# EVOLUTION OF THE WORD ORDER RULES THROUGH COMMUNICATION AND CULTURAL TRANSMISSION: AN ITERATED LEARNING EXPERIMENT

Seiya Nakata[*1,2,4], Akari Ohira[1], and Masanori Takezawa[*1,3,4]

[*]Corresponding Author: snakata@eis.hokudai.ac.jp, m.takezawa@let.hokudai.ac.jp
[1]Department Behavioral Science, Hokkaido University, Sapporo, Japan
[2]Japan Society for the Promotion of Science, Japan
[3]Center for Experimental Research in Social Sciences, Hokkaido University, Japan
[4]Center for Human Nature, Artificial Intelligence and Neuroscience, Hokkaido University, Japan

The grammatical structure found in human language may have emerged through communication and cultural transmission. To focus on the role of communication in the evolution of word order rules, we designed an iterated learning experiment in which participants use a given set of words. The results of the experiment showed that communication facilitates the cultural evolution of word order rules. In cultural transmission chains with communication, the structural language with word order rules emerged rapidly and remained stable throughout the generations. On the other hand, in the cultural transmission chain without communication, the evolution of word order rules was very slow, and the word order distribution didn't show stability. These findings suggest that word order rules culturally evolve by sharing the intention to communicate.

## 1. Introduction

Human languages have a systematic grammatical structure. For example, the order of subject-object-verb (SOV) is determined according to the grammatical rules of each language (Schouwstra, & de Swart, 2014). The systematic ordering of elements by grammatical rules allows for smooth communication among speakers of the same language. The systematic word order rules, rather than simply listing words, are rare in the communication systems of non-human animals (Hauser, Chomsky, & Fitch, 2002; but see also Suzuki, 2014; Suzuki, Wheatcroft, & Griesser, 2017). The question of where the systematic structure found in human language emerges from is one of the central questions in evolutionary linguistics. Studies using iterative learning experiments have shown that the linguistic structure emerges from a trade-off between learnability (cultural transmission) and expressivity (communication) (Kirby et al., 2008; 2015; Winters, 2015). In these experimental studies, reusable parts of strings corresponding to individual meanings evolved from the random strings of the initial language. They used small and simple meaning spaces and showed the emergence of basic compositional structure. Saldana et al. (2019) expanded the semantic space and investigated the emergence of the compositional hierarchical structure of language in the iterated learning experiments. They made the task more complex and implemented communication among participants. Through a series of iterated learning experiments, they found that compositional linguistic structures having hierarchical word order rules inherent to natural languages evolved under cultural transmission and communication.

In the current study, we focused on the cultural evolution of grammatical structures exhibited as word order rules. Initial language in Saldana et al. (2019) was holistic strings each corresponding to a unique stimulus. They found the emergence of two levels of the hierarchy of constituents: word-like forms and sentence-like structures. In our experiment, the initial language was a sequence of randomly ordered words. Each word referred to either an actor or an action. As there was no systematic rule determining word order, a sequence is regarded as holistic language each of which corresponds to a unique meaning. By employing a similar experimental setting as experiment 2 in Saldana et al. (2019), we investigated whether grammatical rules for determining word orders emerge through the processes of cultural transmission and communication.

We were particularly interested in whether communication plays an essential role in the evolution of the word order rules. We compared a solo condition, in which languages were transmitted in a single transmission chain, with a pair condition, in which pairs of individuals communicate with languages. If communication plays an essential role, we will observe the rapid emergence of a word order rule in the pair condition. If the cultural transmission is sufficient for the emergence of word order rules, we won't observe much difference between the conditions.

## 2. Methods

### *2.1. Participants*

120 (51 females and 69 males) undergraduates of Hokkaido University, whose native language was Japanese, were recruited. Eighty and forty participants were assigned to the pair condition and the solo condition respectively. Mean age was 19.93 years (*SD* = 1.55; ranging from18 to 23). Participants received 500 JPY

(about 4.33 USD) and additional payment based on the number of successes in the task (described later).

### *2.2. Stimuli*

Participants were asked to learn a combination of a two-second animation and a sentence. All the animation consisted of two actors and the action of each actor (Fig. 1). A sentence was made of four words; two words represent actors (a red rabbit, a blue rabbit, a red bird, or a blue bird) and the other two words represent an action of actors (a rabbit punching, a rabbit getting angry, a bird punching, or a bird getting angry). Each word was randomly made of two to three Japanese letters and participants could not modify the words. Participants were allowed to see the correspondence table of the words and their meaning during the experiment. The initial sentences presented to the participants in the first generation were created by randomly ordering the four words representing the elements in an animation. The order of words representing actions was fixed for representing the chronological order of the actions – the action word appearing first represents the action of the left actor, who first initiates an action, in the animation.


Figure 1. An example of animation and a sentence. Each animation started with an action of a left actor followed by an action of a right actor (the text in brackets was not shown).

### *2.3. Experimental procedures*

The experiment consists of two phases. In the learning phase, participants were asked to memorize pairs of animations and sentences presented on a computer screen. There were 16 pairs of stimuli and eight were randomly selected for the learning phase. In each round of the learning phase, a pair of a sentence and animation was presented on the computer screen for 6 seconds. At the end of each round of presentation, one of two confirmation tests was randomly administered. In the sentence production test, participants were presented with only the animation again and were asked to reproduce the presented sentence by rearranging the four given words. In the animation selection test, participants were presented with only the sentence again and asked to select the corresponding animation from a given set of four animations. Each stimulus was presented three times and the learning phase consisted of 24 rounds (three blocks of eight stimuli).

Next, participants in the pair condition engaged in a communication phase where two participants worked on the communication task for a total of 64 rounds (four blocks of sixteen stimuli). In the communication phase, a total of 16 stimuli were used, including 8 stimuli that did not appear in the learning phase. The 16 stimuli were randomly ordered and presented in each block. The roles of speaker and hearer were alternated every round so that a participant played the roles of a speaker and a hearer twice for each stimulus. Participant assigned a speaker's role was shown an animation and asked to produce a corresponding sentence by rearranging the four words presented. Participants assigned a hearer's role were presented with a sentence produced by the speaker and were asked to choose a corresponding animation from four candidate animations. If the animation selected by the hearer matched the one presented to the speaker, both participants received 20 JPY as an additional payment. After each round, both participants received feedback on their communication; the animation presented to the speaker, the sentence produced by the speaker, the animation selected by the hearer were displayed with the cumulative number of successes. After completing the experiment, participants answered a post-questionnaire and received payments based on the number of successes.

Instead of the communication phase, participants in the solo condition engaged in the test phase. Participants were given one of the two tests similar to the ones used in the communication phase of the pair condition – either creating a sentence corresponding to the displayed animation by rearranging the four presented words or choosing the animation corresponding to the displayed sentence out of the four candidates. The number of rounds of the test phase was identical to the communication phase in the pair condition. At the end of every round in the test phase, participants received feedback and 40 JPY was added to the payment for each correct answer.

Participants were organized into an independent transmission chain each with five generations. In the solo condition, 16 pairs of animations and sentences produced by a participant in the third and last blocks of the test phase were transmitted. In the pair condition, one of the participants was randomly selected and the 16 pairs of sentences and animations produced as a speaker in the third and last block of the communication phase were transmitted. In both conditions, eight pairs were randomly selected and displayed in the learning phase of the next generation. The animation was created by Adobe Animate 2021, and the experiment was programmed in oTree v. 3.3.11 (Chen, Schonger, & Wickens, 2016).

## 3. Results

### *3.1. Emergence of word order rules*

In the following, "N1" refers to a word meaning the actor who initiates the action, and "N2" refers to a word meaning the actor reacted to the action of N1. "V1" and "V2" refer to the action of N1 and N2, respectively. All the sentences are made by ordering these four words (N1, N2, V1, and V2). Initial sentences were made by randomly ordering words although V1 always preceded V2 for representing the order of actions. The number of possible orderings is 24. For analyzing the emergence of word order rules, we calculated the Shannon entropy (Shannon, 1948) for a frequency distribution of each order on 16 sentences transmitted to the next generation. If all the 16 sentences were made by a single ordering, the value of entropy becomes zero.

Fig. 2 (left) shows the mean Shannon entropy per generation for each condition. Values in generation zero were calculated based on the initial 8 sentences participants in the first generation learned in the learning phase. In the pair condition, entropy approached zero at the first generation and remained unchanged. In the solo condition, entropy did not change until the third generation and gradually lowered from the fourth generation. It is suggested that a single word order rule quickly emerged in the pair condition while a word order rule very slowly evolved in the solo condition.

We fitted a linear mixed model where conditions, generations, and interaction effect were entered as predicting factor variables (chain was treated as a random effect variable). We then calculated estimated marginal means for post-hoc contrasts. We found that, except for the generation zero, the entropy in the pair condition was significantly smaller than the solo condition in all the generations ($M_{\text{(pair – solo)}}$ = –2.41 ~ –1.06, *p*s = .000). In the pair condition, the entropy in the generation zero was significantly larger than the other generations ($M_{\text{(0 – 1, 2, 3, 4, or 5)}}$ = 2.18~2.51, *p*s = .000). In the solo condition, entropy in the generation 4 and 5 were significantly smaller than the other generations ($M_{\text{(0,1,2 or 3 – 4 or 5)}}$ 0.59~1.38, *p*s = .05 ~ .001).

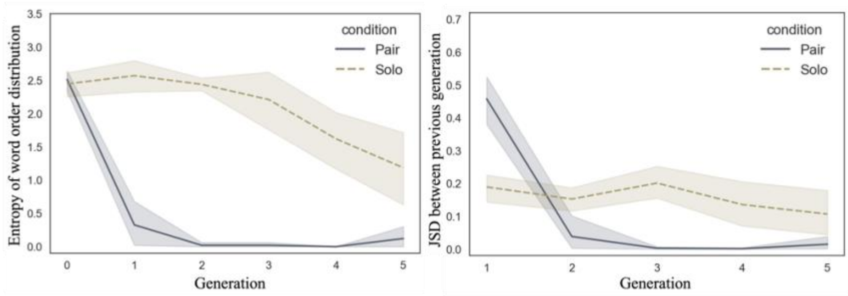


Figure 2. (left) Entropy of word order distribution plotted by generation. Lines and shadows show mean and bootstrapped 95% CI respectively. (right) Similarity of the word order distribution between neighboring generations. Distances between t–1 and t are shown as t on the horizontal axis. Lines and shadows show mean JSD between previous generation and bootstrapped 95% CI respectively.

### *3.2. Word order stability*

For analyzing the similarity of word orders within a chain, we calculated the Jensen–Shannon divergence metric (JSD) as the distance of word order distributions between neighboring generations (i.e., $t$–1 vs. $t$). The lower JSD indicates word order distributions are more similar (Fig. 2 (right)). As we've already discussed, in the pair condition, a single word order rule quickly emerged in generation 1 and remained unchanged for the rest of the generations. This pattern is also exhibited in the trend of JSD distance in the pair condition. On the other hand, in the solo condition, the language changed a little each generation, and no stable word order distribution emerged.

We fitted the similar LMM on the JSD and found that, in the solo condition, JSD were not significantly different across generations ($M$ = –0.05~0.09, $p$s = .082~.999). In the pair condition, JSD in the first generation significantly larger than the other generations ($M_{(1-2\sim5)}$ = 0.42~0.46, $p$s = .000).

### *3.3. Evolved word order rules reflected causal order in the referent*

Table 1 shows the most frequently used word orders in the final generation of each chain. In both conditions, "N1-V1-N2-V2" and "V1-N1-V2-N2" (italics in the table) were the most frequently used although word orders were more diverse in the solo condition. Although any word order could achieve successful communication in the pair condition, we found that the word orders reflecting the order of action in the animation were more likely to evolve in the experiment, which was similar to that of the participants' native language, Japanese.

Table 1. The most frequently used word orders utilized in the final generation (numbers in parentheses are frequencies).

| | *Pair condition* | *Solo condition* |
|---|---|---|
| Chain 1 | *V1-N1-V2-N2* (75%) | *N1-V1-N2-V2* (38%), V1-N1-N2-V2 (38%) |
| Chain 2 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (100%) |
| Chain 3 | *V1-N1-V2-N2* (100%) | N1-N2-V1-V2 (25%), N1-V1-V2-N2 (25%), V1-N1-N2-V2 (25%), *V1-N1-V2-N2* (25%) |
| Chain 4 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (50%) |
| Chain 5 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (100%) |
| Chain 6 | *N1-V1-N2-V2* (97%) | N1-V1-V2-N2 (94%) |
| Chain 7 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (31%), N1-V1-V2-N2 (31%) |
| Chain 8 | *N1-V1-N2-V2* (97%) | *V1-N1-V2-N2* (50%) |

## 4. Discussion

We investigated the hypothesis that communication facilitates the cultural evolution of the word order rules. Our iterated learning experiments showed that communication is a powerful pressure on the evolution of word order rules. In the pair condition where communication is necessary, structural languages with word order rules emerged rapidly. The word order distributions in the pair condition showed stability throughout the generation. In contrast, in the solo condition, the evolution of word order rules was much slower, and the word order distribution changed from generation to generation.

Specific word orders became dominant in the final generation of all chains in the pair condition. On the other hand, in the solo condition, various word orders remained even in the final generation. Throughout the experiment, participants frequently produced word orders to reflect the order of the actions in the animation. In our experimental design, participants could communicate even with the initial random word order sentences because they could see the correspondence table of the words and their meanings. Nonetheless, it is interesting to note that specific word order rules were preferred. In other words, the emergence of word order rules could be explained by the intent of the communication, not by the need to understand the meanings of sentences. It is suggested that the rapid emergence of word order rules in the pair condition is due to the shared intention to communicate appropriate meaning to each other. However, the impact of an incentive to reproduce the initial language in the solo condition (which would not be present in the pair condition) should also be considered in the future. Finally, the participants' acquired languages could attribute to the high frequency of causal order emergence. Therefore, it would be necessary to conduct future experiments with participants from various language families at various stages of language development.

### Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17H06383. We would like to thank Koji Fujita on the helpful advice for the experimental design.

### References

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88-97.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

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., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.

Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional hierarchical structure evolves through cultural transmission: an experimental study. *Journal of Language Evolution*, *4*(2), 83-107.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431-6.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, *87*, 59-65.

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature communications*, *7*(1), 1-7.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415-449.