

THE ROLES OF SEGMENTATION ABILITY IN LANGUAGE EVOLUTION

KAZUTOSHI SASAHARA

*Laboratory for Biolinguistics, Brain Science Institute, RIKEN, Japan
sasahara@brain.riken.jp*

BJORN MERKER

*Department of Psychology, Uppsala University, Sweden
gyr694c@tinet.se*

KAZUO OKANOYA

*Laboratory for Biolinguistics, Brain Science Institute, RIKEN, Japan
okanoya@brain.riken.jp*

We focus on segmentation ability as a prerequisite of language, studying what part it plays in language evolution with a simple computational model. Language is mediated by distinct sounds and has a characteristics of 'duality'. For such a structure, it is necessary to have segmentation ability, which is an ability to find out discrete units in continuous sound sequences.

To model segmentation ability, we review some experimental findings. In songbirds, it has been found that male Bengalese finches have songs with duality; a chunk consists of phonemes and a song consists of chunks (Okanoya, 2002). A male juvenile of the Bengalese finch learns a song from his father within a certain period. To do so, he must detect the discrete parts (e.g. song elements and chunks) in the flow of the song sample. In infants, a number of experiments have shown that infants are able to find discrete patterns in the flow of adults' utterances by detecting word frequency, the transition rate of sounds, accent patterns, and so on (Tomasello, 2003). Both cases have two common features: (i) the statistical cues in strings contribute to segmentation; and (ii) the dyadic interaction (i.e. father and juvenile, mother and infant) is a 'leader-follower' kind in which one of the two is a well-versed agent and the other is not; hence there is an asymmetry of information flows.

In light of above consideration, we model an evolution of discourse where agents utter strings by turns. Let's suppose a society of N agents, each of which can produce long sound strings and has simple statistical ability. Each agent is modeled by a recurrent network (RNN) that studies the transition rate of sound

elements in the sound strings it hears. In initial state, all network weights of every agent are randomly initialized. Then two agents are randomly chosen to engage in conversation. The utterances of the agents consist of the outputs of their RNN and they are translated into letters (here, *A, B, ..., J*), considered sound elements. When one agent utters a sequence of sounds, the other agent hears it one-by-one and predicts the next sound element in the utterance. After that, the agent's RNN is trained with supervised learning in such a way that it can predict the transition of sound elements better. Then the agents take turns uttering and hearing. The procedures are repeated over within a certain number of discourses.

With this model, we demonstrate how common shared words (i.e. frequently used sound patterns) emerge and how the distribution of sound elements changes from random initial state as common words increase. In the early stages of the evolution, common words were rare in the artificial society because the patterns of sounds were almost random. However, once common patterns emerged in sound strings, some of them came to stay in the discourses of the agents. Furthermore, we consider how the leader-follower interaction of agents contribute to the emergence of words. Self-organization the leader-follower interaction in our agents was difficult just because of statistical cues in discourse. Our results show that if agents have simple statistical ability, the frequently appearing patterns in sound strings may become established as words through the interaction of the agents and that emerging words may affect succeeding discourses in evolution.

So far the certain patterns of sounds have been described as words; however these are not exactly the same as words due to the lack of meaning. If we take the following 'mutual segmentation' hypothesis into account, our model may deal with syntax and semantics within a single framework (Merker & Okanoya, 2005). Suppose a society without language. When agents with segmentation ability collaborate, the common parts of behavioral, environmental and social context they face and the common parts of sound strings they utter could be mutually segmenting, and the segmented small parts of sound strings could link to ever more specific contexts; a word and a meaning could emerge into co-existence. Our model at present doesn't have any context. We plan to extend it by introducing behavioral context of agents (e.g. sensory-motor experience) to explore that hypothesis.

References

- Merker, B., & Okanoya, K. (2005). Contextual semanticization of songstring syntax: a possible path to human language. *Proceedings of Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC 2005)*, 72–76.
- Okanoya, K. (2002). Sexual display as a syntactical vehicle. In *The transition to language* (pp. 46–63). Oxford University Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.