Advances in Complex Systems
Vol. 15, Nos. 3 & 4 (2012) 1150021 (20 pages)

DOI: 10.1142/S0219525911500214

LANGUAGE DYNAMICS IN STRUCTURED FORM AND MEANING SPACES

BART DE BOER* and TESSA VERHOEF

Amsterdam Center for Language and Communication,
University of Amsterdam,
Spuistraat 210 Amsterdam, 1012 VT, Netherlands
**b.g.deboer@uva.nl*

Received 10 May 2011
Revised 7 November 2011
Accepted 8 November 2011
Published 13 March 2012

This paper reviews how the structure of form and meaning spaces influences the nature and the dynamics of the form-meaning mappings in language. In general, in a structured form or meaning space, not all forms and meanings are equivalent: some forms and some meanings are more easily confused with each other than with other forms or meanings. We first give a formalization of this idea, and explore how it influences robust form-meaning mappings. It is shown that some fundamental properties of human language, such as discreteness and combinatorial structure as well as universals of sound systems of human languages follow from optimal communication in structured form and meaning spaces. We also argue that some properties of human language follow less from these fundamental issues, and more from cognitive constraints.

We then show that it is possible to experimentally investigate the relative contribution of functional constraints and of cognitive constraints. We illustrate this with an example of one of our own experiments, in which experimental participants have to learn a set of complex form-meaning mappings that have been produced by a previous generation of participants. Theoretically predicted properties appear in the sets of signals that emerge in this iterated learning experiment.

Keywords: Evolution of speech; compositional structure; experimental iterated learning.

1. Introduction

Categorical perception [22] is central to human language. Both linguistic forms (e.g. [8]) and meanings (e.g. [3]) are perceived categorically. Nevertheless, underlying these categories are signal spaces and semantic spaces with rich structure. This means that some of the categories are more alike than the others. To illustrate this for signals: the sound [f] of English "fought" is closer to the [θ] of "thought" than it is to the [b] of "bought". For meanings, the color red for example is closer to orange than it is to blue. Even at the abstract level of syntax, it is clear that categories are not simple, abstract entities: some words, for example, are more prototypical of their syntactic category than others. "Stone", for instance is a very typical noun, and "quickly" is a very typical adverb, but "yesterday" falls somewhere in between.

For understanding the evolution and emergence of language and of languages, it is important to take into account the structure of the signal and meaning spaces. The exact nature of these spaces determines which signals and which meanings can or cannot easily be confused. A lot of this structure is much older than *Homo sapiens*. Smith and Lewicki [60] show that cat auditory neurons appear optimized for speech, and explain this by arguing that speech has adapted to properties of general mammalian hearing. Similarly, Berlin and Kay [3] argue that the distribution of basic color terms they find in human languages can be explained by evolutionary much older properties of primate color vision. On the other hand, Martínez *et al.* [41] argue on the basis of fossil evidence that human hearing has evolved to adapt to speech while Steels and Belpaeme [69] argue that color terms in language cannot be explained by reference to primate color vision alone. The relation between properties of the signal and meaning spaces on the one hand, and cultural and biological properties of language on the other is therefore not straightforward.

Although linguists and cognitive scientists are aware of these issues, quite often a choice is made to abstract away from the complexities of the signal and meaning spaces and to take discrete categories as given. This is especially true in the study of language dynamics which already has to deal with complexity at the individual and collective levels. This approach has been very successful in both computational and mathematical modeling [5, 30, 31, 37, 44, 66, 67] and experiments with humans [17, 32, 58, 65], but we aim to show that taking into account the complexities of signal and meaning spaces is both feasible and leads to surprising results.

We will discuss the consequences of structured form and meaning spaces on language emergence and evolution. We start with a discussion of theoretical issues, and present a formalization of what we mean by structured form and meaning spaces. On the basis of this formalization, we show that in the case of structured form, only a limited number of discrete signals can be distinguished because of the possibility of confusion. This implies that for expressing an unlimited number of meanings, combinatorial structure in the signal space is needed. Another effect is that the signals that are used will become dispersed over the available signaling space in order to minimize the probability of confusion. Interestingly, dispersion may lead to the beginning of combinatorial structure in the form space.

We will also show that in structured form and meaning spaces, it is possible to have unlimited productivity without recombination of discrete symbols. However, this is only possible when the topology of the meaning space is a subset of that of the form space. If there is no such match (and this is generally the case when communicating using speech) a conventionalized discrete system is necessary. In order to be productive, such a system needs to use combinatorial structure. Although using combinatorial structure in the *meaning* space probably requires cognitive adaptations, it turns out that in certain structured *form* spaces, combinatorial structure may emerge from optimization for distinctiveness without specialized cognitive adaptations.

Although much of linguistic structure may be the result of optimization of linguistic function, it is more than likely that humans do have specialized cognitive adaptations for dealing with language. The recent trend to stress the importance of domain-general cognition over language-specific cognition [6, 15] focuses its critique on highly specialized adaptations proposed in some current linguistic theories (e.g. [1]). However, the adaptations we expect are to aspects of language that are theoretically expected to occur in any sufficiently complex communication system, such as the ones mentioned above. Examples of such cognitive adaptations are the tendency to perceive continuous input categorically (both in form and meaning spaces), the ability to deal with combinatorial structure in both form and meaning space and the ability to learn many arbitrary form-meaning mappings.

Fortunately, this is not just a philosophical question, but one that can be tested experimentally. We will discuss a number of ways in which this can be done, focusing briefly on the study of emergence of new sign languages, and more in depth on the experimental cultural learning paradigm [20, 32, 58, 63]. We will present an example of an experiment to investigate the emergence of combinatorial structure in a form space as a case study.

2. Theoretical Issues

We focus on communication systems in which the meaning and/or form spaces are structured in the sense that some pairs of forms or pairs of meanings can be more readily confused than other pairs. What this means exactly and what its consequences are for the dynamics of communication systems is set forth in this section.

2.1. *Basic formalisms*

A useful definition of the value v of a communication system is:

$$v = \sum_{\forall m_r \in M} \sum_{\forall m_s \in M} v(m_r|m_s)p(m_r|m_s)p(m_s), \tag{1}$$

where M is the set of possible meanings, m_s is the sent meaning, m_r is the received meaning, $p(m_r|m_s)$, the transmission matrix, is the probability of receiving meaning m_r when meaning m_s was sent, and $v(m_r|m_s)$, the value matrix, is the value of meaning m_r being received when meaning m_s was intended. It should be noted that there are a number of assumptions in this definition: it is only valid for one sender–receiver pair, or for a population in which all senders and all receivers are equal. It also assumes that the value of a meaning and its probability of being misinterpreted only depend on the meanings themselves, not on the context, or on which meanings came before. These assumptions are not necessarily valid for human language or for animal communication systems, but for the purpose of illustrating the issues that this paper addresses, they are acceptable.

Given that meanings are not transferred directly, but are transferred from speaker to hearer in a separate signal space, the transmission matrix can be calculated as follows:

$$p(m_r|m_s) = \sum_{\forall\, f_r, f_s \in F} p(m_r|f_r)p(f_r|f_s)p(f_s|m_s), \tag{2}$$

where F is the space of possible signals (forms), $p(f_s|m_s)$ is the send-matrix, $p(f_r|f_s)$ the confusion matrix and $p(m_r|f_r)$ is the receive matrix.

In the context of the study of language evolution, send and receive matrices were first used by Hurford [27] and similar formulations have been used in many other papers [37, 45, 61, 66, 76]. The use of a value matrix with off-diagonal elements (such that receiving the wrong message can still be valuable) was probably first modeled in the context of language evolution by Zuidema and Westermann [78].

It should be noted that both the form and the meaning space can be continuous (even though the notation of (1) and (2) suggests discrete spaces, but this is purely for convenience). All objects that are referred to as matrices can also be seen as functions with two continuous parameters.

2.2. *Optimal communication strategies*

In general, value of communication will be optimized when only one form is coupled with every meaning in the send matrix and only one meaning is coupled with every form in the receive matrix. More formally:

$$p(f|m) = \begin{cases} 1, & \text{if of all } f \in F, \quad f \text{ is most likely perceived as } m, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$$p(m|f) = \begin{cases} 1, & \text{if of all } m \in M, \quad m \text{ most likely produced } f, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

(this is true intuitively, as such a communication system concentrates all communicative effort where it is most likely to succeed). In optimal communication systems, mappings between forms and meanings are therefore *not* stochastic but deterministic. However, it is possible that when there is a lot of noise, in the optimal system certain meanings that are very rare or have very little value cannot be communicated at all, in order to create more bandwidth for more frequent or more valuable meanings.

Some observations can be made about how signals can be mapped to meanings in an optimal way. A trivial case occurs when all *forms* are equivalent: then it does not matter which forms are mapped to which meanings. A similarly trivial case occurs when all *meanings* are equivalent (all their *a priori* probabilities and all perception and confusion values are equal) and the number of forms is equal to the number of meanings: then the form to meaning mapping can again be arbitrary.

Things become more interesting when neither all forms nor all meanings are equivalent or when there are more forms than there are meanings to express. When neither forms nor meanings are all equivalent, an optimal communication system will couple the most important meanings (the ones that are most frequent and have the highest value) with the most robust signals (the ones that are most likely to be transmitted correctly). An optimal system will also couple meanings that are close (in the sense that if they are confused, they still have relatively high value) with forms that are close (in the sense that they are likely to be confused). In other words, there will be a topology-preserving mapping between forms and meanings. This was shown experimentally by Zuidema and Westermann [78].

When the number of forms is greater than the number of meanings, a subset of forms must be chosen such that the value of the communication system is maximized. In general this is not a simple matter of choosing the best forms (the forms that have the highest probabilities of being transmitted as themselves) because one has to take into account how forms are confused with each other. The simple strategy of choosing the best forms only works if the probability of confusing a form with another form is constant for each form. The confusion matrix then has the form:

$$p(f_r|f_s) = \begin{cases} p_r & f_r = f_s, \\ \dfrac{1-p_r}{|F|-1} & f_r \neq f_s. \end{cases} \tag{5}$$

(this is for example the form of the confusion matrix as used by Nowak *et al.* [44]). In the more general case, the probability that a form is confused with other forms is not constant. In an optimal communication system, the forms that are used will be minimally confused with each other. In addition, forms associated with important meanings will be less easily confused with other forms than forms associated with less important meanings. The probability of discovering the optimal set of forms is far from trivial and has been the subject of considerable research both in linguistically based form spaces [11, 29, 38, 39, 48, 57] and in more abstract form spaces [45, 56, 77].

Differences between real signals, which exist in the continuous physical world (in the auditory or visual domains, for example), are often expressed in terms of a distance function. Confusion probabilities are then modeled by shifting signals with a noise distribution. The search for distance functions that properly model human linguistic behavior is an important aspect of both phonetic science [4, 38, 40] and speech recognition [25]. However, it must be kept in mind that these distance functions are an abstraction of the confusion probabilities between signals. Only the probabilities can be observed experimentally. It is not even guaranteed that a confusion matrix can be modeled by a distance function. This is, for example, impossible if the confusion matrix is not symmetric. Confusion matrices for real stimuli are indeed not symmetric [9].

2.3. *Consequences for form-meaning mappings*

The first consequence of the above observations is that the exact choice of forms only makes a difference for the value of a communication system if the different forms behave differently. For real signals, this is always the case: if physical signals are used, some signals are just more similar to each other than to other signals. In certain computer models, such as those in [27, 66, 68, 70, 74], no difference is made between the forms, but this can be an acceptable simplification if one is not really interested in the exact forms that are used.

The second consequence is that if the topologies of the signal space and the meaning space are similar, an optimal communication system will reflect this. This correspondence between form and meaning spaces is like iconicity, except that it does not need to reflect any other property of the meaning except the topological correspondence. To illustrate this point, we can look at the alleged correlation between perceived pitch of vowels in words for expressing size [43]. Words with vowels whose perceived pitch is high ("teeny") are used to express smallness, while words with low pitch ("huge") are used to express largeness (of course this correlation is not perfect). This relation is iconic, as resonance frequencies of small objects are higher than those of large objects. But the correspondence also works because both size and pitch have the same topology (they are one-dimensional, continuous scales). If topology matching were the only criterion, however, low frequencies could equally well correspond to small objects and high frequencies to large objects. Topology correspondence therefore appears to be a necessary condition, but not a sufficient condition for iconicity.

Correspondence between topology in form and meaning can also occur in more than one dimension. Bee dances are an example. Bees communicate heading and distance to a food source by performing a "waggle dance" in which angle with the vertical corresponds to angle with the sun and duration of the waggle to distance [75]. Both forms correspond topologically to the meanings they express. Whether bee dances are examples of combinatorial structure depends on the definition one uses, but in any case they do not use recombination of *discrete* units.

An important property of form-meaning mapping where the topologies match is that it is possible to express meanings and understand signals that have not been expressed or observed before. In this sense a form-meaning mapping with matching topologies can be considered *productive*. If there is no match between the topology of the form space and the meaning space, an iconic mapping between the two spaces becomes more problematic. If the dimensionality of the meaning space is smaller than that of the form space, a subspace of the form space can be used. If the dimensionality of the meaning space is larger than that of the form space, the form space may fold up to cover the meaning space as well as possible.

To the best of our knowledge, such an experiment has not been done yet in the context of communication systems, but the process is very similar in nature to that of representing an input space by a self-organizing map, such as a Kohonen map [35].

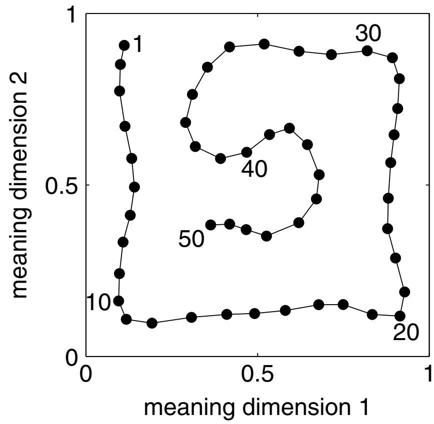


Fig. 1. Example of how a one-dimensional form space can fold up to map onto a two-dimensional meaning space. For illustration purposes, a one-dimensional Kohonen map was used to map a two-dimensional input space. Nodes of the map are indicated with black circles, and their neighborhood relations with black lines. Node numbers in the map are indicated.

It has been observed that if the dimension of the Kohonen map is lower than that of the input space, it will fold up to cover the input space [23]. This is illustrated in Fig. 1. It is expected that the same will happen when a low-dimensional form space is used to express a higher-dimensional meaning space. The mapping between a one-dimensional space and a two-dimensional space, as shown in Fig. 1, illustrates how such a mapping is less iconic than a mapping between spaces with equal dimensionalities. Although for parts of the trajectory there is a relation between the position along the form space and the position along one dimension in the meaning space, this relation is different for different parts of the form space. However, the mapping between the form space and the meaning space can be re-interpreted as robustly discretizing the meaning space: forms 1–11 express a low value for dimension 1, forms 11–20 indicate a low value along dimension 2, forms 46–50 express a low medium value along dimension 2, etc. Although it is no longer possible to communicate exact values[a] in the meaning space, it is now possible to communicate a discrete number of values robustly. A possible robust mapping is illustrated in Table 1.

For expressing a position in two-dimensional meaning space, a sequence of two forms is now needed. This is a possible avenue toward combinatorial meaning. As for combinatorial *forms*, Zuidema and de Boer [13, 77] have shown that optimization of distance between signals that are extended in time leads to what they call a superficial combinatorial structure. This is combinatorial structure that can be observed by an outside observer, but of which the users of the signals are not aware. Zuidema and de Boer [77] argue that after the superficial combinatorial

[a]It should be kept in mind that if noise is present, communicating exact values is also impossible in the case that form and meaning topologies match.

Table 1. Robust form-meaning mapping based on Fig. 1.

Form prototype	Meaning
6 (1)	Low dimension 1
15	Low dimension 2
24	High dimension 1
32	High dimension 2
36	Low mid dimension 1
40	High mid dimension 2
44	High mid dimension 1
48 (50)	Low mid dimension 2

Note: The meaning space is discretized into four steps along both dimensions, and there is a distance of at least four forms between prototypical forms.

structure emerges in a set of signals, the users of the signals will have an evolutionary advantage when they develop adaptations to productively use the combinatorial structure in perception, learning and production.

2.4. *Consequences for language dynamics*

The form-meaning mappings discussed above add a layer of complexity to language dynamics. The topologies of the form space and the meaning space also need to be taken into account. While mathematical analysis of the emergence of form-meaning mappings in spaces where either all forms are equivalent or the number of forms is equal to the number of meanings has had some success [2, 21, 33, 37], mathematical analysis of more complex form-meaning mappings has to the best of our knowledge not been done, yet. It is a rather intractable mathematical problem, as it is closely related with the sphere packing problem [7]. The sphere packing problem is about how spheres can be fitted into a given space, and how many will fit in a given volume. This is similar to finding the optimal set of signals in a Euclidean signal space such that the minimal distance between the signals is maximized. This can be understood by imagining the signals as the centers of spheres and their radii as the distance along which they can be shifted by noise. If one can find the optimal arrangement of signals, one has solved the sphere packing problem, and if one knows how to solve the sphere packing problem, one can find the optimal arrangement of signals.

The mathematical complexity of complex form and meaning spaces does not mean that they cannot be meaningfully investigated: computational models have been quite successful. Liljencrants and Lindblom [38] have shown that small vowel systems can be predicted by optimization of acoustic distinctiveness. Using an agent-based model, de Boer [10–12] has shown that optimization can be the effect of self-organization in a population, and that population pressure can also lead to stabilization of sub-optimal systems. Oudeyer [49] has used population-based models, in which there is not even an explicit communicative goal, to investigate emergence of discretization and phonemic structure. Nowak *et al.* [46] have used a game-theoretic approach to investigate emergence of combinatorial structure and Zuidema and de Boer [13, 77] have shown that structure can emerge without agents being aware of it. These simple models only work for systems with a small number of signals. Larger systems show regularities that cannot be predicted by simple optimization of acoustic distinctiveness. Schwartz *et al.* [57] have investigated additional perceptual factors and Schwartz *et al.* [56] have investigated the effect of Ohala's maximal use of available distinctive features (MUAF) principle [47]. This principle states that human languages tend to use available distinctions maximally, and therefore reflects cognitive factors. As for semantic spaces, Steels and Belpaeme [69] have shown that distribution of color terms in languages can be explained by a similar process of maximization of perceptual distinctiveness. Interestingly, Schoonhoven *et al.* [55] have shown that at least in signed languages there appears to be a correlation between complexity of the sign and rarity of the color term.

The bottom line of these different lines of research is that there are factors related to perception and production, cognitive factors, communication-theoretical factors and cultural processes that determine what a communication system will end up looking like. Which of these factors have been most important in shaping actual human languages must be determined experimentally. Fortunately, there is now a growing body of research that addresses just this question.

3. Experimental Techniques

Knowledge about individual behavior is fundamental to understanding the interaction between individual cognition and cultural processes. As the rich body of work on individual learning of speech has been thoroughly reviewed elsewhere [24, 36, 50, 51] we will not go into this topic here. An important assumption underlying much work on individual learning of language is that individual biases are directly reflected in universal properties of language. That this relation may be more complex has been made clear by mathematical and computational modeling of iterated cultural learning [5, 21, 30, 33, 76]. This work shows that depending on the exact way in which language learners reproduce the variation that exists in the language they learn, individual biases may or may not be good predictors of the properties of the linguistic conventions in the population. More specifically, if language learners reproduce linguistic variation with the same frequency distribution as with which they have observed it (they are then called *probability matchers*) then individual biases will be reflected straightforwardly in the language. If, on the other hand, they preferably reproduce the utterances that they have observed most frequently (they are then called *maximizers*), then there is a more complicated relation between the properties of the language and the individual learning biases [64]. It is likely that real humans are maximizers to some extent, because they tend to overgeneralize observed patterns and sometimes regularize inconsistent input (e.g. [26]). There are two ways to investigate this hypothesis: one is to look at emergence of new languages, and the other is to test it in a laboratory setting.

Emergence of a new language is the best natural experiment to investigate how the properties of human languages take shape. However, it is a very rare occurrence. Creole languages have formed a number of times in the past (e.g. [42]) and although these are very interesting from a sociolinguistic point of view, and to investigate what happens with syntax, morphology and lexicon, they are perhaps less interesting to investigate the basic properties of speech, as it is assumed that the source languages for creoles already have the same basic properties as all other human languages.

Combinatorial signaling systems do emerge from scratch, however, when new sign languages appear in recently formed groups of deaf people. Two examples of such emergence have been studied: that of Nicaraguan Sign Language [59] and the emergence of Al-Sayyid Bedouin Sign Language (ABSL) [54]. Especially in the latter case, it has been observed that a (signed) language can have complex semantic structure, without clear combinatorial phonological structure [28, 53]. Unfortunately these natural experiments are too rare to allow generalizations, and in addition it is very difficult to control for outside influences. In particular, it is always controversial how much of the emerging structure is due to the influence of already existing sign languages (e.g. [52]).

Fortunately, experiments with cultural learning can be done in a laboratory setting. In this newly emerged paradigm of experimental cultural evolution, human participants learn language-like stimuli and reproduce these. Two variants exist: one in which a group of participants develops a communication system in interaction. This is called social coordination and it models emergence of linguistic norms. In the second variant, participants learn stimuli from a previous generation of participants, and their own productions are learned by the next generation of participants. This is called iterated learning (or sometimes a diffusion chain) and it models language change over the generations. Existing early work as well as methodological issues have been reviewed elsewhere [18, 32, 58, 63].

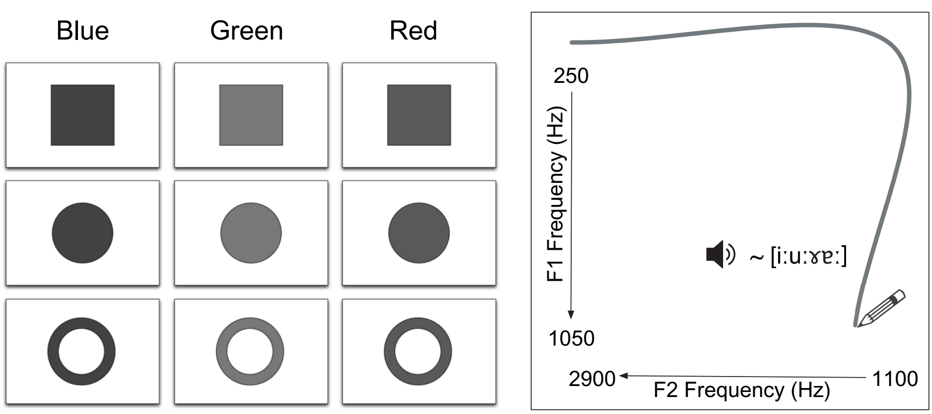
Most of the experimental work has focused on syntactical and morphological issues. Nevertheless, a small number of experiments has focused on how the form of the signals emerges and changes. Most of this work has focused on graphical symbols [14, 16, 19, 20] rather than on acoustic signals. Fay *et al.* [16] have looked at how (graphical) signals become conventionalized and simplified in a horizontal transmission (social coordination) cultural learning experiment based on Pictionary — guessing meanings on the basis of drawings. Garrod *et al.* [20] have contrasted this with what happens in vertical transmission (iterated learning). However, they only looked at complexity, accuracy and convergence, not at combinatorial structure or iconicity. Only del Giudice *et al.* [14] and Galantucci *et al.* [19] have looked at the emergence of combinatorial structure. Del Giudice *et al.* [14] used an iterated learning experiment with transmission of visual signals and showed that combinatorial structure emerges under pressure of articulatory and memory constraints of the learners. Galantucci *et al.* [19] used an interactive game setting in which the visual communication medium had different levels of rapidity of fading. Combinatorial structure was found to be stronger in the signaling systems that emerged in the case of more rapid fading. Below we present an example of an experiment that investigates the emergence of combinatorial structure in acoustic signals. The experimental iterated learning paradigm is not just used to address the issue of the role of individual biases in explaining language universals. It also aims to investigate what strategies humans use to create and extend systems of signals, and to understand how and under what conditions combinatorial structure emerges. As has been pointed out in the previous sections, human languages appear both to maximize distinctiveness of signals and to optimize the economy of the use of distinctive features (the MUAF-principle [47]). Doing experiments with the emergence of systems of signals may help us elucidate what the exact role of both these mechanisms are.

3.1. *Scribble: A case study*

Recently we have made the first steps toward applying the experimental iterated learning paradigm to the emergence of structure in sound systems. In this section, one of these experiments will be presented as a case study, to illustrate the possibilities and challenges for investigating speech. Technical details of the experiment and more detailed results can be found in [72].

3.1.1. *Methods*

The experiment involves iterated learning and reproduction in which each participant learns from the signals that their predecessor has produced. The stimuli consist of meaning–signal pairs. The meaning space is small, discrete, and has a very clear structure in two dimensions: color and shape. Figure 2(a) shows the complete



(a) Meaning space

(b) Scribble to sound mapping

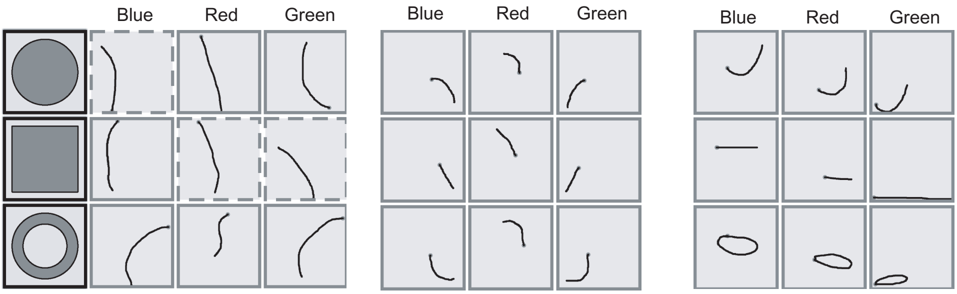
Fig. 2. Visual representation of (a) the meaning space and (b) the signal space.

meaning space. The signal space consists of sounds that are produced by making gestures in a two-dimensional visual space. Participants use the mouse to scribble these gestures and the gesture-sound mapping converts two-dimensional position to a signal with acoustic energy concentrated around two frequencies, resulting in vowel-like sounds. Signals are therefore continuous trajectories (or scribbles) through a vowel space. Such a mapping was used in order to prevent participants from simply writing or drawing conventional signs. Figure 2(b) shows a visual clarification of the scribble-sound mapping. Before the experiment started, a vocabulary was formed by matching each meaning to a random sound (created by generating a random trajectory in scribble space). The first participant had to learn this initial vocabulary and then each succeeding participant, up to ten consecutive generations, learned from the output of the previous participant.

The complete task consisted of three rounds of learning and testing. Each round started with a learning phase in which the participants were exposed to the training set six times, each time in a different random order. In this phase they saw the meaning, heard the sound that labeled this meaning and were given one chance to imitate the sound. Then, in round one and two a short test of five items followed in which only the picture was shown and the participants had to reproduce the right sound from their memory. After the third training phase, a longer test followed which included all nine examples. The signal productions in this last test were used as input for the next participant. Previous work on iterated learning with computer models and laboratory experiments revealed the importance of the transmission bottleneck [32, 62]. When a learner is not exposed to every possible expression during acquisition, there is a transmission bottleneck and often expressions for new items are constructed by generalization over the learned signals. As a result of this, structure has been shown to emerge through cultural transmission [32, 62]. In the case study described in this section, the learning bottleneck was introduced by training the participants on only six out of the total of nine sound–meaning pairs in the learning phase, but testing them in the final test on all nine pairs. The participants were thus not only forced to *learn*, but also to *generalize*. Based on earlier iterated learning work (e.g. [32, 34, 62]), we expected to find an increase of structure toward the end of the transmission chains. We would call this structure combinatorial if it consists of a systematic reuse of basic building blocks in the sounds.

3.1.2. *Results*

The combinatorial structure we expected to find did not emerge in the four chains of the experiment described above. The structure we found appeared to reflect the combinatorial structure of the meaning space, but the forms themselves did not have clear combinatorial structure. Moreover, the structure did not persist reliably over generations. For a discussion on why this happened, see [72]. Here, we focus on the aspects of our results that relate to the issues raised in this article and present



(a) Chain one, generation one (b) Chain two, generation two (c) Chain two, generation four

Fig. 3. (Color online) Example trajectory sets. (a) The dotted borders indicate items that did not appear in training and for which the participant could not have learned the signal. (b) Note that the location of the trajectory indicates the color of the object in the meaning space. (c) Note that the shape of the trajectory appears to express the shape of the object, while the position of the trajectory expresses the color of the object.

a closer analysis of the results to show that in a number of instances, topology mappings like those described in the theoretical section appeared.

Analyzing the results qualitatively, it can be observed that a first step in developing structure often involves similar (or exactly the same) signals being used for objects that share either color or shape. From the very beginning, there seems to be a tendency for participants to search for patterns and apply generalizations. Often features such as the duration of the sound, or the location of the trajectory in the scribble space are linked to colors or shapes in the pictures. For instance in generation one of chain one [Fig. 3(a)], the trajectories that were created for the three unobserved meanings were often based on, or almost the same as the observed ones with identical color or shape. The red square, for instance, gets a trajectory that goes down, as does the red circle and the blue square, while the green square gets a trajectory that goes up, as does the green circle. In chain two, we observe a more developed system by generation two [Fig. 3(b)]. In this system, the location of the scribbles is clearly linked to the colors of the pictures in the meaning space. Red objects are always linked to scribbles in the upper half of the scribble space, green objects are linked to scribbles in the lower left corner and blue objects are linked to scribbles in the lower right corner. This seems to be a first step toward matching the topologies of the signal and the meaning spaces. The meaning space has two discrete dimensions and for one of the relevant dimensions in the meaning space (color), regions creating audibly distinct sounds are used in the signal space. Location in the scribble area has thus been discretized into three regions, which correspond to the different colors.

In generation four of the same chain, this color-location mapping becomes simplified because now the location distinction is made solely on the basis of the vertical dimension. In addition, more structure emerges because the shape of the scribble is also used to make a meaningful distinction between different shapes in the meaning space [Fig. 3(c)]. In this stage both dimensions of the meaning space are matched to a dimension in signal space. The structure that appeared in generation four was learned almost perfectly by the next person, and persisted for the most part upto generation six, in which the structure is learned perfectly and even the sounds created for the meanings that were not in the training set are reproduced correctly, because the participant could generalize from the examples in the learned set. Although structure in the signal–meaning mappings emerged in this way, most of the time it did not persist for long in any chain, because recognizing the signals turned out to be harder than expected.

3.1.3. *Discussion*

The experiment described is an illustration of how the evolutionary dynamics of speech-like signals can be investigated empirically. However, it is also an illustration of the fact that such experiments do not always work as predicted. We did, for instance, not observe the emergence of combinatorial structure. This was due to two factors: first of all, the participants had great difficulties learning how to imitate sounds, and this hindered correct transmission over the generations. Secondly, participants focused mostly on regularities in the meaning space, and made corresponding regularities in the sets of forms they used. Therefore the regularities that could be observed did not show the theoretically predicted [77] emergence of combinatorial structure, but a one-to-one mapping of meaning and form space, as predicted by the theory presented in Sec. 2.3. The experiment and the reasons it did not work out as predicted are discussed in considerable detail in [72].

In order to deal with the problems that characterized this first experiment, we currently focus our efforts on a slightly different approach, as described in [73] and [71], in which participants produce sounds with a more intuitive device than the scribble area, namely a slide whistle. In addition, the absence of meaning in this study allows us to investigate combinatorial structure independent of influences from semantic iconicity or compositionality. In this experiment, combinatorial structure does emerge and this happens in a similar way to what is described by the MUAF-principle [47] mentioned earlier in this article.

4. Conclusion

We have investigated the implications of structured form and meaning spaces for language dynamics. We have shown with Eqs. (1) and (2) that structure can be added to the ordinary mathematical formulation of a signaling game in a straightforward way by using a confusion matrix for the signals and a value matrix for the meanings. This formulation is more general than a formulation using a form (or meaning) space with a distance metric in which signals (or meanings) are shifted randomly: a confusion (or value) matrix can be derived from a distance measure and a noise level, but generally not the other way around.

When either all meanings or all forms are equivalent, any form-meaning mapping where each meaning is mapped to a unique form and vice versa is equivalent, so the structure of the form and meaning spaces can be ignored. If, on the other hand, not all meanings or all forms are equivalent, the structure of the form and meaning spaces determines which form-meaning mappings are best for a robust communication system. An important special case occurs when the topologies of the form and the meaning space are the same (more precisely: when the meaning space is a subspace of the form space). In that case, a direct mapping between the two spaces is possible, and the communication system can be said to be iconic. Such a system turns out to be productive because previously unseen forms can be interpreted and new meanings can be expressed.

When the topology of the form space does not match the topology of the meaning space, a purely iconic productive communication system is not possible. Instead, a conventionalized system must be used. The analysis of such mappings is highly nontrivial, and computer simulations must in general be used. It turns out that mappings that are optimal for robust communication have properties that are found in modern human languages. They show combinatorial structure and the meaning space becomes discretized. Moreover, when using realistic production and perception, resulting sets of signals reflect universals of human sound systems. The degree of realism that can be attained in such simulated systems is limited if one only looks at ease of production and perception. Apparently modern humans also have cognitive adaptations for learning and using repertoires of signals. One of these adaptations is that humans use as small a set of articulatory actions as possible to produce a set of distinctive signals of a given size [47]. The relative importance of functional and cognitive factors can be investigated experimentally through observing the emergence of sets of signals in human groups.

We have argued that although it is possible and worthwhile to investigate this question using data from research into real languages, for instance through the study of emerging sign languages or emerging creole languages, the resulting data is sparse and hard to control. Fortunately, it is also possible to study these questions in a controlled laboratory setting using the recently developed experimental iterated learning paradigm. We have illustrated this procedure using one of our own experiments. In this experiment, participants developed a set of artificial signals using the computer program "Scribble". We have shown that participants do appear to exploit a mapping in topology between the form and meaning spaces. The experiment we presented here in some detail illustrates how theories about (cultural) dynamics of speech can be tested in the laboratory. However, it also serves as a cautionary tale that we do not quite understand the dynamics of the emergence of structure in speech yet. We hope that this will serve as an encouragement to further research within the paradigm of experimental cultural learning.

Our case study has focused on iterated learning (vertical transmission); but we like to stress that for understanding cultural evolution and dynamics, investigation of social coordination (horizontal transmission) is equally important.

Garrod *et al.* [20] find that social coordination leads to simpler and more symbolic signals, whereas iterated learning does not. However, del Giudice *et al.* [14] and Verhoef *et al.* [71, 73] do find that iterated learning leads to structure. Clearly the last word has not been said about these issues.

The theoretical analysis of structured signaling games combined with the new experimental techniques that we have illustrated can help to clarify what cognitive adaptations modern humans have for dealing with the complex form-meaning mappings of human language. Such insight is necessary to understand how these cognitive adaptations can have evolved and how they influence the dynamics of modern language.

Acknowledgments

We thank Simon Kirby, Alex del Giudice, Carol Padden, Jelle Zuidema and Wendy Sandler for helpful discussions and suggestions. This research was funded by NWO Vidi Project 016.074.324 "Modeling the evolution of speech".

References

[1] Baker, M. C., *The Atoms of Language* (Oxford University Press, Oxford, 2002).

[2] Baronchelli, A., Felici, M., Loreto, V., Caglioti, E. and Steels, L., Sharp transition towards shared vocabularies in multi-agent systems, *Journal of Statistical Mechanics: Theory and Experiment* (2006) P06014.

[3] Berlin, B. and Kay, P., *Basic Color Terms: Their Universality and Evolution* (University of California Press, Berkeley, CA, 1969).

[4] Bladon, R. A. W. and Lindblom, B., Modeling the judgement of vowel quality differences, *J. Acoust. Soc. Am.* **69** (1981) 1414–1422.

[5] Brighton, H., Kirby, S. and Smith, K., Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable, in *Language Origins: Perspectives on Evolution*, Tallermann, M. (ed.) (Oxford University Press, Oxford, 2005), pp. 291–333.

[6] Christiansen, M. H. and Chater, N., Language as shaped by the brain, *Behav. Brain Sci.* **31** (2008) 489–558.

[7] Conway, J. H. and Sloane, N. J. A., *Sphere Packings, Lattices and Groups*, 3rd edn. (Springer Verlag, New York, 1998).

[8] Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. and Gerstman, L. J., Some experiments on the perception of synthetic speech sounds, *J. Acoust. Soc. Am.* **24** (1952) 597–606.

[9] Cutler, A., Weber, A., Smits, R. and Cooper, N., Patterns of English phoneme confusions by native and non-native listeners, *J. Acoust. Soc. Am.* **116** (2004) 3668–3678.

[10] de Boer, B., Emergence of vowel systems through self-organisation, *AI Commun.* **13** (2000) 27–39.

[11] de Boer, B., Self organization in vowel systems, *J. Phonetics* **28** (2000) 441–465.

[12] de Boer, B., *The Origins of Vowel Systems*, Studies in the Evolution of Language (Oxford University Press, Oxford, 2001).

[13] de Boer, B. and Zuidema, W., An agent model of combinatorial phonology, *Adapt. Behav.* **18** (2010) 141–154.

[14] del Giudice, A., Kirby, S. and Padden, C., Recreating duality of patterning in the laboratory: A new experimental paradigm for studying emergence of sublexical structure, in *The Evolution of Language*, Smith, A. D. M., Schouwstra, M., de Boer, B. and Smith, K. (eds.) (World Scientific, Hackensack, NJ, 2010), pp. 399–400.

[15] Evans, N. and Levinson, S. C., The myth of language universals: Language diversity and its importance for cognitive science, *Behav. Brain Sci.* **32** (2009) 429–492.

[16] Fay, N., Garrod, S. and Roberts, L., The fitness and functionality of culturally evolved communication systems, *Philos. Trans. R. Soc. London B* **363** (2008) 3553–3561.

[17] Galantucci, B., An experimental study of the emergence of human communication, *Cognitive Sci.* **29** (2005) 737–767.

[18] Galantucci, B. and Garrod, S., Experimental semiotics: A review, *Front. Hum. Neurosci.* **5** (2011) 1–15.

[19] Galantucci, B., Kroos, C. and Rhodes, T., The effects of rapidity of fading on communication systems, *Interact. Stud.* **11** (2010) 100–111.

[20] Garrod, S., Fay, N., Rogers, S., Walker, B. and Swoboda, N., Can iterated learning explain the emergence of graphical symbols? *Interact. Stud.* **11** (2010) 33–50.

[21] Griffiths, T. L. and Kalish, M. L., Language evolution by iterated learning with bayesian agents, *Cognitive Sci.* **31** (2007) 441–480.

[22] Harnad, S. (ed.), *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press, Cambridge, 1987).

[23] Hertz, J. A., Palmer, R. G. and Krogh, A. S., *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity (Addison-Wesley, Redwood City, CA, 1991).

[24] Holt, L. L. and Lotto, A. J., Speech perception as categorization, *Atten. Percept. Psycho.* **72** (2010) 1218–1227.

[25] Huang, X., Acero, A. and Hon, H.-W., *Spoken Language Processing* (Prentice-Hall PTR, Upper Saddle River, NJ, 2001).

[26] Hudson Kam, C. and Newport, E. L., Getting it right by getting it wrong: When learners change languages, *Cognitive Psychol.* **59** (2009) 30–66.

[27] Hurford, J. R., Biological evolution of the Saussurean sign as a component of the language acquisition device, *Lingua* **77** (1989) 187–222.

[28] Israel, A. and Sandler, W., Phonological category resolution: A study of handshapes in younger and older sign languages, in *Cadernos de Sade*, Vol. 2, Special Issue Linguas Gestuais, Castro Caldas, A. and Mineiro, A. (eds.) (UPC, Lisbon), pp. 13–28.

[29] Ke, J., Ogura, M. and Wang, W. S.-Y., Optimization models of sound systems using genetic algorithms, *Comput. Linguist.* **29** (2003) 1–18.

[30] Kirby, S., *Function, Selection and Innateness: The Emergence of Language Universals* (Oxford University Press, Oxford, 1999).

[31] Kirby, S., Natural language from artificial life, *Artif. Life* **8** (2002) 185–215.

[32] Kirby, S., Cornish, H. and Smith, K., Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language, *Proc. Natl. Acad. Sci.* **105** (2008) 10681–10686.

[33] Kirby, S., Dowman, M. and Griffiths, T. L., Innateness and culture in the evolution of language, *Proc. Natl. Acad. Sci.* **104** (2007) 5241–5245.

[34] Kirby, S. and Hurford, J., The Emergence of linguistic structure: An overview of the iterated learning model, in *Simulating the Evolution of Language*, Cangelosi, A. and Parisi, D. (eds.) (Springer Verlag, New York, 2002), pp. 121–148.

[35] Kohonen, T., Self-organized formation of topologically correct feature maps, *Biol. Cybern.* **43** (1982) 59–69.

[36] Kuhl, P. K., Early language acquisition: Cracking the speech code, *Nat. Rev. Neurosci.* **5** (2004) 831–843.
[37] Lenaerts, T., Jansen, B., Tuyls, K. and De Vylder, B., The evolutionary language game: An orthogonal approach, *J. Theor. Biol.* **235** (2005) 566–582.
[38] Liljencrants, J. and Lindblom, B., Numerical simulations of vowel quality systems, *Language* **48** (1972) 839–862.
[39] Lindblom, B., MacNeilage, P. F. and Studdert-Kennedy, M., Self-organizing processes and the explanation of language universals, in *Explanations for Language Universals*, Butterworth, M., Comrie, B. and Dahl, Ö. (eds.) (Walter de Gruyter, Berlin, 1984), pp. 181–203.
[40] Mantakas, M., Schwartz, J.-L. and Escudier, P., Modèle de prédiction du 'deuxiéme formant effectif' f2'-application à l'étude de la labialité des voyelles avant du français., in *Proc. 15th Journées D'étude Sur la Parole* (Société Française d'Acoustique, 1986), pp. 157–161.
[41] Martínez, I., Rosa, M., Arsuaga, J.-L., Jarabo, P., Quam, R., Lorenzo, C., Gracia, A., Carretero, J.-M., Bermúdez de Castro, J.-M. and Carbonell, E., Auditory capacities in middle pleistocene humans from the sierra de atapuerca in spain, *Proc. Natl. Acad. Sci.* **101** (2004) 9976–9981.
[42] Mufwene, S. S., *Language Evolution: Contact, Competition and Change* (Continuum International Publishing Group, London, 2008).
[43] Newman, S. S., Further experiments in phonetic symbolism, *Am. J. Psychol.* **45** (1933) 53–75.
[44] Nowak, M. A., Komarova, N. L. and Niyogi, P., Evolution of universal grammar, *Science* **291** (2001) 114–118.
[45] Nowak, M. A. and Krakauer, D., The evolution of language, *Proc. Natl. Acad. Sci.* **96** (1999) 8028–8033.
[46] Nowak, M. A., Krakauer, D. and Dress, A., An error limit for the evolution of language, *Proc. R. Soc. London* **266** (1999) 2131–2136.
[47] Ohala, J. J., Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation, in *Proc. ICPhS IX*, Vol. 3 (Copenhagen, Institute of Phonetics, 1980), pp. 181–185.
[48] Oudeyer, P.-Y., Coupled neural maps for the origins of vowel systems, in *Proc. Int. Conf. Artificial Neural Networks*, Lecture Notes in Computer Science 2130, Dorffner, G. and Bischof, K. H. (eds.) (Springer Verlag, Berlin, 2001), pp. 1171–1176.
[49] Oudeyer, P.-Y., The self-organization of speech sounds, *J. Theor. Biol.* **233** (2005) 435–449.
[50] Peperkamp, S., Phonological acquisition: Recent attainments and new challenges, *Lang. Speech* **46** (2003) 87–113.
[51] Pierrehumbert, J. B., Phonetic diversity, statistical learning, and acquisition of phonology, *Lang. Speech* **46** (2003) 115–154.
[52] Polich, L., *The Emergence of the Deaf Community in Nicaragua* (Gallaudet, Washington, DC, 2005).
[53] Sandler, W., Aronoff, M., Meir, I. and Padden, C., The gradual emergence of phonological form in a new language, *Natural Language and Linguistic Theory* **29** (2011) 503–543.
[54] Sandler, W., Meir, I., Padden, C. and Aronoff, M., The emergence of grammar: Systematic structure in a new language, *Proc. Natl. Acad. Sci.* **102** (2005) 2661–2665.
[55] Schoonhoven, M. B., de Boer, B. and Pfau, R., Colours on hands: Phonological markedness of sign language color terms, in *10th Conf. on Theoretical Issues in Sign Language Research* (West Lafayette, IN, 2010), pp. 173–175.

[56] Schwartz, J.-L., Boë, L.-J. and Abry, C., Linking dispersion-focalization theory and the maximum utilization of the available distinctive features principle in a perception-for-action-control theory, in *Experimental Approaches to Phonology*, Solé, M. J., Beddor, P. S. and Ohala, M. (eds.) (Oxford University Press, Oxford, 2007), pp. 104–124.
[57] Schwartz, J.-L., Boë, L.-J., Vallée, N. and Abry, C., The dispersion-focalization theory of vowel systems, *J. Phonetics* **25** (1997) 255–286.
[58] Scott-Phillips, T. C. and Kirby, S., Language evolution in the laboratory, *Trends Cogn. Sci.* **14** (2010) 411–417.
[59] Senghas, A., Kita, S. and Ozyurek, A., Children creating core properties of language: Evidence from an emerging sign language in Nicaragua, *Science* **305** (2004) 1779–1782.
[60] Smith, E. C. and Lewicki, M. S., Efficient auditory coding, *Nature* **439** (2006) 978–982.
[61] Smith, K., Brighton, H. and Kirby, S., Complex systems in language evolution: The cultural emergence of compositional structure, *Adv. Complex Syst.* **6** (2003) 1–22.
[62] Smith, K., Brighton, H. and Kirby, S., Iterated learning: A framework for the emergence of language, *Artif. Life* **9** (2003) 371–386.
[63] Smith, K., Kalish, M. L., Griffiths, T. L. and Lewandowsky, S., Introduction cultural transmission and the evolution of human behaviour, *Philos. Trans. R. Soc. London* **363** (2008) 3469–3476.
[64] Smith, K. and Kirby, S., Cultural evolution: Implications for understanding the human language faculty and its evolution, *Philos. Trans. R. Soc. London B* **363** (2008) 3591–3603.
[65] Smith, K. and Wonnacott, E., Eliminating unpredictable variation through iterated learning, *Cognition* **116** (2010) 444–449.
[66] Steels, L., A self-organizing spatial vocabulary, *Artif. Life* **2** (1995) 319–332.
[67] Steels, L., The synthetic modelling of language origins, *Evol. Commun.* **1** (1997) 1–34.
[68] Steels, L., The spontaneous self-organization of an adaptive language, in *Machine Intelligence*, Furukawa, K., Michie, D. and Muggleton, S. (eds.), Vol. 15 (Oxford University Press, Oxford, 1999), pp. 205–224.
[69] Steels, L. and Belpaeme, T., Coordinating perceptually grounded categories through language: A case study for colour, *Behav. Brain Sci.* **28** (2005) 469–489.
[70] Steels, L. and McIntyre, A., Spatially distributed naming games, *Adv. Complex Syst.* **1** (1999) 301–324.
[71] Verhoef, T. and de Boer, B., Cultural emergence of feature economy in an artificial whistled language, in *Proc. 17th Int. Congr. Phonetic Sciences*, Zee, E. and Lee, W. (eds.) (City University of Hong Kong, Hong Kong, 2011), pp. 2066–2069.
[72] Verhoef, T., de Boer, B., del Giudice, A., Padden, C. and Kirby, S., Cultural evolution of combinatorial structure in ongoing artificial speech learning experiments, *CRL Tech. Rep.* **23** (2011) 3–11.
[73] Verhoef, T., Padden, C. and Kirby, S., Cultural emergence of combinatorial structure in an artificial whistled language, in *Proc. 33rd Annual Conference of the Cognitive Science Society*, Carlson, L., Hölscher, C. and Shipley, T. (eds.) (Cognitive Science Society, Austin, TX, 2011), pp. 483–488.
[74] Vogt, P., Anchoring of semiotic symbols, *Robot. Auton. Syst.* **43** (2003) 109–120.
[75] von Frisch, K., Decoding the language of the bee, *Science* **185** (1974) 663–668.

[76] Zuidema, W., How the poverty of the stimulus solves the poverty of the stimulus, in *Advances in Neural Information Processing Systems 15*, Becker, S., Thrun, S. and Obermayer, K. (eds.) (MIT Press, Cambridge, MA, 2003), pp. 51–58.

[77] Zuidema, W. and de Boer, B., The evolution of combinatorial phonology, *J. Phonetics* **37** (2009) 125–144.

[78] Zuidema, W. and Westermann, G., Evolution of an optimal lexicon under constraints from embodiment, *Artif. Life* **9** (2003) 387–402.