

Katarzyna Rogalska-Chodecka

Nicolaus Copernicus University in Poland, Department of Italian
kasia@umk.pl

The Role of Interaction in Language Evolution. Evidence from Two Iterated Learning Experiments with Human Participants

Abstract. Experiments conducted with the use of the iterated learning methodology have recently become a rich source of information on the hypothetic ways in which the human language might have evolved. Basing on the groundbreaking studies of the Language Evolution and Computation Centre in Edinburgh, it is possible to utilise the methodology in question in increasingly more advanced experiments, both with and without human participants, using various designs and referring to different theories. The present paper discusses the role of interaction in iterated learning experiments, comparing the results obtained in two studies: the first one conducted with only one participant at a time (without communication) and the second one with three participants interacting with each other. As results suggest, the intuition that communication constitutes one of key factors in language evolution might be misleading.

Keywords: language evolution; iterated learning; communication; interaction; human participants.

Introduction

The way in which human language evolved constitutes one of the greatest mysteries of modern science (cf. Christiansen & Kirby, 2003, p. 30). Consequently, the very beginning of the 21st century marks an increasing interest of scientists coming from various fields in the study of language

origins. From the linguistic perspective, one of the most promising and valuable insights into the transmission of linguistic structures and the consequent evolution of the linguistic code is the methodology of iterated learning. Initially elaborated for mathematical and computational models, the methodology in question is now commonly used in studies involving human participants. Based on a simple “alien fruit” experiment originated by Kirby and his collaborators (known to a wider audience thanks to a BBC Two Horizon TV series entitled “Why do we talk?”, shot between 2009 and 2010), it is possible to overcome the most commonly held accusation against evolutionary linguistics, namely that as languages do not fossilise, they cannot be studied empirically (e.g. Fitch, 2000, p. 262). Experiments utilising the iterated learning methodology allow for the observation of artificial mini-languages emergence and development within just several hours instead of several thousand or even hundreds of thousands of years (Waciewicz, 2013, p. 11). The experiments conducted by the Language Evolution and Computation Centre in Edinburgh, both those with (e.g. Cornish, 2011) and without (e.g. Kirby & Hurford, 2002; Smith et al., 2003; Brighton et al., 2005) human participants, can undoubtedly be considered a great success. According to their authors, artificial mini-languages produced in such studies develop to become organised, combinatorial, learnable and easy to reproduce (e.g. Verhoef et al., 2014, p. 66). Consequently, they should be considered an example of the way in which the human language might have developed.

The present paper analyses two experiments conducted with the use of the iterated learning methodology, based on the model described by Kirby and Hurford (2002, p. 122). The methodology in question involves three stages: training, testing, and transmission. In the first stage, participants are supposed to familiarise themselves with the names of several “alien” objects, in the case of the experiments to be described here, alien fruits. They are presented as labelled images, displayed one by one. Each of the images shows one, two, or three identical fruits (three shapes are possible in total) in one of three possible colours: blue, red, or yellow. In the testing stage, the participants’ ability to remember the signal-meaning pairs is verified by asking them to name the fruits presented in a different order and with minor replacements. Finally, in the transmission phase, the signals produced by the previous participant become the input for the next participant of the experiment. Each chain of such iterations finally leads to the emergence of a mini-language. Importantly, the participants are not aware of the fact that they acquire language from the previous participant, creating a diffusion chain (Cornish, 2011, p. 7). Consequently, the notion of “a diffusion chain”

refers to a sequence of iterations, where the output of one subject becomes the input for the next one.

The first experiment to be analysed below constitutes a replication of the experiment known from the BBC documentary mentioned above. As there is no publication referring directly to the experiment in question, the design was mainly based on the observation of the version presented in the documentary, as well as on publications referring to other, relatively similar experiments (e.g. Kirby et al., 2008; Cornish, 2011). The main aim of the replication was to check whether results similar to those reported in the BBC documentary, i.e. increasing learnability and compositionality of the mini-language produced, will be observed. On the other hand, the second experiment, involves a similar design as in the case of the first one to be described here, with one major change: instead of individual participants in each generation, there are groups of three interacting participants creating the signals. It is expected that communication will positively influence the learnability and compositionality levels of the mini-language produced. The introduction of interaction into the experimental design seems to be a natural consequence of the intuition that language develops through communication. Similar conclusions may be found in studies referring to animal communication (e.g. Dor et al., 2014; Reboul, 2015; Scott-Phillips, 2015). Following Knight, language “evolved in the context of uniquely human strategies of social cooperation” (2000, p. 19). Therefore, it can be suggested that communication between experiment participants might be an important factor determining the learnability and compositionality of the mini-language produced in the second experiment.

Data from the Experiment Without Interaction

The attempt at replicating the original “alien fruit” experiment was successfully conducted with 18 human participants between 22 and 47 years of age, ten males and eight females, who created two diffusion chains, with nine generations of participants in each of them. The initial input was a set of randomly created 3-syllable words with a CVCVCV structure to describe “alien” fruits varying in three aspects: colour, shape, and number. The participants were instructed to familiarise themselves with words describing nine alien fruits (or groups of fruits), after which they were asked to name the fruits (the testing phase). They were not aware of the fact that the experiment was related to language evolution and believed that only their ability to memorise the items would be tested. They also did not know that

the words they produced would serve as input data for further generations of participants. Finally, the participants were not informed that there was a “trap” in the experiment design, as in the testing phase only six of the fruits seen previously by the participants were familiar, and three of them have not been shown before. In consequence, the participants did not have a chance to learn the words to describe them, which remained unnoticed by most of them, and had to come up with their own names, which constituted the most important part of the experiment from the language evolution perspective. As the experiment did not involve interaction, it is worth mentioning that during the study, the experimenter was present in the room, but not involved in any kind of conversation with the participants after the instructions were given. Still, the participants were encouraged to read the names of “alien fruits” aloud in order to enhance memorisation.

The experiment in question was aimed at controlling the most significant effect obtained by in the original “alien fruit” experiment, i.e. the fact that over several generations of users, the mini-language they produce becomes structured, although initially it was completely random. In consequence, it tested the learnability and compositionality hypotheses known from Cornish’s iterated learning experiments (2011, p. 109). They can be presented as follows:

1. The learnability hypothesis: languages will become easier to learn as a result of iterated learning;
2. The compositionality hypothesis: pressure to generalise to novel stimuli will result in languages evolving to become compositional.

The first hypothesis was tested using Levenshtein Distance (known as LD), also suggested by Cornish, which can be done using an online tool available at <http://www.let.rug.nl/kleiweg/lev/>. It allows for comparing distances between signal strings, calculating both substitutions between two different signals, and insertions as well as deletions of letters (Levenshtein, 1966). The closer the value is to 0, the lesser transmission error occurred. On the other hand, if the value is equal to 1, it means that the language was not learned at all. To sum up, in order to see whether learnability increases, one needs to check whether the number of errors made by the participants decreases. This simple method seems to be particularly useful when signals shown to participants are visual, not acoustic. The second hypothesis was tested by measuring the number of distinct signal strings found in the mini-languages across generations. If the mini-languages remained compositional until the final generation, there would be no drop in the number of distinct signals given by the participants, and there would still be nine such signals, each describing a different meaning.

When it comes to the results¹ obtained in the experiment, Levenshtein Distance was calculated for six generations in each chain, i.e. for every complete generation (as three of them were incomplete as a result of the experiment design). A significant drop in the number of errors between generations could be observed, especially in the final iteration. It leaves no doubt about the learnability increase in both chains. However, despite the fact that the final generation showed a significant increase in signals' learnability, the falling trend in error scores cannot be called gradual. This in turn may suggest that, although the final generation seems to be corroborating the learnability hypothesis, in the course of iterations it is rather hard to talk about a consistent tendency. As certain signals were transferred to different meanings, yet managed to stay within the mini-language, the number of distinct signals was checked to provide information as to whether the language generated by the participants was compositional. If the mini-languages remained compositional until the final generation, there would be no drop in the number of distinct signals given by the participants, and there would still be nine such signals, each describing a different meaning. In fact, there was a gradual drop, which indicates a relevant simplification in both chains, as the same number of meanings is described by a lower number of signals. It is also meaningful that in the course of iterations, the emergence of "trendsetters" can be noticed in both chains. The notion of a "trendsetter" refers to an emergent signal exhibiting entrenched linguistic structures (ELS) bias that immediately proves highly adaptive and successful across the subsequent generations (Rogalska, 2015, p. 227). The term "entrenched linguistic structures" refers to the language stored in participants' minds. In the case of the experiment discussed, data analysis shows that the mini-languages are almost entirely shaped by ELS in both chains. As ELS forms appear to be more salient throughout the generations than those unrelated to ELS, they are highly responsible for the appearance of the final mini-languages.

The analysis of the two chains of iterations conducted above makes it possible to state that entrenched linguistic structures, i.e. previously acquired knowledge of various linguistic forms, coming either from the mother tongue of the participants or from foreign languages, are present in iterated learning experiments' results and play an important role in the process of learning through the iterated learning design, as well as they influence the process of simplification of these structures by means of underspecification. Consequently, the compositionality of the mini-languages decreases due to

¹ Detailed results of this experiment were published in Rogalska, 2015.

repetitions and the fact that the number of distinct signal strings drops. The conclusion coming from the first experiment may be formulated as follows: the emergent signals which exhibit ELS bias enhance learnability of a given mini-language. Such signals should be rejected in the second study. However, to avoid their manual elimination, which would lead to results' manipulation, it seems reasonable to introduce interaction into the experiment design. The expected effect of communication between participants would be, apart from enhancing the learnability and compositionality of the mini-languages, the elimination of ELS-biased signals, as working in groups, participants should control one another better and not try to "smuggle" signals that were not present in the learning phase.

Data from the Experiment with Interaction

The second experiment was carried out with adolescents, all of whom were secondary school students. The main aim of experiment with interaction was to compare the results of the individual non-interactive experiment, in which language evolved without communication, with those interactive and closer to the possible way in which human language evolved hundreds of thousands years ago. Being a communicative system which was designed without a designer (cf. Cornish, 2011), language must have been created through communicative interactions. The basis for the experiment is once again the "alien fruit" experiment, in which learners are asked to memorise the names of nine "alien" fruits, all of which are randomised CVCVCV signals paired with pictures presenting three different kinds of non-existent fruits, differing in colour and number. The most important difference between the two experiments lies in the fact that instead of one learner in each generation, there were three of them. They were encouraged to discuss their answers and were asked to always work out a compromise if they could not agree on a given item. There were 54 participants in total, all native speakers of Polish and secondary school students aged between 15 and 17. The main reason behind conducting the experiment with a less-varied group when it comes to the age of participants than in the case of the first experiment was that the participants were schoolmates. As they were acquainted with one another, there were no barriers in expressing their thoughts, which would probably have been the case if they had not known each other.

The present experiment was designed to control hypotheses very similar to those used in the previous experiment; yet, somewhat modified, as the

expected effect was to be always stronger due to the communicative design of the study:

1. The learnability hypothesis: languages in the communicative experiment will become easier to learn than in individual ones, as a result of iterated learning and interaction between participants.
2. The compositionality hypothesis: pressure to generalise to novel stimuli will result in languages evolving to become compositional. The evolution will be observable as a result of communication within generations of users.
3. The ELS vs. interaction hypothesis: interaction will enhance the corroboration of the two previous hypotheses, as language evolves communicatively; the effect of entrenched linguistic structures in the emerging mini-language will not be visible as a result of mutual control among participants.

Similarly to the previously analysed experiment, the learnability hypothesis was controlled by measuring Levenshtein Distance between consecutive signals. The results are somewhat puzzling. It was hypothesised that communication between participants would enhance signal learnability. However, there were continuous leaps in errors made by the participants. Still, the final value is rather low. Unfortunately, the decrease in the number of errors made by the participants was definitely not a gradual and constant one. The situation is very similar in both chains, which confirms the fact that the leaps were not just a matter of blind chance, but need to be understood in terms of a general tendency in the experiment. Having obtained communicative freedom, some groups of learners might have become distracted, while others were motivated by that. In consequence, the learnability hypothesis can be neither corroborated nor falsified.

The falsification or corroboration of the compositionality hypothesis is based on an analysis of the final states of the two mini-languages produced in both chains.² Almost all of the signals produced have the form of CVCVCV, with two exceptions, where one more CV syllable was added. Shape of the fruits is not marked in any way in the signals obtained. When it comes to the colour, the signals are varied in a very similar manner, leaving no doubt about any particular elements not being assigned to a concrete colour. Finally, the numbers of fruits are also not marked in any way. Therefore, it can be concluded that the final state of the mini-language in the first chain exhibits no signs of compositionality. In consequence, in the case

² Tables containing final states of the mini-languages and complete mappings across generations from both experiments can be found in Rogalska, 2016.

of the first chain, the third hypothesis concerning unbiased signals should be falsified. What is more, although their contribution to the creation of the mini-language was minimal, previously acquired structures were still present and became immediate trendsetters. Still, the mini-language produced in the first chain is by no means compositional, despite the fact it contains certain characteristic features (however, it is very difficult to assign them to the qualities of the presented fruits). Whether such an effect of communication within generations is a rule will be controlled by means of analysing the second chain.

As far as the second chain is concerned, certain features of signals from the final state of the mini-language can be noticed immediately. Almost all of the obtained signals follow the CVCVCV pattern, with one exception, which follows a slightly diversified pattern of CVCVCVV. As far as the colour structure of the signals obtained is concerned, the results are also quite satisfying. It turns out that the fact that the third fruit has two possible beginnings is justified here, as the fruits starting with “na” are always blue, while the remaining ones start with “pa”. No other clear colour distinction can be noticed. Finally, when it comes to the numbers of fruits, it could be claimed that the suffix “o” is present only if the fruit is plural; however, it seems to be an overgeneralisation or a matter of coincidence. Still, analysing only the final state of the mini-language produced in the second chain, it is possible to state that it is to a certain extent compositional. Such a claim can be made especially after comparing the second chain with the first one. This, in turn, means that the results obtained in the two initially identical chains brought different results as far as the compositionality hypothesis is concerned. Therefore, the hypothesis in question can be neither corroborated nor falsified on the basis of the present experiment solely.

In the case of the second chain, the third hypothesis concerning unbiased signals is corroborated. Despite the fact no strong trendsetters could be found, the mini-language obtained in the second chain does not exhibit any obvious similarities to Polish or to the most popular foreign languages, which could be known by the experiment’s participants, especially taking into account the fact that they had to work in groups. Although in the chain in question, the participants tried to associate the signals obtained in the training phase with already existing words, it turned out that the task was too difficult for them and they withdrew from using this technique. It even happened several times that one participant suggested to relate the items to Polish or English words, but the two other participants said it was not right as the words were not supposed to be similar to any known language.

In conclusion, the communicative version of the experiment brought somewhat different results from those expected. It turned out that working in groups did not enhance the learnability of the signals to a major extent, which might be due to the fact that an ongoing discussion distorted the memorisation abilities of the participants. As far as the compositionality hypothesis is concerned, it can be only partially accepted: the final state of the two mini-languages leads to contrasting results. In the first chain, the mini-language does not seem to be compositional, while in the second one it does. Consequently, the hypothesis concerning interaction can be falsified, especially taking into account the fact that ELS-biased signals were still present in the first chain.

Discussion: Interaction in Language Evolution Studies

The present paper does not constitute the first instantiation of the introduction of interaction in experiments devoted to language evolution. In fact, it has been under scrutiny of many other research centres recently. Among studies investigating the role of interaction in language emergence it is worth mentioning Schouwstra and colleagues who discuss the experiment by Winters et al. (2016), involving a communication game (but not mutual production of signals), and another one by Feher et al. (2016), where the input language was already marked with morphological features. Still, the results proved that “language structure does change as a result of interaction and that interaction can have a lasting effect on language structure” (2016, p. 2).

When it comes to iterated learning experiments involving interaction, the one conducted by Kirby et al. in 2015 seems to be an important contribution. The experiment in question aimed at investigating the effects of communicative interaction and cultural transmission through iterated learning involving both simulations and an experiment with human participants. Still, the participants did not produce language simultaneously. As a result, making comparisons between this experiment and those discussed in this paper may be difficult. Nevertheless, several relevant aspects can be noticed. First of all, Kirby et al. claim that only communicative experiments can lead to the mini-language produced becoming more expressive, i.e. communicatively functional and providing a distinct, unambiguous form for every meaning (2015, p. 94). There are, however, also certain reservations: no reference to the influence of entrenched linguistic structures was made. What is more, the reader has no access to complete mappings across generations; therefore, it is impossible to

state whether such influences were in fact present. As known from previous experiments of the Edinburgh Group, it was often the case that the output of the participants was carefully selected before it was transmitted to the next generation. It is important to note that such methods were not used in the experiments discussed in the previous sections of this paper.

Another interactive experiment involving human participants and the iterated learning methodology was conducted by Carr et al. in 2017. The results were juxtaposed with those coming from an identical experiment with individual participants, and it turned out that the number of unique labels produced by pairs of participants was higher than in the case of individuals. Such an observation, however, cannot be confirmed in the case of the two experiments analysed in this paper.

When it comes to the conclusions coming from the abovementioned experiments, several significant observation can be made in the context of the two experiments described in the previous sections of the paper. Interestingly, none of them involved a design similar to the one utilised in the interactive experiment analysed in the paper. In all of them, at least one part of the experiment was conducted individually (usually the learning phase); what is more, in most of them one individual learnt from the other, and they were not involved in the learning process as a group. This is particularly important, as in fact they did not communicate in the way allowing for creating or recalling the signals together. The aspect in question is of much importance, as the motivation behind conducting the interactive experiment was to see the participants negotiate the signals and choose the most appropriate one for a given meaning together. This could not be observed in any of communicative experiments with human participants known to the author of the present paper.

Coming back to the experiments discussed in the previous sections of the paper, one major difference between the results obtained in the individual experiment and those from the interactive one must be noticed, namely the entrenched linguistic bias. While in the experiment where participants were asked to learn and recall a part of the mini-language individually the influence of previously known structures on their answers was robust, the introduction of communication in the experiment reduced the effect. In consequence, it can be stated that communication and collective learning are a successful way of eliminating entrenched linguistic bias from iterated learning experiments' results.³ This, however, does not influence

³ An extensive discussion of the problem of ELS bias in iterated learning experiments can be found the author's PhD thesis (Rogalska, 2016).

the learnability and compositionality of the mini-languages produced to the expected extent, and when it comes to entrenched linguistic structures, biased signals can be simply filtered out. However, by doing that, the results are no longer that reliable, as they do not show a natural language evolution process, but a process directed by the experimenter choosing the right signals to stay in the language. Although iterated learning experiments are aimed at discovering the mechanism of language evolution, not at observing the emergence of individual mini-languages, such an observation may indicate certain problems in the design. In order to observe the mechanism it is necessary to make the process as similar to the way in which language could have evolved as possible. Therefore, the first step should be to reject all manipulation from the experimental design. Secondly, mini-languages should evolve with the use of communication, where signals are collectively chosen by the users. Only in such a way is it possible to obtain a process similar to what can be intuitively inferred about the mechanism of language evolution.

Conclusion

The present paper was aimed at demonstrating whether interaction might have played a significant role in human language evolution. In order to test that, two iterated learning experiments were conducted, both of whom utilised the “alien fruit” design. The first one was carried out with individual participants; to be more precise, there was only one individual responsible for producing the signal in each iteration of the chain. On the other hand, in the second experiment, interaction between participants was allowed, as there were three participants in each generation of iterations, all of whom were encouraged to discuss and negotiate the signals produced. Contrary to the initial intuition, it occurred that interaction played a minor role in the evolution of the mini-language produced by the participants of the second experiment. However, it should be noticed that in the first experiment, data that point to stored-language bias in the emergent mini-language were obtained. The results of the second experiment show that the design involving interaction partially eliminated the ELS bias. When the participants were encouraged to discuss all the options to fulfil their task correctly, the evolving mini-languages were not based on or very similar to entrenched linguistic structures, allowing for a process more similar to what language evolution could have looked like than in the case of the first non-communicative experiment.

References

- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2, 177–226.
- Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*, 41(4), 892–923.
- Christiansen, M. H., & Kirby, S. (2003). *Language evolution: The states of the art*. Oxford: Oxford University Press.
- Cornish, H. (2011). *Language adapts: Exploring the cultural dynamics of iterated learning* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1842/5603>
- Dor, D., Knight, C., & Lewis, J. (Eds.). (2014). *The social origins of language*. Oxford: Oxford University Press.
- Feher, O., Smith, K., Ritt, E., & Wonnacott, N. (2016). Communicative interaction leads to the elimination of unpredictable variation. *The Evolution of Language: Proceedings of the 11th International Conference (EVO LANG 11)*. Retrieved from <http://evolang.org/neworleans/papers/137.html>
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Science*, 4, 258–267.
- Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). Cambridge: Cambridge University Press.
- 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 of the United States of America*, 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.
- Knight, C. (2000). The evolution of cooperative communication. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 19–26). Cambridge: Cambridge University Press.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklady*, 10, 707–710.
- Reboul, A. (2015). Why language really is not a communication system: A cognitive view of language evolution. *Frontiers in Psychology*, 6, 1434.
- Rogalska, K. (2015). Investigating glossogeny via the iterated learning methodology: The effect of entrenched linguistic system(s) in human agents. In P. P. Chruszczewski, R. L. Lanigan, J. R. Rickford, K. Buczek, A. R. Knapik, & J. Mianowski (Eds.), *Languages in Contact 2014*

- (pp. 219–232). Wrocław-Washington, D.C.: Wydawnictwo Wyższej Szkoły Filologicznej we Wrocławiu.
- Rogalska, K. (2016). *Linguistic bias in the cultural evolution of language observed via iterated learning experiments* (Doctoral dissertation). Toruń: Nicolaus Copernicus University.
- Schouwstra, M., Motamedi, Y., Smith, K., & Kirby, S. (2016). Improvisation, interaction and iteration in the evolution of linguistic structure. *Proceedings of EvoLang XI, Language Adapts to Interaction Workshop*. Retrieved May 16, 2016, from http://evolang.org/neworleans/workshops/papers/LATI_2.html
DOA
- Scott-Phillips, T. (2015). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Basingstoke: Palgrave Macmillan.
- Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9(4), 371–386.
- Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43C, 57–68.
- Waciewicz, S. (2013). Ewolucja języka – współczesne kontrowersje. In P. Stalmaszczyk (Ed.), *Metodologie językoznawstwa I. Ewolucja języka. Ewolucja teorii językoznawczych* (pp. 11–26). Łódź: Wydawnictwo Uniwersytetu Łódzkiego.