Verhoef, Tessa, de Boer, Bart & Kirby, Simon (2012) Holistic or Synthetic Protolanguage: Evidence from Iterated Learning of Whistled Signals, In: Scott-Phillips, T., Tamariz, M., Cartmill, E. A. & Hurford, J. R. (Eds.) *The Evolution of Language: Proceedings of the 9th international conference (EVOLANG9)*, Hackensack, NJ: World Scientific, pp. 368–375

HOLISTIC OR SYNTHETIC PROTOLANGUAGE: EVIDENCE FROM ITERATED LEARNING OF WHISTLED SIGNALS

TESSA VERHOEF & BART DE BOER

ACLC, University of Amsterdam, Spuistraat 210,
Amsterdam, 1012 VT, Netherlands
t.verhoef@uva.nl, b.g.deboer@uva.nl

SIMON KIRBY

LEC, University of Edinburgh
Edinburgh EH8 9AD, UK
simon@ling.ed.ac.uk

Many arguments have been proposed in favor of and against the idea of protolanguage as a set of holistic utterances that were later segmented into words. This paper presents data from a human iterated learning experiment which corroborates arguments in favor of holistic protolanguage and forms a counterexample to some of the arguments that were proposed against it. This experiment involves iterated learning and recall of a set of (initially) holistic whistles produced with a slide whistle. Over generations, the whistle sets become easier to learn and discrete, combinatorial structure emerges.

1. Introduction

Did human protolanguage consist of holistic utterances that were segmented into words (Wray, 1998; Arbib, 2005) or did it start with simple words that were combined into more complex structures (Bickerton, 1990; Tallerman, 2007)? Both positions are plausible from a comparative perspective: although all human languages use complex combinatorial structure, most primate communication systems do not. In such systems one call corresponds to one meaning, and calls do not have discernible internal structure. At most do calls of varying intensity correspond to varying intensity of the message (Slocombe & Zuberbühler, 2006). These are hallmarks of a holistic system. However, examples have also been observed of systems where combinations of calls appear to have different meanings than the individual calls in isolation (Zuberbühler, 2002). This indicates a more synthetic system.

Arguing against holistic protolanguage, Tallerman (2007) points out why holistic protolanguage would be unlikely to lead to compositional syntax. First of all, holistic protolanguage would by definition be irregular. In addition, if it is learned, it is expected to change rapidly over generations, just as modern language does. Tallerman therefore argues that this would not provide a sufficiently stable target for analysis into components. In response to this argument, it has been pointed out that learners often overgeneralize and that inconsistencies in the input do not necessarily prevent the discovery of regularities (K. Smith, 2008; Fitch, 2010). This argument is backed up by computer simulations showing that regularization does indeed happen in cultural transmission (Kirby, Smith, & Brighton, 2004), as well as by cultural learning experiments with humans that show emergence of compositional structure from initially holistic sets of utterances (Kirby, Cornish, & Smith, 2008).

Tallerman's (2007) response to these arguments and to the computational models is that they already assume that there are building blocks – the target words are built up out of a set of discrete segments. She assumes that the task of segmenting signals into relevant elements is impossible if there are no predefined segments, because in that case even the tiniest distinctions between signals should be considered potentially significant. She therefore does not find demonstrations that rely on pre-existing segments convincing.

Here we present experimental results that address the above-mentioned concerns. The data were obtained from a human iterated learning experiment (based on Kirby et al., 2008) intended to investigate the emergence of combinatorial structure in sets of continuous signals. The experiment starts with sets of *continuous* signals that have no regularities. After they have been transmitted over a number of experimental generations, it turns out that human overgeneralization causes the sets of signals to become regularized. Also, the system becomes more constrained and predictable which allows learners to focus on relevant signal variations (Verhoef, Kirby, & Padden, 2011).

2. Cultural transmission of an artificial whistled language

The experiment described below involved learning and recall of twelve distinct whistle sounds that were produced with a slide whistle (see figure 1). The reason for using the slide whistle was to minimize the influence of pre-existing knowledge of speech and phonology. Four parallel diffusion chains were created of ten generations (40 subjects participated in total) in which each participant was exposed to either the recall output of the previous participant in the same chain or the initial whistle set. The initial set was the same in all four chains and these

Figure 1. A slide whistle.

whistles were selected from a database which was created in a pilot study. They were chosen to be holistic and not to exhibit combinatorial structure (more details are reported in Verhoef et al. (2011)). The whistles were not connected to any referent because the goal was to simulate the emergence of 'bare phonology' (Fitch, 2010) and we wanted to control for possible influences of semantics such as iconicity or compositionality.

2.1. *Qualitative observations*

Remembering twelve distinct whistles after limited exposure is not easy and participants usually do not recall all of them. In their productions, they overgeneralize the structure that appears present to them and this results in an introduction of mirrored, combined or repeated versions of existing whistles. The signals consequently share more and more elements, yielding increased combinatorial structure. Recall accuracy also increases, indicating that the more structured sets are more easily remembered.

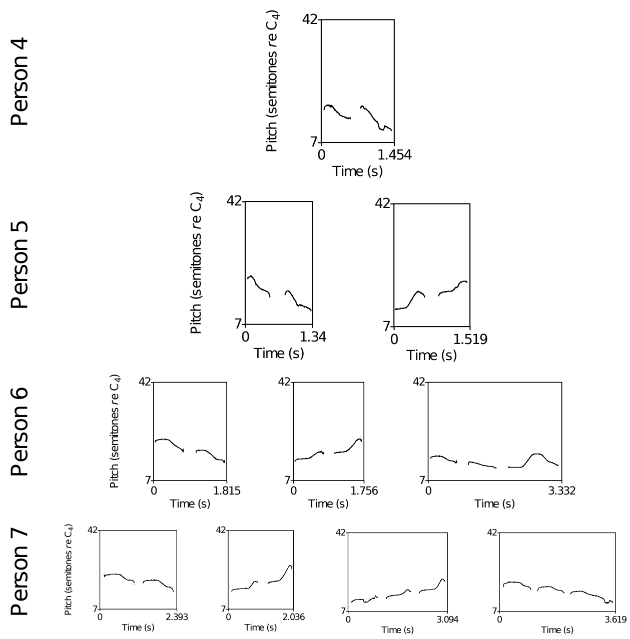

Figure 2. An example of cumulative mirroring, repetition and borrowing. Person 5 mirrors the whistle from the previous set, then person six borrows one of the two in a new whistle and finally this new whistle becomes generalized to fit the pattern of the original two, but repeated. This predictable system is stable towards the end of the chain. The whistles are plotted as pitch tracks.

An example of such cumulative increase of structure over generations is shown in figure 2. In this example combined mirroring, repetition and borrowing results in a predictable system that is stable and persists after its innovation. In generation four there is not yet a whistle that resembles the whistle shown, but in generation five a mirrored version of this whistle appears. Then, in generation six one of these is borrowed and combined into a new whistle. This one may have been interpreted by person seven as having meant to be a repetition of the 'slide down' element present in the original two, because suddenly a '3 slides down' version appears as well as a mirrored '3 slides up' version which fills a gap in this regular system.

Figure 3 shows a fragment of a whistle set that emerged in one of the chains in generation ten. This set exhibits a set of basic building blocks (short level tones and down-up slides) that are systematically reused and combined. Qualitatively, basic elements and systematic recombination can be observed in all four emerged sets of whistles.

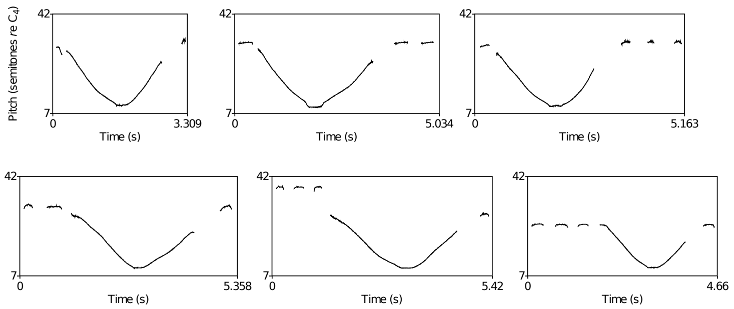

Figure 3. Fragment of the whistles plotted as pitch tracks in the last set of a chain. Basic elements can be identified that are systematically recombined.

2.2. *Quantitative measurements*

In order to quantify these observations, the recall errors of all participants were measured as well as the entropy (Shannon, 1948) of the whistle sets at each generation.

To determine the recall error we measured the distance between two whistles using Derivative Dynamic Time Warping (Keogh & Pazzani, 2001) on the whistle's plunger movement trajectories. Then, the recall error for a participant was measured by comparing the whistles of this person's learned input and recalled output. This was done by pairing the whistles of the two sets in all possible ways, then computing the distance between the paired signals and taking the lowest sum of distances, which belongs to the most optimal pairing. Figure 4 shows this measure for the four chains, with the generations on the horizontal axis. A significant cumulative decrease in recall error was measured using Page's (1963) trend test (L=1318, m=4, n=10, $p<0.05$), implying an increase of learnability and reproducibility of the whistle sets.

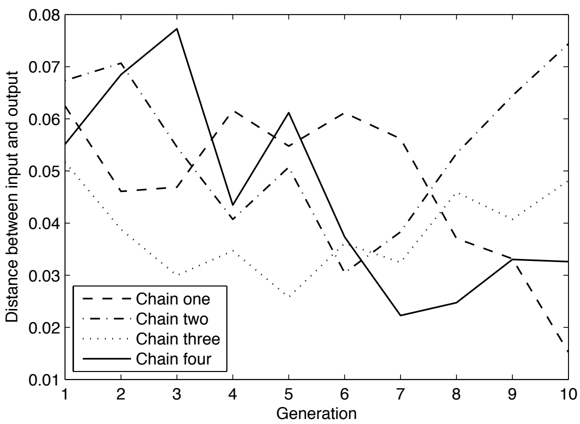

Figure 4. Recall error on the whistle sets over generations for all four chains.

To compute the entropy, the whistles were divided into segments, taking silences as segment boundaries. Then, using all segments that occur in the set of twelve whistles, hierarchical clustering was used to find segment building blocks. Clustering continued until there was no pair of segments left with a distance smaller than 0.03. Equation (1) from Shannon (1948) was then used to compute entropy, where p_i is the probability of occurrence of building block i.

$$H = -\sum p_i \log p_i \tag{1}$$

Figure 5 shows this measure for the four chains, with the generations again on the horizontal axis and 0 referring to the initial set. A significant cumulative decrease in entropy (L=1840, m=4, n=11, $p<0.001$) was measured and this implies an increase of structure and predictability as well as more efficient coding and compressibility.

3. Discussion

The results show that modern human learners quickly generate structure in initially structureless (holistic) sets of continuous signals. Because the initial set is

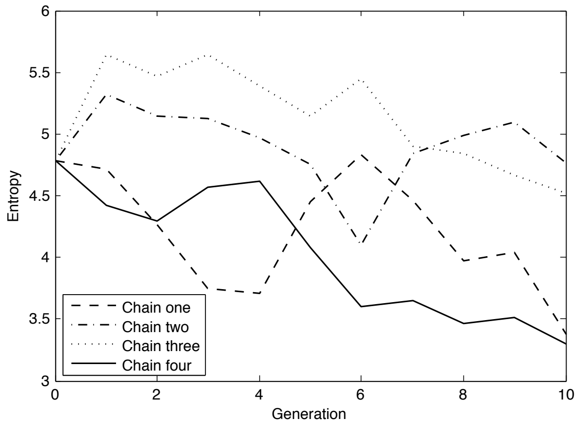

Figure 5. Entropy of the whistle sets over generations for all four chains.

too difficult to learn precisely, learners tend to overgeneralize the structure they (think they) observe. This introduces reuse of a small set of building blocks and increases the learnability of the set of signals. This is in line with the observations of iterated learning of sets of utterances that consisted of discrete symbols (Kirby et al., 2008). Interestingly, the sets of signals were so small that it is unlikely that the emergence of the combinatorial structure was due to any pressure to keep the signals distinct, as was originally proposed by Hockett (1960) and analyzed by Nowak, Krakauer, and Dress (1999) and Zuidema and de Boer (2009). It appears that the emergence of combinatorial structure in the experiment is rather due to properties of the way humans learn and generalize signals.

The results address Tallerman's (2007) objections against the possibility of a holistic protolanguage. Apparently, modern humans have no problems finding (apparent) structure in holistic, continuous utterances. Of course, the utterances in the experiment had no meaning associated with them, so the question of whether regularities in the signals will be associated with regularities in the meaning is not directly addressed. However, the experiments show that participants introduce combinatorial structure very rapidly. Thus, continuous, holistic signals can transform into signals built up out of discrete segments, independent of meaning. Subsequently, as Kirby et al. (2008) have shown, signals built up of discrete elements with an associated meaning (but without a systematic form-meaning mapping) can transform into systems with a systematic form-meaning mapping.

As for the argument about rapid change: rapid change does happen in the experiments, but it leads to structure and better learnability. This is an example of how repeated application of a learning system with limitations creates a linguistic system that is learnable (Zuidema, 2003). Therefore the argument that the changeability of a culturally transmitted holistic system would prevent emergence of structure is not supported by empirical evidence from modern human behavior.

Of course, these observations do not necessarily generalize to ancestral hominins, who may have had very different cognitive adaptations. However, as K. Smith (2008) has pointed out, research with cotton-top tamarins (Hauser, Newport, & Aslin, 2001) has shown that at least some non-human primates already have simple abilities for segmenting streams of speech. It is therefore possible that the ability to find regularities in speech is much older than the split between humans and the other apes, and that the crucial innovation was in the ability to reproduce and generalize learned signals. However, given a pre-existing ability to analyze, we can expect that re-use of regularities could have been possible at the earliest stages of protolanguage.

Although our findings refute arguments against holistic protolanguage, we are not necessarily arguing for an extended holistic protolanguage phase. It appears that such a system would perhaps not be stable for many consecutive generations, because combinatorial structure at the signal level would emerge rapidly. We also do not exclude the possibility that the holophrases were concatenated into larger constructs simultaneously, following a synthetic scenario. In fact, it would be possible to have aspects of both holistic and synthetic protolanguage in one system: holistic phrases break up into smaller units, while at the same time words could be combined into short utterances. As A. Smith (2008) suggested, there is no fundamental contradiction between these two points of view.

At our present state of knowledge, these ideas are still speculative and preliminary. However, we hope to have shown that iterated learning experiments can throw empirical light on the debate between holistic and synthetic protolanguage – a debate which for a long time has been mostly philosophical. Our study is just the first step in a possible series of more refined experiments.

Acknowledgements

We thank Carol Padden, Alex del Giudice, Jelle Zuidema and Wendy Sandler for helpful suggestions and discussions.

References

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, *28*(2), 105–124.

Bickerton, D. (1990). *Language and species.* The University of Chicago Press.

Fitch, W. (2010). *The evolution of language.* Cambridge University Press.

Hauser, M., Newport, E., & Aslin, R. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

Keogh, E. J., & Pazzani, M. J. (2001). Derivative dynamic time warping. In *First SIAM International Conference on Data Mining (sdm2001)* (pp. 1–11).

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, *28*(3), 587–607.

Nowak, M., Krakauer, D., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London*, *266*, 2131–2136.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *58*(301), 216–230.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, *72*(5), 989-999.

Smith, A. (2008). Protolanguage reconstructed. *Interaction Studies*, *9*(1), 99–115.

Smith, K. (2008). Is a holistic protolanguage a plausible precursor to language? *Interaction Studies*, *9*(1), 1–17.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, *117*(3), 579–604.

Verhoef, T., Kirby, S., & Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 483–488). Cognitive Science Society.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, *18*(1), 47–67.

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*(2), 293–299.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15* (pp. 51–58). Cambridge, MA: MIT Press.

Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*(2), 125–144.