THE EFFECT OF PHYSICAL ARTICULATION CONSTRAINTS ON THE EMERGENCE OF COMBINATORIAL STRUCTURE

HANNAH LITTLE & KEREM ERYILMAZ

Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2
1050 Brussels, Belgium
hannah@ai.vub.ac.be, kerem@ai.vub.ac.be

1. Introduction

Language has "duality of patterning", which is structure on both a compositional and a combinatorial level. Compositional structure is the combination of meaningful elements into bigger meaningful structures. Combinatorial structure is the phonological combination of small meaningless units into a potentially infinite number of meaningful units.

Despite "duality of patterning" being named by Hockett (1960) as one of the basic design features of human language, empirical work exploring the emergence of combinatorial structure is still very much in its infancy. Techniques to test existing hypotheses regarding the emergence of phonological structure have only recently been developed, and the strengths and weaknesses within this ongoing work are generating new hypotheses which also need to be tested. The current contribution will outline the existing hypotheses on how combinatorial structure first emerged in language before focusing on hypotheses pertaining to the modality, size and shape of the articulation space. We will then outline existing experimental and computational work which tests the effects of physical articulation constraints on the emergence of combinatorial structure, along with our own ongoing work, and the scope for future work in this area.

2. Existing Hypotheses

Hockett (1960) hypothesised that the emergence of structure on a phonological level is the result of pressures for expressivity and discriminability imposed when the number of meanings increases, as language needs a more efficient way to create new word forms. More recently, Verhoef (2012) has shown experimentally that combinatorial structure can emerge as the result of cognitive learning constraints and biases. However, recent evidence from Al-Sayyid Bedouin Sign Language, which is a newly emerging language, suggests that languages can have thousands of words without a level of phonological patterning (Sandler, Aronoff, Meir, &

Padden, 2011). In a recent paper, Del Giudice (2012) considers that the lack of phonological patterning in emerging sign languages could be because the articulation space in sign languages is much larger than that used in spoken languages, and this allows for a greater number of distinct signals without the need for combinatoriality. This hypothesis is dismissed by Del Giudice (2012) as established sign languages have been shown to have a similarly sized phoneme inventory to those found in spoken languages (Rozelle, 2003). However, this is not evidence to suggest the size of articulation space, as well as other physiological factors, are not important factors in the *emergence* of combinatorial structure in language. Hypotheses regarding the effects of the modality, shape and size of an articulation space have yet to be empirically tested which is what we aim to rectify with this contribution.

3. Experimental Work

Artificial language learning experiments are often used in evolutionary linguistics to show how structure emerges on a compositional level. Work is now appearing on emerging combinatorial structure, started by Verhoef (2012) who used signals created by slide whistles in an iterated learning paradigm. Whistled signals are ideal for the purposes of investigating the emergence of speech as they use a continuous articulatory space, but limit interference from participants' existing linguistic knowledge. In Verhoef's (2012) experiment, participants learned whistled signals and their resulting reproductions became the input for the next participant. Del Giudice (2012) has since carried out a similar iterated experiment where participants created graphical symbols using a moving stylus which limited the use of iconic representation, and found that participants did not use the entirety of the signal space as one would expect if Hockett's (1960) hypothesis were true.

To test the effects of the size of articulation space on the emergence of combinatorial structure, we extended Verhoef's (2012) experiment by running a new condition where the slide whistle was restricted with a stopper, as well as an unrestricted condition. The shape of the whistle's articulation space was kept the same, only differing in size on one dimension. Comparison of combinatoriality between conditions eliminated the problem of an articulation space having some trajectories which are more likely to be produced, which is a problem for analysis when only one condition is being tested. We show that the size of articulation space does indeed have an effect on the emergence of combinatorial structure.

There is a large scope for future experimental work on the effects of physical articulation constraints. A whole host of electronic musical instruments and digitally generated signals are enabling more easily manipulated signal spaces and easily analysable signals. Our next steps are to experimentally test the effects that modality and the dimensionality of a signal space have.

4. Computational Work

The computational work deals with four main issues: the representation of signals, the selection process through which some signals persist while others fall into disuse, the distance and similarity measures between signals, and measures of structure.

4.1. *Signal Space and Signals*

Earlier models of the evolution of combinatorial structure abstract away from the internal structure of signals, representing them as unique symbols (Nowak, Plotkin, & Krakauer, 1999). In such models, the variation in signals necessary for evolution arises from errors in probabilistic learning, and not from comparison of the signals involved. To deal with structure, many later models use signals represented as points or trajectories in an N-dimensional feature space, which may be abstract and not correspond to any actual features of an acoustic signal (de Boer & Zuidema, 2010). The current work deals exclusively with the interplay between the shape of an artificial feature space and the combinatorial structure of signals in that space, abstracting away from the acoustic nature of the features. Each signal consists of a fixed number of ordered points in the feature space, forming a trajectory.

4.2. *Signal Selection*

The signals evolve within a multiagent imitation game. Agents start with a fixed number of randomised signals, and utter them with small, random, shape-preserving mutations as described by de Boer and Zuidema (2010). All signals are further subject to environmental noise but preserve their shape. As in de Boer and Zuidema (2010), each round, a chosen performer agent utters their repertoire $\mathbb{L}$, then the imitating agents utter the closest signal they know to the performer's signal. If the imitation is closer to the original signal than any other in the performer's repertoire, the round is successful. If more imitators are successful using the performer's mutated signal than using the original signal, the performer replaces the original with the modified signal.

4.3. *Signal Distance and Confusion*

For signals represented as trajectories, the easiest distance metric is point-to-point Euclidean distance. However, this may result in overestimation of the distance between similar signals with different timings. We estimate the distance between signals using Dynamic Time Warping (Sakoe & Chiba, 1978), also used in the analysis of some experimental studies. When a signal, X, is emitted, the probability of that signal being identified correctly varies with its distance d to the original position of the signal. This probability is chosen from a Gaussian distribution around X, with the spread δ (i.e. noise level), as in de Boer and Zuidema (2010).

$$f(d) = \int_{x=\frac{1}{2}d}^{\infty} \frac{1}{\sqrt{2\pi}\delta} e^{\frac{-x^2}{2\delta^2}} dx$$

The probability of perceiving the uttered signal X as $Y \in \mathbb{L}$ becomes:

$$P(Y_{perceived}|X_{uttered}) = \frac{f(d(X,Y))}{\sum_{Z\in\mathbb{L}} f(d(X,Z))}$$

4.4. *Measures of Structure*

We propose investigating the amount of structure in the agents' repertoires based on measures motivated by information theory. Specifically, we claim that for signals that can be well-represented by a few data points per signal, such as those in this study, entropy rate of an agent's repertoire is a feasible measure of combinatorial structure.

Choosing a measure of combinatorial structure is far from trivial. It is possible to assume that combinatorial building blocks have greater power to predict what comes next than non-building blocks. However, combinations of these building blocks can also have considerable predictive power. Conversely, trends that appear on very small time scales as opposed to communicatively relevant time scales (combinatorial building blocks) can be artefacts of the articulatory apparatus (or a mathematical or computational proxy). To create a balance between problems at these two extremes, we propose focusing on quantifying the predictability of the signal-generating process per unit time, instead of the predictability of individual signal occurrences. More formally, we propose using a weighted mixture of variable-depth context trees to estimate the entropy rates, given different maximum context depths (Kennel, Shlens, Abarbanel, & Chichilnisky, 2005). By looking at the changes in the estimated entropy rate under different context depths, it is possible to estimate the maximal length of the building blocks. Any part of a signal longer than the longest building block will contain at least two (possibly partial) building blocks. Building blocks have less internal variation than combinations of building blocks, since the blocks themselves do not contain combinatorial parts. Thus, a notable decrease in the estimated entropy rate at a certain depth increment, which is not followed by a comparable decrease at the next depth increment, can be used to estimate the maximum length of a building block.

Theoretically, it is also possible to have an unbounded tree that uses complete trajectories instead of bounded contexts etracted from parts of signals. However, for inventory sizes greater than three or four, such trees becomes impractical both in memory and time complexity, as the context tree can consist of A^{D^D} nodes for an alphabet of size A and a maximum depth of D, depending on the contexts observed.

5. Conclusion

We have argued that physiological constraints are important factors affecting the emergence of combinatoriality within different modalities. We have also outlined problems in existing work which use proxies for articulatory spaces to investigate the emergence of combinatorial structure, and shown how recent experimental and computational techniques can be implemented to test hypotheses pertaining to how physiological constraints can affect the emergence of combinatorial structure. The evolution of speech, as a field, is currently divided between work dealing with the emergence of phonological structure and the cognitive capacity for speech, and work dealing with human phonetic capabilities and the physiological capacity for speech. Fitch (2002) states that some researchers do not even regard phonological evolution as part of speech evolution at all. However, we show that it is important to consider phonetic capabilities when considering the emergence of combinatorial structure.

References

de Boer, B., & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, *18*(2), 141-154.

Del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and cognition*, *4*(4), 381-418.

Fitch, W. T. (2002). Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx. In A. Wray (Ed.), *The transition to language* (p. 21-45). Oxford University Press.

Hockett, C. D. (1960). The origin of speech. *Scientific American*.

Kennel, M. B., Shlens, J., Abarbanel, H. D., & Chichilnisky, E. J. (2005). Estimating entropy rates with Bayesian confidence intervals. *Neural Computation*, *17*(7), 1531–1576.

Nowak, M. A., Plotkin, J. B., & Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, *200*(2), 147–162.

Rozelle, L. G. (2003). *The structure of sign language lexicons: Inventory and distribution of handshape and location.* Unpublished doctoral dissertation, University of Washington.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43–49.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*, *29*(2), 503-543.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and cognition*, *4*(4), 357-380.