Combinatorial structure and iconicity in artificial whistled languages

Verhoef, T.; Kirby, S.; de Boer, B.

[Link to publication](#)

Combinatorial structure and iconicity in artificial whistled languages

Tessa Verhoef (t.verhoef@uva.nl)
ACLC, University of Amsterdam
1012 VT, Amsterdam, Netherlands

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

Bart de Boer (bart@ai.vub.ac.be)
AI lab, Vrije Universiteit Brussel
1050 Brussels, Belgium

Abstract

This article reports on an experiment in which artificial languages with whistle words for novel objects are culturally transmitted in the laboratory. The aim of this study is to investigate the origins and evolution of combinatorial structure in speech. Participants learned the whistled language and reproduced the sounds with the use of a slide whistle. Their reproductions were used as input for the next participant. Cultural transmission caused the whistled systems to become more learnable and more structured. In addition, two conditions were studied: one in which the use of iconic form-meaning mappings was possible and one in which the use of iconic mappings was experimentally made impossible, so that we could investigate the influence of iconicity on the emergence of structure.

Keywords: iterated learning; iconicity; combinatorial structure; phonology; cultural evolution; duality of patterning

Introduction

Duality of patterning, one of Hockett's (1960) basic design features of language, has recently received increased attention (de Boer, Sandler, & Kirby, 2012). This feature describes how, in speech, a limited number of meaningless sounds are combined into meaningful words and those meaningful words are combined into larger constructs. How this feature emerged in language is currently still a matter of debate, but it is increasingly being studied with the use of a variety of different techniques such as computer simulations and laboratory experiments. In one of these laboratory experiments, the emergence of combinatorial structure was studied through transmission of artificial whistled languages in the laboratory (Verhoef, Kirby, & Padden, 2011). For this study the experimental iterated learning paradigm (Kirby, Cornish, & Smith, 2008) was used. Iterated learning refers to the process of cultural transmission, in which individuals acquire a social behavior by observing the performance of others who also acquired it from observation (Kirby et al., 2008). The results demonstrated that sound systems, when passed on in a transmission chain, adapt to cognitive biases and become easier to learn (Verhoef et al., 2011). Combinatorial structure emerged and the whistled systems became more efficiently coded over generations (Verhoef, 2012). These results demonstrated a possible route towards the emergence of structure in the sounds of speech. The findings challenge the hypothesis that Hockett (1960) introduced when he linked the emergence of structure to vocabulary expansion and signal dispersal. Even in the case where only a small set of sounds is transmitted and the signal space is not maximally used, combinatorial structure emerges (Verhoef et al., 2011; Verhoef, 2012). In that experiment the influence of semantics was controlled for and the signals did not refer to any concrete meanings. Obviously, in natural human languages meanings are important and the role of semantics in the evolution of linguistic structure should not be ignored (Schouwstra, 2012; Dingemanse, 2012). Would the introduction of semantics influence the emergence of combinatorial structure at the level of phonology? In this article a new experiment is presented in which artificial whistled languages are culturally transmitted and the whistled signals refer to meanings.

Combinatorial structure versus iconicity

Arbitrariness was another design feature Hockett (1960) listed for language. This feature refers to the arbitrary mapping between words and their meaning. Hockett uses the words 'whale' and 'microorganism' as an example: 'whale' is a short word for a large animal, while 'microorganism' is the reverse. It has been argued that non-arbitrariness is rare in modern languages and that it is irrelevant for understanding linguistic structure (Newmeyer, 1992). More recently, however, researchers began to realize that non-arbitrary form-meaning mappings may be more widespread than initially thought. When exploring beyond Indo-European languages, non-arbitrariness seems to play a role in many languages (Imai, Kita, Nagumo, & Okada, 2008; Dingemanse, 2012). This involves classes of words where for instance the shape, complexity, sound or some other characteristic of the meaning expressed is mimicked or iconically represented in the word. Examples have been identified as 'ideophones', 'mimetics' or 'expressives' and the phenomenon is often called sound symbolism.

Sound symbolic mappings in language have been connected to cross-modal mappings in the human brain (Simner, Cuskley, & Kirby, 2010; Ramachandran & Hubbard, 2001). There appear to be many cognitive biases in cross-modal perception that are shared by most people. The bouba/kiki ef-

fect is one famous example that shows a strong preference to relate sharp shapes to the name 'kiki' and round shapes to the name 'bouba' (Ramachandran & Hubbard, 2001). Such shared biases have been argued to play an important role in the evolution of language, by forming a starting point for the initial emergence of grounded speech (Ramachandran & Hubbard, 2001).

A newly emerging sign language, Al-Sayyid Bedouin Sign Language (ABSL), has recently been described, in which the evolution of duality of patterning can still be observed (Sandler, Aronoff, Meir, & Padden, 2011). Even though ABSL is a fully functional and expressive sign language, its combinatorial structure appears to be less discrete (Sandler et al., 2011). Could it be the case that this young sign language was able to survive up to now without duality of patterning because the manual modality allows for much iconicity and the language is learnable and transmissible enough without phonological structure? When most mappings are transparent, they may be more intuitive and easier to remember as holistic entities. On the other hand, it has been shown that there is actually an advantage for arbitrary mappings in acquiring word meanings in context (Monaghan, Christiansen, & Fitneva, 2011). A secondary objective of the experiment described below is to investigate how iconic form-meaning mappings influence the emergence of combinatorial sub-lexical structure. The experiment is similar to the experiment described by Verhoef et al. (2011) and Verhoef (2012), but with meanings attached to the whistled signals. Two conditions were studied: one in which the use of iconic form-meaning mappings is possible and one in which the use of iconic mappings is experimentally made impossible. This was expected to provide insight into the possible role of iconicity in the emergence of duality of patterning since it could reveal whether a situation that allows for more iconicity, can 'survive' without the emergence of combinatorial structure longer.

In summary, the objective of this study is as follows. First and foremost, we investigate whether the addition of meanings leads to a result that is similar to what was found in the whistle experiment without meanings, to see if combinatorial structure also emerges in the presence of semantics. Second, we search for any differences between the two conditions to see whether iconicity could cause a delay in the emergence of structure.

Method

In this experiment participants were asked to learn and reproduce whistled signals with a slide whistle (see figure 1) as names for objects they saw on a computer screen. There were twelve whistled signals in the training set in total.

The meanings were unusual objects that look like possible mechanical parts, but they are novel objects for which we do not have words in existing languages. The objects were a subset from those created by Smith, Smith, and Blythe (2011) that were slightly modified. To reduce the structure in

Figure 1: Slide whistle

the meaning space, all objects were transformed to blue tone and could therefore not be grouped by their color. They also did not share shapes or parts and were not structured in any other obvious way. This was needed to limit the emergence of semantics-related compositional structure. A few examples of objects that were used are shown in figure 2.
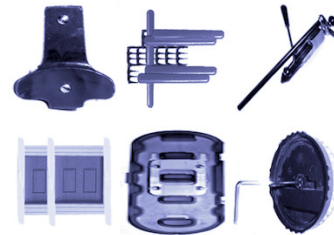

Figure 2: Examples of novel objects used in the experiment

Following the paradigm of experimental iterated learning (Kirby et al., 2008), the last whistle sounds that a participant produced for each object were used as the input given to the next participant. However, this is the point where the two conditions differ from each other. In one condition, the 'intact' transmission, the next participant was exposed to the output of the previous participant exactly as it was produced. The mapping from whistled signals to objects was kept intact. In the other condition, the 'scrambled' transmission, the output of the previous participant was altered before it was given to the next person. The produced form-meaning mappings were broken down by scrambling the mappings at each change of generation and by using a different set of objects between consecutive generations. In this way, if there were any iconic relations to emerge in the sets, it would only be helpful for the participants in the first condition and any semantics-related structure is broken down in between the transmission steps. Only the signal sets themselves stay intact.

Procedure

During the experiment participants completed three rounds of learning and recall, which were alternated by 'guessing game' rounds. In the learning phase the objects and their corresponding whistle were presented one by one in a random order, and participants recorded an imitation of the whistle. In the recall phase a panel was shown with a button for each object and the participant had to choose each of the objects once to record the right whistle for it from memory. The guessing phases were introduced to encourage people to keep paying attention to the mapping between whistle sounds to objects. In this phase the whistles were played one by one in a random order and for each whistle the participant had to choose

the right object from a panel. This was done with half of the whistle-object pairs after the first learning phase and with the other half after the second. The whistles from the last recall phase were used as training input for the next participant, depending on the condition either with intact or scrambled whistle-object mappings. Transmission was continued from person to person until there were eight generations in each chain and four chains per condition.

Initial input sets

Two separate initial whistle sets were constructed. Each set was used as the starting point for half of the chains in each condition. The whistles were taken from a database of whistles that were collected during a pilot study. During this pilot, people were asked to freely record a number of whistles. The two initial sets were constructed so as not to exhibit combinatorial structure. To achieve this, the entropy measure for quantifying combinatorial structure from Verhoef (2012) was used. Sets of twelve whistles were generated randomly from the database until two sets were found with no overlap, that had a comparable and relatively high measured entropy.

Participants

In total 64 participants took part in the experiment. They were divided over eight transmission chains, four in each condition. Participants were recruited from the University of Amsterdam community through posters and e-mail invitations. All participants were between the ages of 19 and 41 years old, 43 were female and 21 male. In each chain either two or three men participated. They were compensated for their time with a cash payment of 10 euros.

Qualitative results

This section describes qualitative observations as a first impression of the data. The internal structure of the whistle sets is investigated as well as the role of iconicity.

Internal structure in whistle sets

On the level of the signals, independent of the objects they refer to, it can be observed that structure develops in a manner that is very similar to what could be observed in the experiment without meanings. Like in that experiment, whistles were introduced that were clearly related in some way to the form of some whistles that already existed in the set. Mirrored versions, combinations of existing whistles, repetitions of the same pattern within a whistle or whistles with similar shapes but different whistle manners for instance appeared. Figure 3 shows an example from one of the chains in the scrambled condition. Here, at generation four, two whistles were in the set that followed approximately the same shape in pitch contour (down and up), but were whistled in a different manner. One of them was whistled in a smooth and unbroken fashion and the other was more staccato-like and broken into pieces. In generation five, one half of each of these whistles is borrowed and reused to form a new whistle. The left part
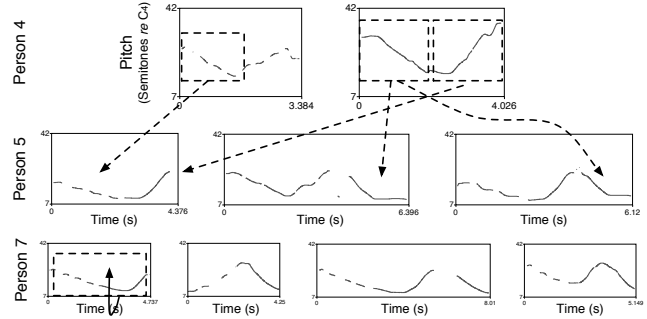

Figure 3: Development of structure in a chain from the scrambled condition. See main text for an explanation.

of the smooth whistle is also reused and combined with existing whistles. In later generations, these are reproduced and all kinds of other variations on this appear, such as ones that are mirrored again as a whole.
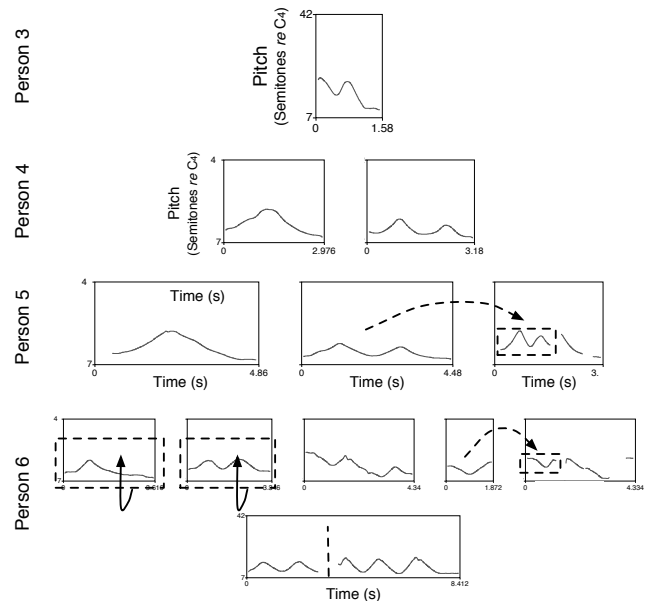

Figure 4: Development of structure in a chain from the intact condition. See main text for an explanation.

Figure 4 shows an example from one of the chains in the intact condition. In this example one whistle from generation three seems to be the inspiration for two new whistles in the next generation: one with one 'bump' and another with two. In generation five the 'two bump' whistle is started to be reused and combined with another pattern and in generation six both the one bump and two bump whistles are being reused, mirrored and recombined more widely. An existing whistle with several up and down movements is even segmented into two parts, where the first part is again the two bump whistle.

To examine the final result of these gradual changes in the chains, we can look at the set of whistles produced by the eighth and last participant in a chain. Figure 5 shows a frag-
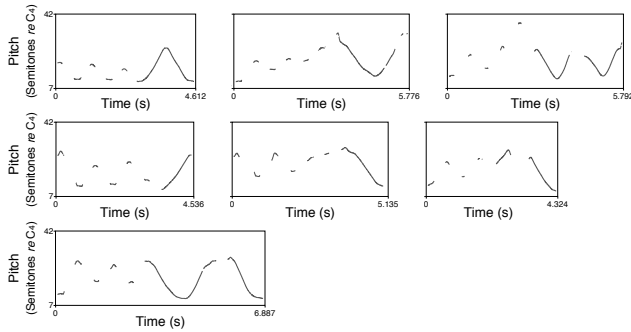
Figure 5: Fragment from the whistle set produced by the last participant in a chain from the scrambled condition.

ment of such a set from the scrambled condition and here we can identify a clear combinatorial structure. There is a set of building blocks (short level notes, falling-rising slides, rising-falling slides, etc) and these are reused and combined in a systematic way to create the whistles in the set. For some of the whistles, there is another version that is mirrored vertically and a pattern of short notes of alternating pitch height seems to be a recurring theme. The set has become constrained as well, for instance in terms of the complexity of the falling-rising patterns and the overall variation in the type of building blocks that are left.

Segmenting whistles into building blocks

As compared to the whistle experiment described in Verhoef (2012), the emergence of discretization is not always the same in the current study. In the previous experiment, the silences (or pauses in the air stream) in the whistles were used as indicators for where one segment ended and another one began. In the current study this appears not to be the only obvious manner in which discretization can be observed. Here, we also find structures that are combinatorial, but non-sequential or sequential without silences. Figure 6 shows an example, where the same whistle shape, or movement, is reused several times, but each time with some parts realized in a different whistle manner (broken or smooth). This observation is taken into account in the quantitative analysis described in the next section.
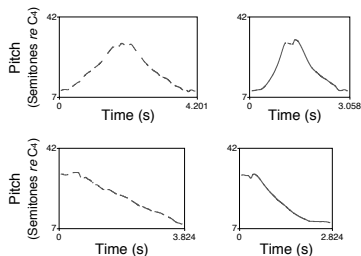

Figure 6: A structure where recombination is not solely sequential.

Iconic whistle-object mappings

When talking about mappings between whistle sounds and objects one may wonder how a whistle sound can iconically depict a visual object. This is difficult to identify as an outside observer, since iconicity is subjective and depends on experience and individual history. However, some examples could be found in the form-meaning pairs in the current data and iconicity could take several different forms in these examples. Most often, the shape of the whistle, or the pitch contour, would mimic certain features in the object. This could for instance be the overall shape of the object, the orientation of the object or the amount of visually distinctive parts on the object. It needs to be noted though that these are subjective observations and that it is not necessarily the case that the participants would be aware of these structural similarities. Judging from the observations, iconic mappings were not found to be widespread throughout the whole experiment. Figure 7 shows a few examples of clearly iconic form-meaning mappings that were encountered.
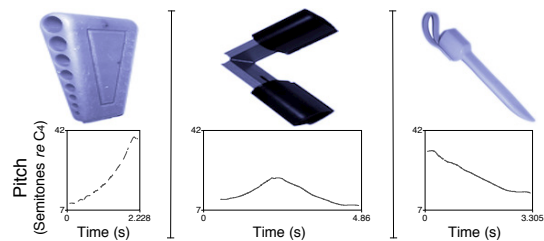

Figure 7: Examples of iconic whistle-object pairs in the data.

Participants filled out a post-participation questionnaire in which they were asked to describe their specific strategy (if any) for recall and whether they thought the whistles and objects fit well together. Often participants reported strategies in line with the observations described in the previous paragraph. Other strategies that were reported involved: imagining how the object would sound and linking this with the whistle, imagining how the object would move and linking the pitch contour with that, or linking the object with some real object they know and linking the whistle with the sound that object would make. These reports further illustrate the subjectivity of form-meaning resemblance.

In summary, the structures that emerged in the sets of whistled signals resemble the discrete and combinatorial structure that emerged in the experiment without meanings (Verhoef et al., 2011; Verhoef, 2012). Qualitatively, no difference could be observed between the structures in the two conditions and iconicity did not seem to play a large role.

Quantitative results

This section describes a quantitative analysis to assess whether the observed patterns are consistent across the data. First, the learnability is investigated by computing how well participants were able to recall the set of whistle-object pairs.

Then, the development of combinatorial structure is measured.

Recall error

To find out whether the sets of whistle-object pairs became easier to learn and reproduce, we measured how well participants were able to recall the right whistle for each of the objects. The recall error was measured by comparing each whistle that a participant produced for an object with the whistle linked to that specific object in the input. To determine the distance between two whistles, the same distance measure was used as the one used by Verhoef (2012). This measure compares plunger movement tracks (pitch tracks converted to plunger displacement) with the use of derivative Dynamic Time Warping (Keogh & Pazzani, 2001).

Figure 8 shows the data for this measure of recall error for the four chains in both conditions, with increasing generations on the horizontal axis. The mean over the four chains for each condition and standard error are plotted.

A linear mixed effects analysis was performed (with *lme4* in R) to explore the effect of generation on the recall error, with an intercept for chain as random effect. Likelihood ratio tests of this model against a null model excluding the effect of generation revealed a significant trend affecting error ($\chi^2(7) = 14.08, p = 0.0498$), decreasing it by about 0.18 ± 0.056 (standard errors) from generation to generation for the intact condition, as well as for the scrambled condition ($\chi^2(7) = 27.25, p = 0.0003$) with a decrease of about 0.1 ± 0.025 (standard errors). This suggests that there is an increase in the reproducibility of the form-meaning pairs over generations in both conditions.

blocks that are more widely reused and combined. One adjustment had to be made to the measure as it was described in Verhoef (2012). Based on the qualitative observation that there was clearly no one 'right' segmentation that could be used to describe the discretization of the signal space, three different types of segmentation were defined. The whistles were segmented in all three ways and the entropy was computed for each of the three sets of basic building blocks that resulted from the segmentations. The lowest entropy value that was measured was then considered to be the best minimal description length approximation and was used as the measure for (dis)order. The first type of segmentation used silences as segment boundaries. The second type used the minima and maxima in the plunger movement track as segment boundaries and the third used the points of minimal and maximal velocity.

Figure 9 shows the development of entropy for the four chains in both conditions, where 0 refers to the initial whistle set. Again, the mean over the four chains and standard error for each condition is plotted.

A linear mixed effects model was constructed in the same way as for recall error, but with entropy as the test variable. A significant trend ($\chi^2(7) = 19.73, p = 0.006$) of decreasing entropy by about 0.69 ± 0.20 (standard errors) from generation to generation for the intact condition was found, as well as for the scrambled condition ($\chi^2(7) = 17.22, p = 0.016$) with a decrease of about 0.52 ± 0.19 (standard errors). These findings imply that the process of iterated learning in both conditions caused structure to emerge in the sets of whistles that refer to meanings.
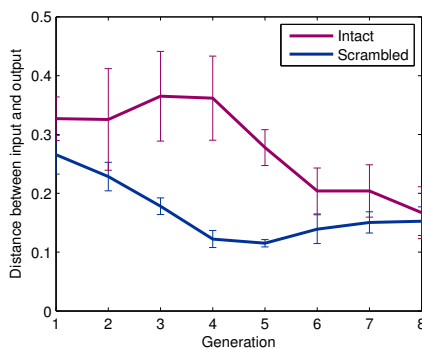

Figure 8: Recall error on the whistle-object pairs over generations in both conditions, showing the mean and standard error. Recall error decreases significantly in both conditions.
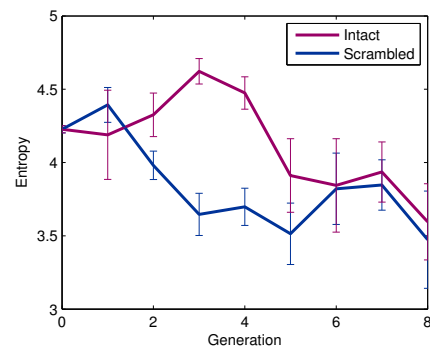

Figure 9: Entropy of the whistle sets over generations in both conditions, showing the mean and standard error. Entropy decreases significantly in both conditions.

Combinatorial structure

To investigate whether the sets of whistles gradually become more structured over generations, the entropy measure that was used by Verhoef (2012) was applied to the current data. This measure makes use of the notion of entropy from information theory and is based on the idea that a set with more combinatorial structure is composed of less basic building

Transparency

Although it is difficult to assess the actual role iconicity played in the two conditions, the results of the guessing game phases could indirectly reveal a potential influence. If the mappings were more transparent in the intact condition, we would expect participants in that condition to score higher on the identification task after only very little exposure to the

data. A linear mixed effects analysis was performed to explore the effect of condition on the scores, with round number (half of the items appeared in a guessing game round after the first exposure to the data and the other half after the second exposure) as fixed effect and intercepts for chain and generation as random effects. Likelihood ratio tests of this model against a null model excluding the effect of condition showed that condition does not affect performance in the guessing game ($\chi^2(1) = 0.210, p = 0.647$). This could suggest the role of iconicity was minimal in both conditions, or at least did not play a large enough role in the intact condition to boost identification scores. However it needs mentioning that the participants had been exposed to the data before the guessing game phases, which is expected to have influenced the scores.

Discussion

The experiment presented in this article shows that cultural evolution in the laboratory causes a system of whistled words for novel objects to become more learnable and more structured over time. This work expands a previous finding that showed the same result for whistled systems without meanings (Verhoef et al., 2011; Verhoef, 2012). For two different situations, one with transmission of intact form-meaning pairs and one with scrambled pairs, we showed that the transmitted whistled systems develop from holistic towards having discrete and combinatorial structure. Sets of building blocks are efficiently reused and combined, similar to the structures of speech. In addition to the data presented by Verhoef et al. (2011), the current data forms another example to show that the emergence of combinatorial structure is not necessarily driven by vocabulary expansion and dispersal as was proposed by Hockett (1960).

As a secondary objective we explored the tension between combinatorial structure and possible iconic mappings. It appeared that the potential for iconic mappings did not prevent the emergence of structure in this experiment. However, when looking at the development of entropy in the two conditions, we can see that the main 'drop' in entropy in the intact condition took place approximately from generation four to eight, while in the scrambled condition, this was sooner, approximately from generation one to five. This could hint at a slight delay in the emergence of combinatorial structure in the intact condition, which would follow the expectation, but the current data from the guessing game phase does not suggest a large influence of iconicity. A more detailed analysis is needed. When we have a better picture of the actual development of iconicity, the development of structure can be linked to it more directly. Finding an objective measure for quantifying iconicity in the data is not trivial however and this will be part of our future continuation of this research.

To conclude, this article provides additional evidence to show that combinatorial structure in language can emerge through cultural evolution. The influence of iconicity in this evolutionary process still needs to be investigated in more depth, but with this study we provide an experimental plat-form that can be used to tackle this issue. In the future, additional experiments are expected to lead to more insights.

References

de Boer, B., Sandler, W., & Kirby, S. (2012). New perspectives on duality of patterning: Introduction to the special issue. *Language and Cognition*, *4*(4), 251–259.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, *6*(10), 654–672.

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

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.

Keogh, E., & Pazzani, M. (2001). Derivative dynamic time warping. In *the 1st siam international conference on data mining (sdm-2001), chicago, il, usa*.

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

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology-General*, *140*(3), 325.

Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language*, *68*(4), 756–796.

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

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

Schouwstra, M. (2012). Semantic structures, communicative strategies and the emergence of language. *Lot Dissertation Series*, *312*.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? cross-modal mappings across gustation and audition. *Perception*, *39*(4), 553.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cognitive Science*, *35*, 480–498.

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

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