Towards Formal Models of Embodiment and Self-organization of Language

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

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 (fixed) language. We focus on the framework of *language games* and argue that they offer a new 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

There exists a long tradition of formulating and studying formal models of language processing and language learning. These models have generally focused on the the linguistic competence of a single individual (e.g. Chomsky, 1980). They have proven to be appealing because the formalisms offer precision and clarity, they have led to successful technology, and they have allowed for extensive theoretical research to complement empirical work.

However, these competence-models have abstracted away many arguably crucial characteristics of language. These abstractions are viewed with growing uneasiness by cognitive scientists, linguists, and other researchers. Some of their concerns are well-known: competence-theories lack an appreciation of linguistic "performance" and of the communicative function of language, and they place a strong emphasis on symbolic processing and innateness (see e.g. Elman *et al.*, 1998).

Here we focus on a particular criticism: traditional models fail to acknowledge how much of linguistic structure emerges from communication and embodiment. Recent research on natural language pragmatics, for instance, 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*. Researchers in the framework of "emergentism" have argued that the structure of language should be explained as the emergent result of the many interactions between known processes in evolution, development, speaking, listening, and diachronic language change (MacWhinney, 1999).

This type of work thus emphasizes the role of (i) the function of language for communication between individuals ("cooperativity"), and (ii) the biophysical constraints of the human body and its environment ("embodiment") in the explanation for the origin and development of linguistic structure. We are sympathetic to these arguments and share the criticism of a tradition that in some sense equates the *formalisms* of the researcher with the *mechanisms* of the real brain. However, we regret that this general criticism goes hand in hand with a reluctance to use formal models at all. Many researchers have focused instead uniquely on empirical or philosophical approaches, or on building "embodied" robots.

The goal of this paper is to argue that formal models can deal in a meaningful way with embodiment, situatedness and self-organization. They can help to define these concepts and elucidate the role they play in the development of complex language. *Language games*, such as studied in recent years in the field of artificial life, are a prime candidate for this purpose. Language games are models of language change and language evolution in populations of communicating individuals. Although in most of these models "cooperativity" and "embodiment" have not played much of a role, we believe they can be successfully extended to incorporate these aspects.

In the following we will discuss the possibilities and the general format of these models and present a measure for the quality of a lexicon. We will then study a model that is simple, but is nevertheless novel and serves well to illustrate our approach. Finally, we will discuss how simple language games can be gradually extended to incorporate realistic aspects of cognition, embodiment, and communication.

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 that the reader will be convinced that language games offer an appealing framework to study other aspects of language as well. For language games that do incorporate grammar, we refer to (Batali, 1998; Steels, 1998; Batali, 2000; Kirby, 2000).

Language Games

The models of language evolution that we will consider are *multi-agent models*. They define a population of individuals that talk to each other and learn from each other, in a developing 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. The models evaluate the complex relationship between (i) acoustic, cognitive and articulatory constraints, (ii) learning and development, (iii) cultural transmission and 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.

Language game models can be viewed as an extension of the basic communication model that consists of a sender, a message and a receiver. Language games consider 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, f_i) and their meanings (m_j). 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. Consequently, lower scores are preferred over higher ones. E.g., if we have the associations $\langle f_1, m_1, 0.1\rangle$ and $\langle f_2, m_1, 0.6\rangle$, then the form f_1 will be uttered if meaning m_1 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. 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 contrast to such models, 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 initiator finds that the closest match to this (heard) form is the form that it originally used, the game is successful. Otherwise the game 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 maximization 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 simplification 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), but in the model that we present here, meanings are not available to the hearer. The model is in a sense a "standard" naming game, but with a continuous form space that was used only in the imitation game.

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):

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

β is a parameter that determines the speed of adaptation (here: $\beta = 0.1$). Associations with bad scores are seldom used, and associations that are not used often enough are removed. 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 will 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 forms 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:

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

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

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

From this equation 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 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.

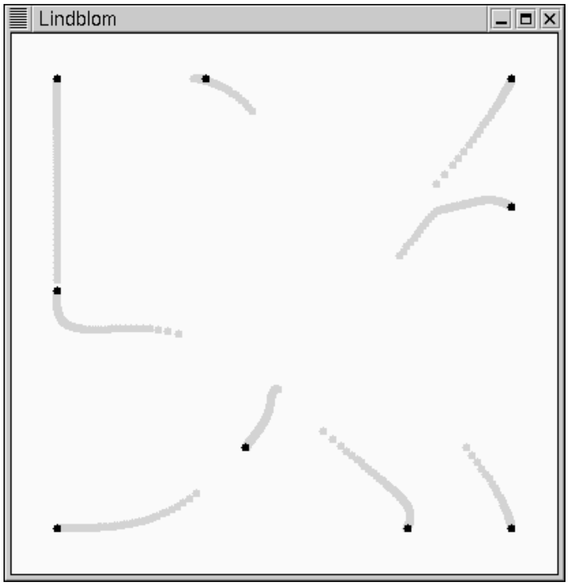
If we assume a simple extension of the model – a flux of agents – we can add a fourth criterion. New agents that come into the population should acquire the lexicon of the population as quickly as possible. In general, learning a mapping between two spaces is easiest if there is a regularity in the mapping,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

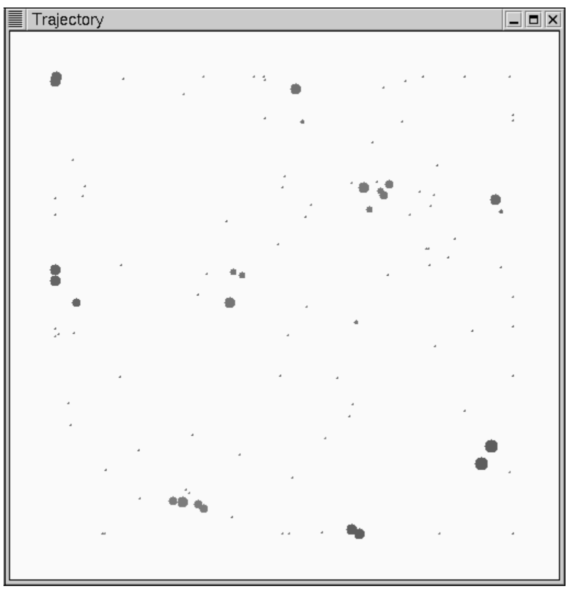
Equations 2 and 3 constitute a highly idealized quality measure for a communication system between individuals (described by the functions S and R), under some "embodied" constraints of articulation and perception (described by function U). The measure is a starting point and can easily be extended to describe more details of human communication. E.g., one could adapt equation 2 such that there are varying degrees of success in the communication about a topic, or that certain meanings are more frequent or more important than others.

Results and Discussion

The main result that we present here is that close approximations of each of the first three properties of the optimal lexicon emerge from the *local interactions* that we have defined above. 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



(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%)*

through local interactions between two communicating agents, expressing 9 different meanings with forms from a 2-dimensional form space. Each of the 9 clusters in this figure show strong associations from two agents for one particular meaning.

The emergence of these properties is not trivial and in fact depends crucially on the characteristics of the model. E.g., without the mutations (i) and (ii) that were described above, none of these results were obtained. Instead, the lexicon collapsed to a single large cluster of forms for only a single meaning.

The conditions for the emergence of an "optimal lexicon" need to be studied in more detail. However, our results already show that there is no necessity for explicit and innately specified "principles" that guarantee specificity, distinctiveness, sharedness and regularity. It is possible in principle that these basic characteristics emerge from simple interactions between agents, a generic learning algorithm and topological meaning and form spaces. That is, they emerge from the embodiment and situatedness of the simulated agents. Of course, the "biophysical constraints" of real humans are different from the ones we implemented in our model. The next step in our research is therefore to evaluate if more *realistic* constraints lead – through similar dynamics – to an emergent language with more *realistic* characteristics.

The fourth property of the optimal lexicon, "regularity", can equally be obtained in a distributed system. We will first discuss how the present model can be extended to be more cognitively plausible, and then mention briefly some preliminary results from a variant of the model described here.

Making the Model more Cognitively Plausible

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. First of all, 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. While such a process cannot be applied to simple language games directly, it does show a structural relationship between words that reflects a semantic relationship between their meanings.

Second, there is growing evidence for the controversial hypothesis that the pronunciation of a word can suggest its meaning ("sound symbolism"). This idea was first mentioned by Plato and has been pursued since then, notably by von Humboldt (1836). 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 version than such "absolute" sound symbolism, is a "relative" sound symbolism that can be directly applied to the creation of new forms in naming games. It is described in (von Humboldt, 1836, p. 74) as "Words whose meanings lie close to one another, are likewise accorded similar sounds", while the sounds themselves bear no direct semantic content. To integrate this type of sound symbolism into the language games played by agents, we have modified the way in which new forms are created by making use of the topology of the form and meaning space. The decoding of the form by the hearer works 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. 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 the study of embodiment and self-organization of language, 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, embodiment and self-organization 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.

CHOMSKY, N. (1980). On cognitive structures and their development: A reply to Piaget. In: *On Language and Learning: The Debate between Jean Piaget and Noam Chomsky* (Piattelli-Palmarini, M., ed.), pp. 35–52. Routledge & Kegan Paul.

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.

ELMAN, J., BATES, E. *et al.* (1998). *Rethinking innateness*. MIT 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*, pp. 577–582. Lawrence Erlbaum Associates.