Optimal communication in a noisy and heterogeneous environment

Willem Zuidema

Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
and Institute for Cell, Animal and Population Biology
University of Edinburgh
40 George Square, Edinburgh EH8 9LL
Scotland, United Kingdom
jelle@ling.ed.ac.uk
http://www.ling.ed.ac.uk/~jelle

Abstract. Compositionality is a fundamental property of natural language. Explaining its evolution remains a challenging problem because existing explanations require a structured language to be present before compositionality can spread in the population. In this paper, I study whether a communication system can evolve that shows the preservation of topology between meaning-space and signal-space, without assuming that individuals have any prior processing mechanism for compositionality. I present a formalism to describe a communication system where there is noise in signaling and variation in the values of meanings. In contrast to previous models, both the noise and values depend on the topology of the signal- and meaning spaces. I consider a population of agents that each try to optimize their communicative success. The results show that the preservation of topology follows naturally from the assumptions on noise, values and individual-based optimization.

1 Major transitions in the evolution of language

Human languages are unique communication systems in nature because of their enormous expressiveness and flexibility. They accomplish this by using combinatorial principles in phonology, morphology and syntax [8], which impose important requirements on the cognitive abilities of language users. Explaining the origins of the structure of language and the human abilities to process it is a challenging problem for linguistics, cognitive science and evolutionary biology. Mathematical and computational models have been invaluable tools for getting a grip on this problem [11].

Jackendoff [8] has laid out a scenario for the various stages in the evolution of human language from primate-like communication, that reflects a growing consensus and can be summarized with the following "major transitions":

1. From situation-specific signals (e.g. alarm calls), to signals that are non-situation-specific but from a closed class;

2. From (1) to an open, unlimited (learned) class of signals and, subsequently, a phonological combinatorial system;
3. From (1) to the concatenation of signals and, subsequently, the use of ordering of signals to convey semantic relations ("compositionality");
4. From (2) and (3), which constitute the ingredients of a protolanguage, to hierarchical phrase structure and recursion,
5. From (4) to modern language, with a vocabulary for abstract semantic relations, grammatical categories, grammatical functions and a complex morphology.

Presumably, all transitions have greatly increased the number of distinct "signs" (signal–meaning pairs) that can be expressed, transmitted, memorized and learned. Jackendoff argues convincingly that modern languages contain "fossils" of each of the intermediate stages. E.g. the compound noun construction in English can be viewed as a fossil of stage (3): the meaning of words like "doghouse" and "housedog" is deducible (but not completely specified) from the meaning of the component words and the order in which they are put.

Less consensus exists on how the transition from each stage to another could have happened. Some have argued for extensive innate, language-specific cognitive specializations that have evolved under natural selection (e.g. [1, 8]. This is an appealing position, in line with dominant "nativist" theories in linguistics and evolutionary biology. Unfortunately, explanations of this type have generally remained much *underspecified*. Jackendoff admits: *"I will not inquire as to the details of how increased expressive power came to spread through a population [...]. Accepted practice in evolutionary psychology [...] generally finds it convenient to ignore these problems."([8], p. 237)*

Ignoring this problem is an unfortunate tradition. Understanding how innovations can spread in a population is the essence of any evolutionary explanation, and a better end-result is neither a sufficient nor a necessary condition for the spread of innovations. Specifically in the case of language, the spread of innovations is not at all obvious, even if the end-result – when the whole population has adopted an innovation – is demonstrably better, because of two important difficulties that arise from the *frequency-dependency* of language evolution : (i) if only the hearers benefit from communication, it is not clear why speakers would evolve as to give away – altruistically – more and more information [15, 21]; (ii) even if both speakers and hearers benefit, it is not clear how there can be a positive selection pressure on a linguistic innovation if that innovation appears in a population that uses a language without it, and, moreover, how that pressure can be strong enough to prevent it from being lost by drift [6, 3, 21].

A number of researchers have explored the possibilities of general learning and cognitive abilities and cultural evolution explaining the transitions instead (see [20, 11] for reviews and references), or, of cultural evolution facilitating the genetic evolution of linguistic innovations [5, 9]. These models are useful in clarifying the conditions for the "major transitions", but face some new difficulties themselves as well: (i) in many cases, the assumed cognitive abilities are much more language-specific than one would like; (ii) cultural evolution, such as the progressively better structured languages in the "Iterated Learning Model" [10, 2], only takes off when there is already some initial, random structure in the language.

Explaining the evolution of aspects of natural language like combinatorial phonology and compositionality thus remain challenging problems because both the genetic and the cultural evolution explanation require a structured language to be already present in the population before the linguistic innovations can successfully spread in a population. In this paper, I focus on compositionality: the property that the meaning of the whole (e.g. a sentence) is a function of the meaning of the parts (e.g. the words) and the way they are put together. I do not study the evolution of compositionality itself, but explore a possible route for a structured language to emerge without the capacity for compositionality present in the population. That structure is *topology preservation* between meaning-space and signal-space, i.e. similar meanings are expressed with similar signals.

In the next section I present a formalism to describe a communication system where there is noise in signaling and variation in the values of meanings. In contrast to previous models, both the noise and values depend on the topology of the signal- and meaning spaces. In section 3 I present a model of a population of agents that each try to optimize their communicative success under these circumstances. The results, in section 4, show that the preservation of topology between meaning-space and signal-space follows naturally from the assumptions on noise, values and individual-based optimization.

2 A formalism for communication under noisy conditions

Assume that there are M different meanings that an individual might want to express, and F different signals (forms) that it can use for this task. The communication system of an individual is represented with a *production matrix* $\mathbf{S}$ and an *interpretation matrix* $\mathbf{R}$. $\mathbf{S}$ gives for every meaning m and every signal f, the probability that the individual chooses f to convey m. Conversely, $\mathbf{R}$ gives for every signal f and meaning m, the probability that f will be interpreted as m. $\mathbf{S}$ is thus a $M \times F$ matrix, and $\mathbf{R}$ a $F \times M$ matrix. Variants of this notation are used by [7, 14] and other researchers.

In addition, following [13], I assume that signals can be more or less similar to each other and that there is noise on the transmission of signals, which depends on these similarities. Further, I assume that meanings can be more or less similar to each other, and that the value of a certain *interpretation* depends on how close it is to the *intention*. These aspects are modeled with a *confusion matrix* $\mathbf{U}$ (of dimension $F \times F$) and a *value matrix* $\mathbf{V}$ (of dimension $M \times M$). This notation is an extension of the notation in [13], and was introduced in [22].

These four matrices together can describe the most important aspects of a communication system: which signals are used for which meanings by hearers and by speakers, how likely it is that signals get confused in the transmission, and what the consequences of a particular successful or unsuccessful interpretation are. Interestingly, they combine elegantly in one simple expression for the expected payoff w_{ij} of communication between a hearer i and a speaker j [22]:

$$w_{ij} = \mathbf{V} \cdot \left(\mathbf{S}^i \times \left(\mathbf{U} \times \mathbf{R}^j\right)\right) \tag{1}$$

In this formula, "×" represents the usual matrix multiplication and "·" represents dot-multiplication (the sum of all multiplications of corresponding elements in both matrices; the result of dot-multiplication is not a matrix, but a scalar).

A hypothetical example, loosely based on the famous Vervet monkey alarm calls [17], might make the use of this formalism and measure clear. Imagine an alarm call system of a monkey species for three different types of predators: from the air (eagles), from the ground (leopards) and from the trees (snakes). Imagine further that the monkeys are capable of producing a number (say 5) of different sounds that range on one axis (e.g. pitch, from high to low) and tahte these are more easily confused if they are closer together. Thus, the confusion matrix **U** might look like in the left matrix of figure 1.

$U =$

	received signal				
sent signal ↓	$1kHz$	$2kHz$	$3kHz$	$4kHz$	$5kHz$
$1kHz$	0.7	0.2	0.1	0.0	0.0
$2kHz$	0.2	0.6	0.2	0.0	0.0
$3kHz$	0.0	0.2	0.6	0.2	0.0
$4kHz$	0.0	0.0	0.2	0.6	0.2
$5kHz$	0.0	0.0	0.1	0.2	0.7

$V =$

	intentions		
interpretations ↓	*eagle*	*snake*	*leopard*
eagle	0.9	0.5	0.1
snake	0.2	0.9	0.2
leopard	0.1	0.5	0.9

Fig. 1. Confusion and value matrices for the monkeys in the example, describing the noise in signaling and the value of intention–interpretation pairs in their environment.

Further, although it is obviously best to interpret a signal correctly, if one makes a mistake, typically not every mistake is equally bad. For example, if a leopard alarm is given, the leopard response (run into a tree) is best, but a snake response (search surrounding area) is better than an eagle response (run into a bush, where leopards typically hide) [17]. Thus the value matrix **V** might look something like the right matrix in figure 1.

$S =$

	sent signal				
intention ↓	$1kHz$	$2kHz$	$3kHz$	$4kHz$	$5kHz$
eagle	1.0	0.0	0.0	0.0	0.0
snake	0.0	0.0	1.0	0.0	0.0
leopard	0.0	0.0	0.0	0.0	1.0

$R =$

	interpretation		
received signal ↓	*eagle*	*snake*	*leopard*
$1kHz$	1.0	0.0	0.0
$2kHz$	1.0	0.0	0.0
$3kHz$	0.0	1.0	0.0
$4kHz$	0.0	0.0	1.0
$5kHz$	0.0	0.0	1.0

Fig. 2. Production and interpretation matrices for the monkeys in the example, describing which signals they use for which situations.

For any given production and interpretation matrix, we can through equation (1) calculate the expected payoff from communication. Assume a speaker i with its $\mathbf{S}^i$ as the left matrix in fig. 2, and a hearer j with its $\mathbf{R}^j$ as the right matrix in that figure. The expected payoff of the interaction between i and j if the constraints on communications are as in **U** and **V** in fig. 1 is, by proper application of equation (1): $w_{ij} = 0.7 \times 0.9 + 0.2 \times 0.5 + 0.2 \times 0.5 + 0.6 \times 0.9 + 0.2 \times 0.5 + 0.1 \times 0.5 + 0.2 \times 0.9 + 0.7 \times 0.9 = 2.33$

In this simple example, the matrices **U** and **V** are very small, and reflect only a 1-dimensional topology in both signal and meaning space. The matrices **S** and **R** are set by hand to arbitrarily chosen values. In contrast, in the simulations of this paper we will consider larger and more complex choices for **U** and **V**, and we will use a hill-climbing algorithm to find the appropriate (near-) optimal settings for **S** and **R**.

3 Distributed hill-climbing

Based on the measure of equation (1), I use a hill-climbing algorithm to improve the communication. To speed up the simulations, I make the simplification that the values in the **S** and **R** matrices are all either 1 or 0, i.e. they are deterministic encoders and decoders, which can be shown to always perform better than their stochastic versions [18, 16]. Hill-climbing in the simulations reported here is *distributed*, i.e. I simulate a population (size 400) of agents that each try to optimize their success in communicating with a randomly selected other agent (see the author's website for details). Experiments (not reported here) suggest that distributed hill-climbing, although orders of magnitude faster, leads to very similar results as global hill-climbing. In some of the simulations, agents are placed on a grid (size 20×20) and interact only with their direct neighbors (8, except for agents at the edge which have less neighbors), but also in this condition very similar results are obtained.

The motivation for this style of optimization is (i) that it is fast and straightforward to implement; (ii) that it works well, and gives, if not the optimum, a good insight on characteristics of the optimal communication system; and (iii) that it shows possible *routes* to (near-) optimal communication systems, and in a sense forms an abstraction for both learning and evolution.

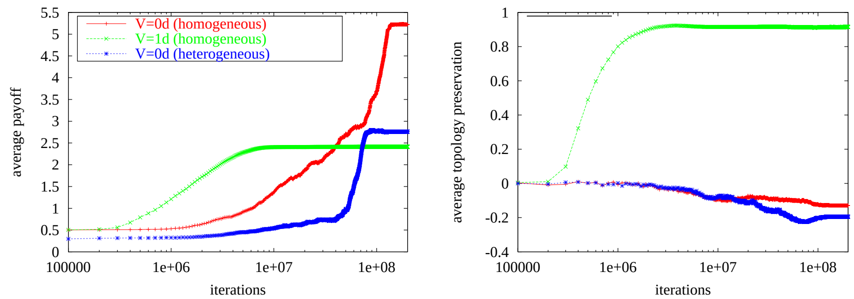
The **V** and **U** matrices can be chosen to reflect all kinds of assumptions about the signal and meaning space. In this paper I vary whether all meanings are equally valuable ($v = 1.0$, labeled "homogeneous"), or get assigned a random value ($0.0 < v \leq 1.0$, labeled "heterogeneous"). I further vary whether or not there is a topology, and if so, of which dimensionality. The diagonal elements of **V** are always v and of **U** always 1.0. Without a topology ("0d"), the off-diagonal elements in **U** or **V** are 0. With a topology, the off-diagonal elements are given by $\mathbf{V}(p,q) = v/(1 + D(p,q))$ and $\mathbf{U}(p,q) = 1/(1 + D(p,q))$, where $D(p,q)$ gives the squared Euclidean distance between the positions of the two meanings or signals i and j. In the 1-dimensional condition ("1d"), the position of a meaning or signal is simply defined as its index. In the 2d condition, the meaning and signal spaces are 2-dimensional surfaces of size $(\sqrt{M} \times \sqrt{M})$ or $(\sqrt{F} \times \sqrt{F})$. The x-coordinate is then given by the largest integer smaller than the root of the index: $x = int(\sqrt{i})$. The y-coordinate by: $y = i$ *modulo* x. After these values are set, the rows of both **U** and **V** matrices are normalized.

I monitor the behavior of the model with two measures. The first is the average payoff, as given by equation (1), averaged over all individuals interacting with all other individuals, both as speaker and as hearer. The second is a measure for the degree of topology preservation between the meaning space and the signal space in the emerging languages. Following [2], I use the correlation ("Pearson's r") between the distance between each pair of meanings and the distance between the corresponding signals:

$$r = \underset{m,m' \in M}{correlation} \left(D\left(m, m'\right), D\left(S[m], S[m']\right)\right), \tag{2}$$

where $S[m]$ gives the most likely signal used to express m according to $\mathbf{S}$.

For 2-dimensional meaning spaces I also visualize the topology preservation by plotting all meanings as nodes in a meaning space, and connecting those nodes where the corresponding signals are (one of maximal 4) neighbors in signal-space.



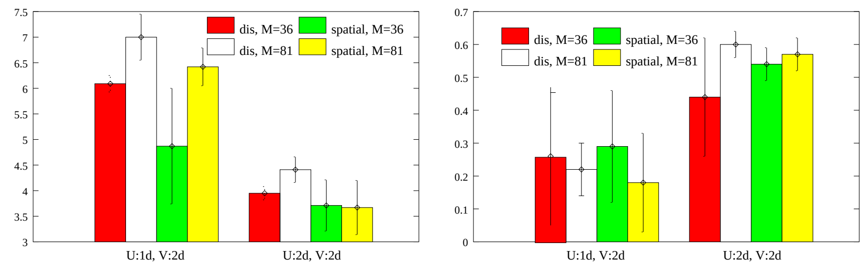
(a) Payoff ($\overline{w}$)   (b) Topology Preservation ($\overline{r}$)

Fig. 3. Average payoff (a) and degree of topology preservation (b) for 2×10^8 iterations under 3 conditions: (1) **V**:0d homogeneous, (2) **V**:1d homogeneous; (3) **V**:0d heterogeneous. The maximum average payoffs that are reached depend on the arbitrary chosen values of the **V** matrices; hence, only the shapes of the curves are important. Common parameters are P=400, M=16, F=49, **U**:1d.

4 Results

Figure 3 shows the average payoff and topology preservation for simulations under 3 different conditions: (i) homogeneous and no topology in the meaning space ("**V**:0d"); (ii) homogeneous and **V**:1d; (iii) heterogeneous and **V**:0d. The results are plotted with a logarithmic x-axis. They show that convergence is more than 10 times faster if there is a topology in the meaning space. Recall that in the topology condition, interpretations with a meaning close to the intention are also rewarded. That fact facilitates establishing conventions regarding which signals to use for which meanings, because it creates more possibilities to break the initial symmetry (when no convention is established, every signal–meaning pair is equally good or bad).

Figure 4 shows the average payoff and topology preservation for 60 simulations where the dimensionality of the signal space is varied, and where hearers are selected randomly from either the whole population ("dis"), or from one of the speaker's 8 neighbors ("spatial"). In all cases, the payoff reaches high levels (when the signal space is 1d) or intermediate levels (when the signal space is 2d and the overall noise-level is consequently higher because each signal has more neighbors). Also, in all cases the topology preservation reaches high levels (when the dimensionalities of meaning and signal space match) or intermediate levels (when the dimensionalities mismatch).



(a) Payoff ($\overline{w}$) (b) Topology preservation ($\overline{r}$)

Fig. 4. Average payoff (a) and degree of topology preservation (b) after 5×10^7 iterations for different parameters. Error-bars indicate standard deviations. Common parameters are P=400, M=36 and **V**:2d heterogeneous.

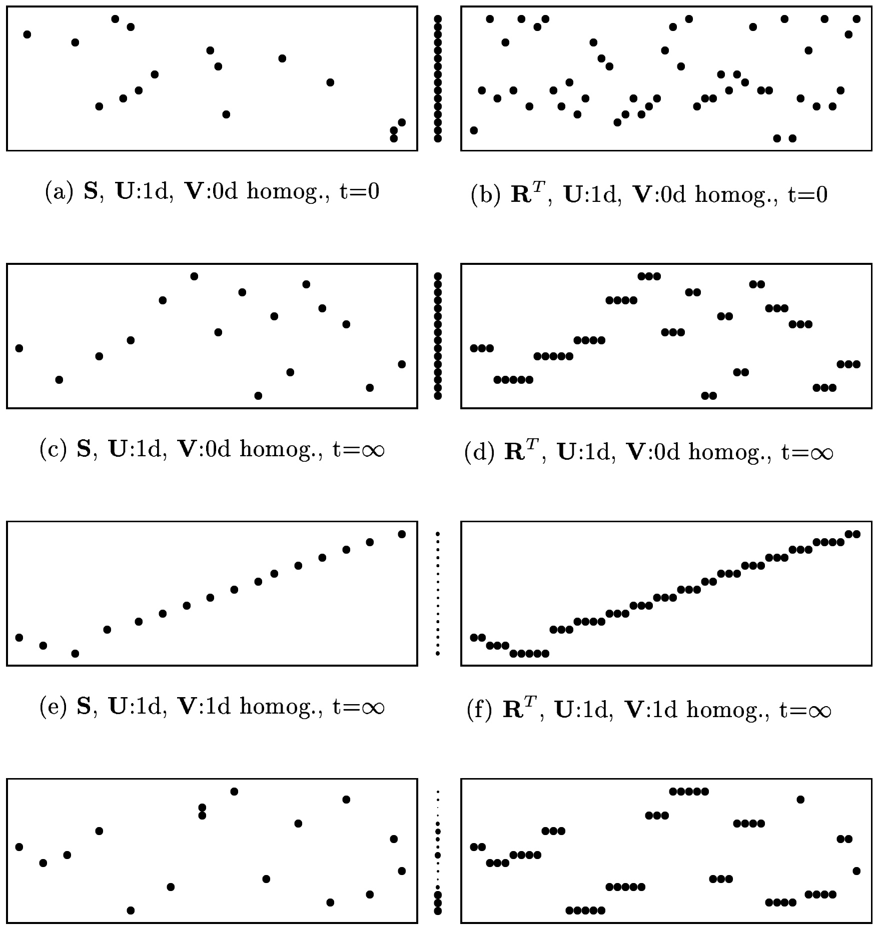
The emerging communication systems are visualized in fig. 5 and 6 and can be summarized with the following properties:

Specificity: every meaning has exactly one signal to express it and vice versa (i.e. no homonyms, and no real synonyms: if different signals have the same meaning they are very similar to each other).

Coherence: all agents agree on which signals to use for which meanings, and vice versa. Specificity and coherence are also found in "language game" models where there is no noise on signaling (e.g. [14, 19]).

Distinctiveness: in the **S** matrices, the used signals are maximally dissimilar to each other, so that they can be easily distinguished (compare figure 5a, at the start of the simulation, with 5c, at equilibrium). In the **R** matrices, clusters of neighboring signals all are interpreted as the same meaning. Typically, the most central signal (except at the edges) in such a cluster is the one that is actually used by the **S** matrix (compare figure 5c with 5d). Distinctiveness is also found in the "imitation game" [4], where no meanings are modeled.

Topology preservation: if there is a topology in both the meaning- and signal-space (as determined by **V** and **U**), similar signals tend to have similar meanings [22]. This preservation is not perfect (there is one major irregularity and several minor ones in the signal–meaning mapping of figure 5e and f. The topology preservation, according to equation (2), is $\overline{r} = 0.915$), but in all simulations performed it is surprisingly high. "Bad" solutions, such as the **S** and **R** of figures 5c



(a) **S**, **U**:1d, **V**:0d homog., t=0

(b) $\mathbf{R}^T$, **U**:1d, **V**:0d homog., t=0

(c) **S**, **U**:1d, **V**:0d homog., t=∞

(d) $\mathbf{R}^T$, **U**:1d, **V**:0d homog., t=∞

(e) **S**, **U**:1d, **V**:1d homog., t=∞

(f) $\mathbf{R}^T$, **U**:1d, **V**:1d homog., t=∞

(g) **S**, **U**:1d, **V**:0d heterog., t=∞

(h) $\mathbf{R}^T$, **U**:1d, **V**:0d heterog., t=∞

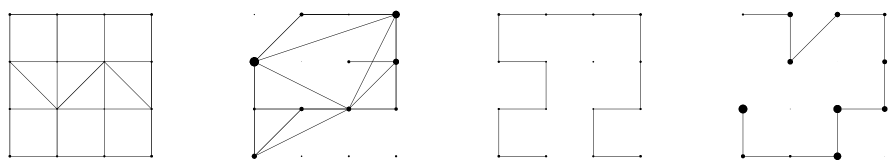
Fig. 5. (a)-(h) Examples of **S** and **R** matrices from the simulations of figure 3. For easy comparison, the **R** matrices are transposed so that in all matrices meanings differ on the vertical axis, and signals on the horizontal axis. Between the matrices the diagonal values of the **V** matrix are plotted, where the diameter of a circle corresponds to value of the corresponding meaning. Common parameters are P=400, M=16, F=49.

and d ($\overline{r} = -0.073$), are stable once established in the population, but have a much smaller basin of attraction. In the case of a two-dimensional meaning space, we can draw plots like figures 6a-d, which show that the topology is almost perfectly preserved if the dimensionalities of the meaning- and signal-spaces match (6a), although it is skewed if different meanings receive very different values

(6b). But even if the dimensionalities do not match, there is a strong tendency to preserve topology as well as possible (6c and d).

Valuable meanings first: When one analyzes the intermediate stages between the random initialization and the equilibrium solutions (not shown here; see author's website), it becomes clear that with a heterogeneous **V** valuable meaning-signal pairs get established first, and change little afterward.

Meanings sacrificed: Finally, when the **V** matrix is heterogeneous (figure 6b and d), or there is a dimensionality-mismatch (figure 6c and d), one can observe that meanings with very low value are sacrificed for the benefit of robust recognition of more valuable meanings (a similar observation was made in [13]). These sacrificed meanings "deliberately" get expressed with a signal that will be interpreted with a meaning that is very close.



(a) **U**:2d, **V**:2d ho., $\overline{r} = 0.67$

(b) **U**:2d, **V**:2d he., $\overline{r} = 0.65$

(c) **U**:1d, **V**:2d ho., $\overline{r} = 0.46$

(d) **U**:1d, **V**:2d he., $\overline{r} = 0.07$

Fig. 6. Topology preservation at equilibrium in 4 simulations with 1d and 2d **U** matrices, and homogeneous and heterogeneous 2d **V** matrices. Nodes are meanings (diameters correspond to value), edges connect neighbors in signal space (several signals can map to a single meaning, such that nodes can have many neighbors; some meanings are not expressed, and the corresponding nodes are not connected). Common parameters are P=400, M=16, F=49.

5 Conclusions

In this paper, I have shown that from simple assumptions about topologies in meaning- and signal-space, and individual-based optimization, communication systems can arise that show a structured mapping from meanings to signals. In a population where such a language is spoken, the fundamental new phenomenon of compositionality can presumably much more easily evolve.

There is no space here to explore the many connection between these simulations and the fields of Information Theory [16] and Evolutionary Game Theory [12]. In a sense, the matrices of figure 5 and 6 describe evolutionary stable strategies, under the constraints of communication over a noisy channel. These connections, and the analytic proofs that can be worked out in these frameworks, will be the topic of future work.

Acknowledgments Funding from the Prins Bernhard Cultuurfonds and a Marie Curie fellowship of the European Commission is gratefully acknowledged. I thank Gert Westermann and Andy Gardner for their remarks.

References

[1] D. Bickerton. *Language and Species.* University of Chicago Press, 1990.
[2] H. Brighton. *Simplicity as a Driving Force in Linguistic Evolution.* PhD-thesis, Theoretical and Applied Linguistics, University of Edinburgh, 2003.
[3] L.L. Cavalli-Sforza and M.W. Feldman. Paradox of the evolution of communication and of social interactivity. *Proc. Nat. Acad. Sci. USA*, 80:2017–2021, 1983.
[4] B. de Boer. Self organization in vowel systems. *J. of Phonetics*, 28:441–465, 2000.
[5] T. Deacon. *Symbolic species, the co-evolution of language and the human brain.* The Penguin Press, 1997.
[6] R.A. Fisher. On the dominance ratio. *Proc Roy Soc Edin*, 42:321–431, 1922.
[7] J. Hurford. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989.
[8] R. Jackendoff. *Foundations of Language.* Oxford University Press, 2002.
[9] S. Kirby and J. Hurford. Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey, editors, *Proceedings 4th European Conference on Artificial Life*, pages 493–502. MIT Press, Cambridge, MA, 1997.
[10] S. Kirby. Syntax without natural selection. In C. Knight et al., editors, *The Evolutionary Emergence of Language.* Cambridge University Press, 2000.
[11] S. Kirby. Natural Language from Artificial Life. *Artificial Life*, 8(2):185–215, 2002.
[12] J. Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, Cambridge, UK, 1982.
[13] M.A. Nowak and D.C. Krakauer. The evolution of language. *Proc. Nat. Acad. Sci. USA*, 96:8028–8033, 1999.
[14] M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for research on language newsletter*, 11(1), 1996.
[15] M. Oliphant. The dilemma of Saussurean communication. *BioSystems*, 37(1-2):31–38, 1994.
[16] J.B. Plotkin and M.A. Nowak. Language evolution and information theory. *Journal of Theoretical Biology*, pages 147–159, 2000.
[17] R.M. Seyfarth and D.L. Cheney. Some general features of vocal development in nonhuman primates. In C.T. Snowdon and M. Hausberger, editors, *Social influences on vocal development*, pages 249–273. Cambridge University Press, 1997.
[18] C.E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:379–423 and 623–656, 1948.
[19] L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language.* Oxford University Press, Oxford, UK, 2002.
[20] L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1:1–35, 1997.
[21] W. Zuidema and P. Hogeweg. Selective advantages of syntactic language: a model study. In Gleitman and Joshi, editors, *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, pages 577–582. Lawrence Erlbaum, 2000.
[22] W. Zuidema and G. Westermann. Evolution of an Optimal Lexicon under Constraints from Embodiment. *Artificial Life*, 2003. (accepted).