans as well as non-humans to explore the basis of language and intelligence; artificial intelligence to see to what extent machines can copy human (linguistic) abilities; comparative biological research to learn from communicative behaviour in animals; neuroscience to find out how the brain is involved and paleobiology to study what our ancestor’s use of symbols and tool making can reveal. This could have been the start of a fruitful collaborative programme but it was not until 1996 before the field really took shape and the first EvoLang conference, an international and interdisciplinary conference on the evolution of language, was organised in Edinburgh. This became a series of biannual meetings with contributions from the different disciplines that were represented at the 1976 meeting, as well as the introduction of other modern and empirical methods. Geneticists for instance now search for unique genes that may explain human linguistic behaviour; computer modellers analyse and simulate evolutionary scenarios and interactions between individuals; linguists head into the field and study newly emerging (sign) languages; cognitive scientists and psychologists conduct experiments in which human participants learn or invent artificial languages and so on. In sum, there is now a wealth of data available and the development of suitable methods for studying language evolution is growing.
As modern data is accumulating, it becomes progressively clear that there are viable alternatives to the theory that assumes an innate language faculty. Computational simulations, laboratory experiments and other methods have yielded results (discussed in more detail in chapter 2) that are in line with the suggestion that language is shaped by the brain (Christiansen and Chater, 2008) and that not only biological evolution but also cultural evolution can explain the emergence of linguistic structure (Deacon, 1997; Kirby and Hurford, 2002; Kirby et al., 2004). As Deacon (1997) wrote in The Symbolic Species, “The structure of a language is under intense selection because in its reproduction from generation to generation, it must pass through a narrow bottleneck: children’s minds” (Deacon, 1997, p.110). An idea that has become increasingly popular is that language is a system that culturally evolves in a way that can to some extent be compared to the process of natural selection in biological evolution. As Kirby and Hurford (2002), Kirby (2002), Zuidema (2003) and others demonstrated, transmission of a language from generation to generation can make the language more learnable and more structured. Each time the language is passed on it is filtered by the brains that are learning it. It is impossible for a learner to be exposed to every possible utterance in a language because languages are open-ended systems, so all learners have to form their own hypotheses about the structure of the system. Only those structures that can be inferred will be reproduced and therefore there is selection on learnable structures. The structures that are easily transmitted pass through the bottleneck and remain part of the language. In addition, typological data on many different languages revealed that languages around the world are much more diverse than originally thought, which makes the assumption of highly specialised biological adaptations even more implausible (Evans and Levinson, 2009).

Traditionally, it was assumed that the nature of language could be unravelled by studying individual language users (Pinker and Bloom, 1990) and by identifying the universal structures found in languages around the world as an indication of what is encoded innately. The newer ideas mentioned in the previous paragraph imply that language should be viewed as a complex adaptive dynamical system (Beckner et al., 2009; Brighton and Kirby, 2001; Kirby, 2002; Steels, 1997b). From this point of view, it follows that it would be naïve to study language as a system independent of culture and context. Language is the result of many systems that all influence each other in complex ways. The characteristics of the linguistic utterances produced by the individual is only a very small part of this system. Language is a complex system because it emerges as a result of interactions between multiple individuals. At the population (macro) level, language is more than a sum of all the utterances produced at the individual (micro) level.

As mentioned before, languages are transmitted over generations and are dynamic; they change over time and adapt to the selective pressures created by constraints on learning, interaction and population
structure. The role of the first has been explained in the previous paragraph. The second, interaction between multiple individuals, can cause the emergence of a conventionalised shared system when individuals align their behaviour. This has been demonstrated with the use of computer simulations (e.g. de Boer, 2000; Steels, 1997b; Zuidema and de Boer, 2009) and can also be observed in real languages when communities with no shared language start to share a living environment and a new language results from the interactions between members of the communities (Bakker, 1994). Third, population structure and social factors have been found to be related to language structure and complexity. Lupyan and Dale (2010) used statistical analysis techniques on a large sample of languages and found that factors such as population size and contact with other languages could predict certain characteristics of the language structure. Languages spoken by more people tend to be less complex. Wray and Grace (2007) similarly proposed that the pattern of language use may be of influence on the structure. In small, cohesive populations where everyone knows each other and the language is rarely used to talk with strangers, the content of what is talked about is expected to be predictable, for instance because roughly the same knowledge is shared by all members of the population. In contrast, in larger populations in which a greater proportion of conversations is held with strangers, the content cannot always be so easily predicted on the basis of a shared cultural background and context. Wray and Grace (2007) therefore argue that languages that are more often used for talking with strangers are more likely to develop towards having predictable structures and being transparent. Languages of isolated populations on the other hand are expected to be more opaque. In summary, different sources of data all indicate that influences of social and cultural factors should be taken into account in the study of language evolution.

The research presented in this thesis builds on the interdisciplinary work that views language as a complex adaptive dynamical system. Two of the relatively novel methods that have been developed in the field of language evolution, experiments with human participants and computer simulations, are central to this thesis. Computer simulations provide an excellent tool for investigating evolutionary processes and help shed light on the non-trivial relation between micro-level behaviours of individuals and macro-level structures in linguistic systems. The outcome of the complex interactions between these levels are hard to predict and simulations may lead to surprising new insights. However, assumptions and simplifications need to be made when creating computer models, which means that computer agent speakers do not necessarily resemble real speakers in every aspect, especially in terms of their cognitive power. Therefore, it is important to incorporate real human participants in research about language evolution as well. The method of experimental iterated learning (Kirby et al., 2008) has proven to be very suitable for this and
1.2 Sound organisation

This thesis focuses on one particular characteristic of human language: the organisation of speech sounds. Speech sounds are part of a discrete repertoire of primitives that are organised in combinatorial structures. Where does this kind of structure come from? Compared to other species, humans are generally able to produce a larger range of different sounds and these sounds are organised and combined more elaborately (Hurford, 2011). In addition, humans are able to speak about an enormously rich set of meanings. Animal communication systems show very little semantics and complex, acquired meanings are rare. Some bird species use their song to convey fitness in the competition for mating and territory (Doupe and Kuhl, 1999), bottlenose dolphins refer to individuals within a group and maintain group cohesion by producing distinct signature whistles (Janik and Slater, 1998) and there are monkeys that associate different alarm calls with the threats of different predators (Zuberbühler, 2000), but none of these examples even remotely resemble the rich compositional semantics human language has (Hurford, 2011).

Unlike complex semantics, combinatorial structure is not something that is strictly unique to human language. At the level of (phonological) combinatorial structure, there are clear analogous structures in animal song systems. As Hurford (2011) shows with a detailed analysis of such systems: “Apart from the obvious lack of compositional, and referential, semantics, these songs are not qualitatively, but only quantitatively, different in their basic combinatorial structure.” (Hurford, 2011, p. 24). Examples are the structures found in the songs of birds, whales and non-human primates. Certain species of birds that typically acquire their song when growing up, such as the white-crowned sparrow or the zebra finch, produce songs that can be analysed into hierarchical structures in which basic building blocks (notes) are combined into syllables and syllables are organised into larger motifs (Doupe and Kuhl, 1999). A similar type of predictive and hierarchical pattern is found in the songs of humpback whales (Payne and Mcvay, 1971). Payne and Mcvay (1971) describe how the structure of the songs of these whales spans a much
1. Introduction

longer duration than those of birds, but also consists of basic sound ‘units’ that are combined into larger constructs called themes, phrases, songs and song sessions. These whale songs have been analysed by Suzuki et al. (2006) with a computerised unit classifier and measures based on information theory to provide additional evidence for the presence of hierarchical combinatorial structure. Within the primate lineage, gibbons are known to produce complex songs as well (Clarke et al., 2006). They use a set of basic vocal units to form complex phrases and songs and individuals engage in ‘duets’ by taking turns in a systematic way. These examples suggest that perhaps very general cognitive structures are involved in processing and dealing with combinatorial structure of this type, and that no language-specific biological adaptations need to be assumed for explaining the emergence of such structure.

The evolution of complex sound systems for speech is investigated here within a framework that recognises the importance of cultural evolution. In this thesis I study how sound systems emerge, develop and are preserved when being transmitted over generations. One of the main aims is to investigate to what extent structures in sound systems for speech can be explained as the result of general cognitive biases and the process of cultural transmission. Several issues are addressed: the influence of cultural transmission on the emergence of phonological structure; the role of referentiality and semantics in such emergence and the way population structure affects the preservation of emerged systems.

1.3 Overview

The next chapter provides a background on a selection of areas in the field of language evolution that are relevant for the main subjects of this thesis. It provides a brief general overview of different views on the nature of human protolanguage, reviews current hypotheses and ideas that have been proposed to explain the emergence of combinatorial structure, summarises different experimental methods that have been used in the field and links these to ideas about efficient coding in the brain. Chapter 3 subsequently describes a first experiment in which the cultural emergence of combinatorial structure is studied. Chapter 4 then describes a more elaborate experimental study in which combinatorial structure emerges through cultural transmission in artificial whistled languages. Chapter 5 describes experiments disguised as online games that were conducted to further analyse the data from chapter 4. In chapter 6 results from a follow-up experiment with artificial whistled languages is described in which semantics is added. Chapter 7 is about a computational model that was used to study the preservation of emerged vowel systems in populations of interacting computer individuals. The thesis ends with a general overall discussion and conclusion.