On the Relevance of Language Evolution Models for Cognitive Science[*]

Willem H. Zuidema
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels
Belgium

Gert Westermann
Sony CSL-Paris
6, rue Amyot, 75005 Paris
France

Abstract

We argue that Cognitive Science can profit from the study of language evolution. Research in language evolution is concerned with the question of how complex linguistic structures can emerge from the interactions between many communicating individuals. As such it complements psycholinguistics which investigates the processes involved in individual adult language processing, and child language development studies, investigating how children learn a given language. We focus on the framework of *language games* and argue that they offer a fresh and formal perspective on many current debates in Cognitive Science, including those on the synchronic vs. diachronic perspective on language, the embodiment and situatedness of language and cognition, and the self-organization of linguistic patterns. We present a model of lexical dynamics that shows the spontaneous emergence of near-optimal characteristics of a lexicon in a distributed population of individuals. Finally, we analyze the shortcomings of our models and discuss how research in Cognitive Science could contribute to improving them.

Introduction

Cognitive Science has a long tradition of formal and computational models of language processing and language learning. These models generally do not consider multiple individuals in interaction; they are therefore restricted to studying language synchronically. Recent years, on the other hand, have seen a growing interest in *language games*: models of language change and language evolution in populations of communicating individuals. We argue that, although these models have not been very wide-spread in the cognitive science community, they can in fact be considered an integral part of this field.

Cognitive Science can profit from the insights that "language games" offer in several ways. In particular, language games offer a fresh and relatively formal perspective on many heated debates in cognitive science: they explicitly deal with the *diachronic* aspect of language and the origins of linguistic structure rather than the processing and acquisition of language; they offer a precise and concise way of incorporating constraints from *embodiment* into the modeling framework; they are *situated* and study language in its *communicative function*; and finally, they are *dynamical systems* in the mathematical sense of the word, showing *self-organization* in a testable and meaningful way.

In the following we will discuss each of these points in more detail. We will then introduce a formalism to describe language games and present some results from simulations of such games. Finally, we will discuss some of the many ways in which findings from Cognitive Science can be incorporated and used to further advance language game modeling.

In our models we restrict ourselves to the development of a common lexicon between individuals, thus skipping the much more complex and controversial issues in syntax. Nevertheless, we hope to show that language games offer an appealing framework to study other aspects of language as well. Language games that do incorporate grammar are being studied and are starting to yield interesting results (Batali, 1998; Steels, 1998; Batali, 2000; Kirby, 2000).

The models we discuss are necessarily and deliberately simple. We do not intend to provide a scenario for language evolution or to simulate a historical development. Rather, we aim at calling attention to the enormous potential for *spontaneous pattern formation* (self-organization) in populations of individuals that learn language from each other, from generation to generation, under the realistic constraints of hearing, articulating and processing. In some sense the essence of our paper is thus that "many simple interactions can lead to complex patterns" — a cliché in the natural sciences, but still underestimated in Cognitive Science.

Relevance

The models of language evolution that we will consider are *multi-agent models*. There is a population of individuals that talk to each other and learn from each other, and there is a language that as a results changes over time. Individuals in the models have limited production, memory and perception abilities, and they have limited access to the knowledge of other individuals. Language in these models is thus studied diachronically, embodied and situated; results from these models show self-organization. The models evaluate the complex relationship between (i) acoustic, cognitive and articulatory constraints, (ii) learning and development, (iii) cultural transmission and

[*] Submitted to the 23d Annual Meeting of the Cognitive Science Society, Edinburgh 2001; please check http://arti.vub.ac.be/˜jelle for reference

interaction, (iv) biological evolution and (v) the complex patterns that are to be explained: the phonology, morphology, syntax and semantics that are observed in human languages. They are thus directly relevant for debates on such issues in Cognitive Science.

Diachrony & Collective Dynamics

An explanatory theory of (some aspect of) language should ultimately not only explain how it is *implemented* in an individual's cognitive apparatus. It should also explain how the individual *acquires* his or her language from the population's language. And, it should explain how the population *created* that language in the first place.

The latter is not a trivial issue. The origin of the languages that we can study today lies in the interplay between the biological evolution of the human brain and cultural processes of transmitting and adapting of language over generations of language users (Deacon, 1997). Models of language evolution have shown that language creation is neither a automatic consequence of language learning nor of Darwinian evolution (Zuidema & Hogeweg, 2000): for the emergence of language one needs both the proper (though not necessarily language-specific) genetic predisposition and the proper cultural dynamics. In machine learning terminology: one needs the proper *inductive bias* (as any realistic learning algorithm does, Mitchell, 1997), and the proper *collective dynamics*.

Just as studies on language acquisition have brought many new and challenging constraints on linguistic theories, we expect studies on language creation – and the collective dynamics of a population of language learners – to have a similar impact.

Embodiment & Optimality

Human language is a very effective way of conveying information. Many of its characteristics are considered to be near-optimal under realistic articulatory, acoustic, cognitive and communicative constraints. For example, phonologists have argued that the distribution of vowels in languages over the available acoustic space is near-optimal from the point of view of distinctiveness (Liljencrants & Lindblom, 1972).

This near-optimality in some sense counters the "arbitariness of the sign": although forms and form–meaning relations are conventionally established and differ from language to language, not every distribution of them is equally good and equally likely to occur.

That observation immediately leads to the question how languages have become "near-optimal" as they appear to be. The fact that each individual language user optimizes his own language is not a sufficient answer: *optimization at the level of the individual does not necessarily lead to optimality at the level of the population*. Take for instance the well-known case of the prisoner's dilemma: if each prisoner optimizes his personal payoff, the collective dynamics lead inevitably to the worst possible situation where neither of the two prisoners cooperate.

Models of language evolution have addressed this issue by showing particular examples of cases where the collective dynamics lead to an optimal or near-optimal language. E.g., (De Boer, 1999) has shown that a population of individuals with realistic production and perception abilities and the task of imitating each other's vowels, can arrive at the near-optimal vowel systems of (Liljencrants & Lindblom, 1972). This model thus shows a specific example of the role of embodiment in explaining language structure.

Self-Organisation vs. "Blueprint Theories"

Explanations for the phonological and grammatical patterns observed in human languages usually postulate a "blueprint" for these patterns in the cognitive apparatus and genetic code of individuals. Underlying such explanations is a strong intuition that the patterns observed in human language are too complicated too arise "spontaneously". However, an impressive amount of examples shows that intuitions about the causes of complex patterns are often flawed. Mechanisms of spontaneous pattern formation in linguistics remain largely unexplored.

Models of language evolution can help to fill this gap in a formal, testable and understandable way. They offer a fresh perspective on the recurring nurture–nature debates, by helping to specify in what way aspects of language are innate or acquired. It might for instance very well be that children use grammatical rules in their speech without ever having encountered them. But such rules don't need to be hard-wired in an infant's genome, if one can show that they are a consequence of the interactions between the infant's brain structures, its (innate) perceptual and motoric machinery, and its physical and cultural environment (MacWhinney, 1999).

Language Games

The most basic communication model consists of a sender, a message and a receiver. Language game models can be viewed as an extension of this basic model, by considering a *population* of individuals ("agents") that can both send and receive. A language game then is a linguistic interaction between 2 or more agents that follows a specific protocol and has varying degrees of success. The types of models that we will consider have the following components: (i) a linguistic representation, (ii) an interaction protocol, and (iii) a learning algorithm.

Linguistic Representation

With "representation" we mean here a formalism to represent the linguistic abilities of agents, ranging from recurrent neural networks (Batali, 1998) or rewriting grammars (Kirby, 2000) to a simple associative memory (Hurford, 1989; Steels, 1996; Oliphant & Batali, 1996; De Boer, 1999; Kaplan, 2000). In the model described in this paper, we will use a simple list of "associations" between linguistics forms (words) and their meanings. Each association has a score that represents the cost (or inversed strength) of that association and guides the choice between associations if several candidates are considered in a certain situation. Lower scores are preferred over higher ones. E.g. if we have the associations $\langle f1, m1, 0.1\rangle$ and $\langle f2, m1, 0.6\rangle$, then the form $f1$ will be uttered if meaning $m1$ needs to be expressed.

In this paper, forms and meanings remain abstract. Other researchers (e.g. Steels, 1998; Batali, 2000) have chosen more concrete representations, such as random strings for forms (e.g. "gugige", "esebodu"), and functional or logical expressions for meanings (e.g. [YCOORD $>$ AVERAGE] for "high", or [$\exists x$ goose(x) sang(x)] for "a goose sang"). However, in these models there are in general no similarity relations between forms and between meanings in the lexicon; i.e. all forms and all meanings have the same distance to each other. Therefore, the form–meaning associations are completely arbitrary (however, associations are not arbitrary in the grammatical expressions of Batali, 2000).

In stead, we assume that there are varying degrees of similarity between forms and between meanings. I.e. there is a topological space of meanings, and a topological space of forms. For the sake of simplicity, in our simulations we choose a 2-dimensional continuous form space and a 1-dimensional discrete meaning space. Adding such a similarity metric is only a first step towards more cognitive plausibility, but already brings fundamental new behaviors.

Interaction Protocol

The agents in the models interact following a simple protocol. In all models two agents are chosen at random. One acts as a speaker or initiator, the other as a hearer or imitator. In the "imitation game" (De Boer, 1999), the initiator chooses a random form from its repertoire and utters it. The imitator then chooses the form from its own repertoire that is closest to the received form and utters it. If the iniator finds that the closest match to this form is the form that it originally used, the game is successful. Otherwise it is a failure. In the imitation game meanings play no role. It serves as a model system for studying the interaction between forms, and the emergent maximisation of the distance between them.

In the "naming game" (Steels, 1996), the meanings do play a role. The speaker chooses a meaning and a form to express that meaning, and the hearer makes, based on the received form, a guess of what is meant. The hearer then receives feedback on the intended meaning, i.e., whether its guess was correct. The game is a success if the speaker's intention and the hearer's interpretation are the same, and a failure otherwise. The naming game serves as a model system for studying the emergence of conventional form–meaning associations and is used for the model in this paper.

In a variant of the naming game, the meaning of the expressed form is immediately available to the hearer (such as in situations where the speakers points at the object that is the topic of a conversation). This variant has been used by most language game models studied so far (e.g. Hurford, 1989; Steels, 1996; Oliphant & Batali, 1996; Batali, 1998; Kirby, 2000; Kaplan, 2000; Batali, 2000).

Learning Algorithm

The learning algorithm that agents use to improve their linguistic abilities is in most models very simple. Most of the algorithms can be considered variants of "stochastic hill-climbing": given a present state of the system a random variation (*mutation*) is tried out. If the performance is better than before, this variation is kept (*selected*), and otherwise it is discarded. For stochastic hill-climbing one has to specify the possible mutations and the quality measure (selection).

In order to be able to try and evaluate many variations at the same time, it is assumed that the different form–meaning associations are in principle independent from each other. Thus, after each interaction, the scores s of the used associations are updated based on the success or failure of that interaction. We use the following update rule, based on (Batali, 2000):

$$\Delta s = \begin{cases} +\beta & \text{in case of failure} \\ -\beta \cdot s & \text{in case of success} \end{cases} \qquad (1)$$

β is a parameter that determines the speed of adaptation (here: $\beta = 0.1$). Associations that are not used often enough are removed, and associations with bad scores are seldomly used. The learning rule therefore implements the selection step of the learning algorithm.

The mutations in the present model occur when an agent has (i) no form associated with a meaning m that needs to be expressed, or (ii) no meaning associated with a form f that is received, and (iii) after every interaction. In case (i) and (ii) a new association is added to the repertoire with the required m or f, a random new form or new meaning and initial score α ($\alpha = 1.0$). In case (iii) every association with a score $s < \alpha$ has a small probability to be duplicated with a small amount of Gaussian noise added to its meaning and form space coordinates. Mutations (i) and (ii) bias the learning algorithm to consider in the first place meanings and forms that are used by other agents. Mutation (iii) allows agents to find better associations, once an approximately correct one is found.

The Optimal Lexicon

We can analyze the model that was outlined above and first derive what would be the "optimal lexicon", i.e. the lexicon that leads to the highest communicative success in the population. To do so, we need a measure for communicative success. Such a measure is presented next; a similar formalism was used in (Hurford, 1989; Nowak & Krakauer, 1999; De Jong, 2000, and other papers). The next step then is to evaluate numerically if the *collective dynamics* can lead to such an optimal situation.

We denote with $S^i(f|m)$ the probability that an agent i uses form f to express meaning m. Similarly, $R^i(m|f)$ is the probability that agent i as a hearer interprets form

f as meaning m. S and R are functions of the lists L of associations of all agents in the population. We assume that there is a finite number $|M|$ of relevant meanings and a finite number $|F|$ of used forms. Further, we assume that there are similarity relations between these meanings and between these forms (i.e. a topology), and that there is some uncertainty about the hearer perceiving the correct form (more similar forms are more easily confused). We denote with $U^i(f^*|f)$ the probability that agent i perceives form f as form f^* (f can be equal to f^*).

Finally, we assume that the communication is successful if the hearer's interpretation equals the sender's intention. The probability of successfully conveying a certain meaning thus depends on the probabilities that the sender uses certain forms and the probabilities that the hearer perceives and interprets these froms correctly.

From these observations, we derive a simple formula that describes the expected success C_{ij} in the communication between a speaker i and a hearer j:

$$C_{ij} = \sum_{m}^{|M|}\sum_{f}^{|F|}\sum_{f^*}^{|F|} S^i(f|m) \cdot U^j(f^*|f) \cdot R^j(m|f^*) \qquad (2)$$

From here it is only a small step to define the communicative success of the whole population of N agents:

$$C = \sum_{i}^{N}\sum_{j \neq i}^{N} C_{ij} \qquad (3)$$

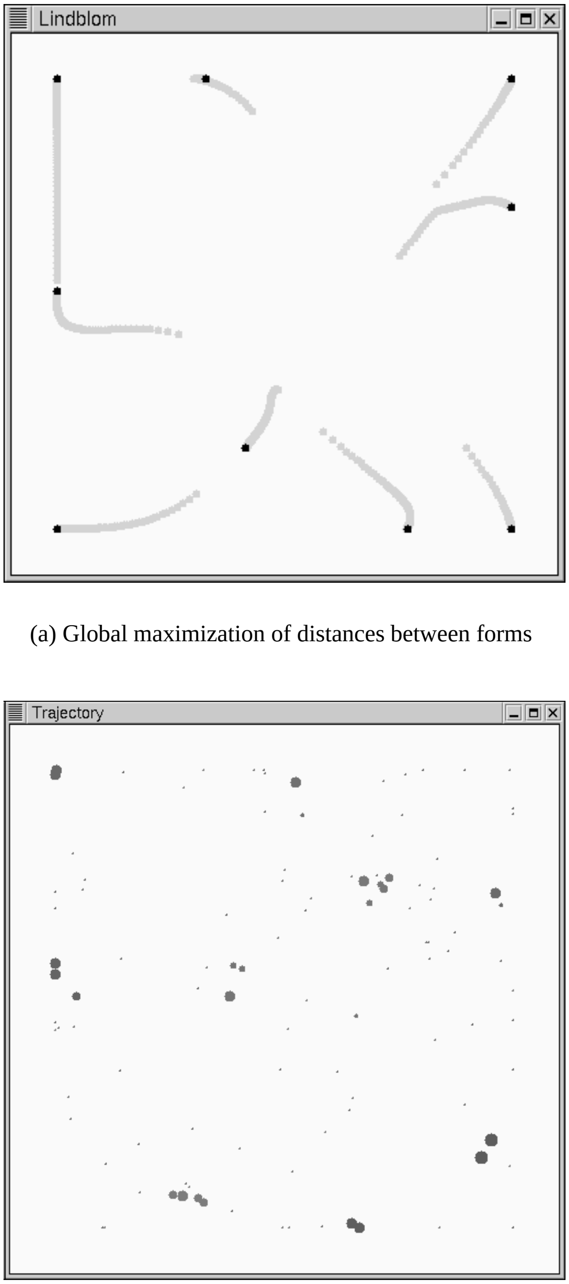
From this formula we can derive under which conditions the communicative success is maximal. Without a formal proof, we state that this is the case if the following conditions hold (provided that $|F| \leq |M|$, and that the U-values are relatively low):

specificity: every meaning has exactly one form to express it, and every form has exactly one interpretation (i.e. no homonyms or synonyms).

distinctiveness: the used forms are maximally dissimilar to each other, so that they can be easily distinguished.

sharedness: all agents use the same forms for the same meanings.

Computational simulations show that close approximations of each of these three properties of the optimal lexicon result from the local interactions that we have defined. Figure 1a shows the trajectories and the final pattern formed with 9 forms in a 2-dimensional form space, randomly initialized, where the distance between the forms is maximized through a simple global heuristic. Figure 1b shows a pattern formed through local interactions between two communicating agents, expressing 9 different meanings with forms from a 2-dimensional form space.



(a) Global maximization of distances between forms

(b) Local interactions: emergence of distinctiveness, sharedness and specificity

Figure 1: *(a) Maximally dispersed forms in a form space, obtained through global stochastic hill climbing (like Liljencrants & Lindblom, 1972). (b) Dispersed forms in form space, obtained through local interactions between communicating agents. Each of the 9 clusters in this figure shows associations from both agents for one particular meaning. Large dots are strong association. (Parameters: 2 agents, 9 meanings, perceptual noise 10%, duplication probability 0.1%, modification 3%)*

If we assume a simple extension of the model – a flux of agents – we can add a fourth criterion. New agents that come into the population should acquire the lexicon of the population as quickly as possible. In general, learning a mapping between two spaces is easiest if there is a regularity in the mapping, and hardest if the mapping in completely random:

regularity: the mapping between meanings and forms shows regularity, such that new agents can generalize from few samples and quickly acquire the lexicon.

Also this property of the optimal lexicon can be obtained in a distributed system. We will first discuss some possible steps towards more cognitive plausibility, and then mention briefly some preliminary results from a variant of the model described here.

Towards more cognitive plausibility

The models we have discussed have been simple and many crucial cognitive details have been left out. In this section we discuss how research in Cognitive Science can contribute to formal models of language evolution and lead to the eventual integration of language evolution research into the Cognitive Science domain.

Some of the contributions we seek concern rather fundamental issues: the questions of how meanings are represented and what a plausible similarity metric is; how forms are perceived and what a plausible similarity metric is for forms; how individuals generalize from few examples; how memory limitations influence the acquisition words.

In the following we will discuss two specific issues from Cognitive Science that have already in part been incorporated into our models.

Cooperativity

Recent research on natural language pragmatics has focused on language as a cooperative phenomenon where communication is viewed as a *joint action* between the participants (Clark, 1996). This view is in contrast to the traditional approach in which speaking and hearing are investigated in isolation as *individual actions*.

This research could be usefully applied to the language games models. An important principle in line with this view of human communication has been formulated by Grice (1975) as the *Principle of Cooperation*: In a conversation, the speaker makes certain assumptions about the expectations of the hearer, and she uses these assumptions to communicate her intended message effectively. This principle involves the provision of enough but not too much information in a message, the relevance of the message to the current conversation topic, and the truthfulness of the information provided. In interpreting the message, the hearer relies on the speaker to have obeyed these principles.

In the context of language game models, we can extend this principle to the cooperative creation of new words: a new form should only be created if no form for the intended meaning already exists. How can this observation be used for improving the language games? In the present language games the speaker creates a new form when he does not have a form for the object to be named, even though the hearer might already have a form for this objcet. In this sense a naming game is not cooperative: both agents know nothing about each other's knowledge and do not make any assumptions, and thus their communication does not conform to Grice's cooperative principle.

In a cooperative setting where both agents take each other's knowledge into account to improve communication, the speaker and hearer could agree on a new name. By querying the hearer for a possible form, the speaker allows himself to make assumptions about the beliefs of the hearer and therefore to engage in a cooperative language game. Such an extension of the language game algorithm is plausible because it views language as a cooperative phenomenon and as a means to maximize the efficiency of communicating intended meanings. It will prevent the creation of an excess of new forms, thereby reducing the number of synonyms and the cognitive load.

Analogy

When an agent creates a new form in a language game it usually randomly assembles phonemes (e.g., Steels, 1996). This mechanism is in line with the claim of the "arbitrariness of the sign" (de Saussure, 1916): the structure of the form has no relationship to the meaning conveyed by it. While this is true for many forms in today's existing languages, there is evidence that suggests that in the creation of new forms the intended meaning should be taken into account:

Compounds and inflections: When new words are created in, for example, English, they are often compounded and derived from existing words to ease their understanding. Thus, someone who eats bananas will be called a "banana-eater" rather than a "manslo" to indicate the semantic relationship with bananas and eaters. And someone who went for a walk last week is said to have "walked" and not "sali", in order to indicate the semantic relation to the root "walk" (there are only two idiosyncratic past tense forms in English, "went" and "was/were"). While these processes cannot be applied to simple language games directly, they do show a structural relationship between words that reflects a semantic relationship between their meanings.

Sound Symbolism: There is growing evidence for the controversial idea that the pronunciation of a word can suggest its meaning. This idea was first mentioned by Plato and has been pursued since then, notably by von Humboldt (1836) who gave examples of *waft*, *wisp*, *wind*, *wish*, and *wobble* where the "wavering, uneasy motion, presenting an obscure flurry to the senses, is expressed by the *w*" (p. 73). Since many vowels and consonants have undergone shifts through the times, this relationship is obscured in today's languages. However, subsequent psycholinguistic research has shown that indeed in the formation of words, certain sounds can represent certain meanings. For example, in assigning the two words *Mil* and *Mal* to images of big and small tables, 80% of subjects chose *Mal* to stand for the larger table and *Mil* for the smaller table, indicating that /a/ suggests big size and /i/ small size (Sapir, 1929). These results have been reproduced and extended by numerous researchers (see e.g., Hinton *et al.*, 1995).

A less controversial type than such "absolute" sound symbolism, is a "relative" sound symbolism, that could be directly applied to the creation of new forms in naming games. It is described in (von Humboldt, 1836, p. 74) as "designation by sound-similarity, according to the relationship of the concepts to be designated. Words whose meanings lie close to one another, are likewise accorded similar sounds; but [...] there is no regard here to the character inherent in these sounds themselves."

Taken together, these findings suggest that sound structure in word creation can be meaningful and could convey information about the word's meaning to the hearer. To integrate these findings into the language games played by agents, the way in which new forms are created could be modified by making use of the topology of the form and meaning space. The decoding of the form by the hearer could then work as follows:

```
Find a meaning for the form f:
for the nearest neighbor f′ of f
   according to the similarity
   metric, find the best meaning m′
associate f with that of the hypothesized
   feature sets which is closest to m′
```

This approach can help to reduce ambiguity in the hearer's lexicon. We implemented this idea in a variant of the naming game. The preliminary results suggest faster convergence of the language than in the original model, due to the emergence of regularities in the form–meaning mapping. Further, we found several examples of parameter settings that would not lead to convergence under the classical settings, but did converge under topological settings. Finally, we find an unexpected delay in the convergence in the final stage, due to "conflicts" between competing partial regularities.

Conclusions

We have discussed the relevance of language evolution models for Cognitive Science and presented a formalism for describing "language games". Language game models are complementary to work that studies language processing and language acquisition. At this point the models are simple; their value is that they make the roles of diachrony, situatedness and selforganization in the emerging linguistic structure explicit and testable. In the final part of the paper, we have raised issues where Cognitive Science can inform language game modeling, and eventually lead to a detailed understanding of how complex language has emerged from many simple interactions.

References

BATALI, J. (1998). Computational simulations of the emergence of grammar. In: *Approaches to the evolution of language: social and cognitive bases* (Hurford, J. & Studdert-Kennedy, M., eds.). Cambridge University Press.

BATALI, J. (2000). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In: *Linguistic evolution through language acquisition: formal and computational models* (Briscoe, T., ed.). Cambridge University Press.

CLARK, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

DE BOER, B. (1999). *Self-Organisation in Vowel Systems*. Ph.D. thesis, Vrije Universiteit Brussel AI-lab.

DE JONG, E. D. (2000). *Autonomous Formation of Concepts and Communication*. Ph.D. thesis, Vrije Universiteit Brussel AI-lab.

DEACON, T. (1997). *Symbolic species, the co-evolution of language and the human brain*. The Penguin Press.

GRICE, H. P. (1975). Logic and conversation. In: *Syntax and Semantics* (Cole, P. & Morgan, J. L., eds.), vol. 3: Speech Acts, pp. 41–58. New York: Academic Press.

HINTON, L., NICHOLS, J. & OHALA, J. J., eds. (1995). *Sound Symbolism*. Cambridge, UK: Cambridge University Press.

VON HUMBOLDT, W. (1836). *On Language*. Cambridge, UK: Cambridge University Press. Translated from the German by Peter Heath. This edition 1988.

HURFORD, J. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua* **77**, 187–222.

KAPLAN, F. (2000). *Objets et Agents pur Systèmes d'Information et de Simulation*. Ph.D. thesis, Université Paris 6, Sony CSL-Paris.

KIRBY, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In: *The Evolutionary Emergence of Language: Social function and the origins of linguistic form* (Knight, C., Hurford, J. & Studdert-Kennedy, M., eds.). Cambridge University Press.

LILJENCRANTS, J. & LINDBLOM, B. (1972). Numerical simulations of vowel systems: the role of perceptual contrast. *Language* **48**, 839–862.

MACWHINNEY, B., ed. (1999). *The emergence of language*. Lawrence Erlbaum Associates.

MITCHELL, T. (1997). *Machine learning*. McGraw-Hill.

NOWAK, M. A. & KRAKAUER, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA* **96**, 8028–8033.

OLIPHANT, M. & BATALI, J. (1996). Learning and the emergence of coordinated communication. *Center for research on language newsletter* **11**.

SAPIR, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology* **12**, 225–239.

DE SAUSSURE, F. (1916). *Course in General Linguistics*. La Salle, Illinois: Open Court. Translated by Roy Harris. This edition 1986.

STEELS, L. (1996). Self-organizing vocabularies. In: *Proceedings of Alife V* (Langton, C., ed.).

STEELS, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence* **103**, 133–156.

ZUIDEMA, W. H. & HOGEWEG, P. (2000). Selective advantages of syntax: a computational model study. In: *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum Associates.