The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems

Luc Steels

VUB AI Laboratory - Brussels

SONY Computer Science Laboratory - Paris.

July 25, 2005

Abstract

The paper discusses efforts to understand the self-organisation and evolution of language from a cognitive modeling point of view. It focuses in particular on efforts that use connectionist components to synthesise some of the major stages in the emergence of language and possible transitions between stages. The paper does not introduce new technical results but discusses a number of dimensions for mapping out the research landscape.

1 Introduction

Understanding the origins and evolution of language and meaning is currently one of the most stimulating areas of research in cognitive science. Because it is such a challenging question that touches on all aspects of cognition, it stimulates creative thinking and helps to consider many fundamental issues in a new light. Almost all subareas of cognitive science are contributing: from anthropology and archeology to genetics, linguistics and psychology. This paper focuses on contributions from the perspective of (connectionist) cognitive modeling. It surveys some of the major issues related to the question what kind of cognitive architectures individuals in a population need in order to see the emergence, sustainance and evolution of language.

Most of the insights discussed in this article come from a trend, which started in the early nineties [52], to use computer models and robotic experiments for testing theories (see [12], [8] for collections of representative work and [69] for a recent survey). Many of these models use at their core connectionist components, such as bi-directional associative memories, radial basis function networks, recurrent neural networks, etc. The novelty from a connectionist point of view lies not in the components themselves but rather in how they are integrated in a more encompassing architecture and how they are used to tackle aspects of social cognition, particularly how a group can arrive at a shared language inventory and repertoire of concepts.

The paper first introduces the view of language as a complex adaptive system, which has become the consensual view in the work surveyed here. Next a first axis for mapping out the research landscape is introduced: a distinction between the sociobiological and the sociocultural approach to the emergence of linguistic order and its increase in complexity. Then a second axis is proposed, which is based on viewing communication in terms of the transmission of information in the tradition of Shannon versus viewing communication as requiring inference for conceptualisation and interpretation. Finally, possible stages in the evolution from lexical to grammatical communication systems are proposed, which provides a third axis for understanding ongoing work.

2 Language as a Complex Adaptive System

Linguistic Capacities and Inventories

There is obviously a distinction between the following three aspects of language:

1. The *biological capacity for language* (also known as the architecture of the language faculty), which is the set of physiological and cognitive components that agents need in order to enter and participate in a language community. For example, speech requires a vocal tract with fine-grained motor control and auditory sensors and associated rapid signal processing, a lexicon requires some sort of bi-directional associative memory for storing and retrieving words, syntax requires the ability to recognise and produce sequential patterns, etc.

2. The *language inventory of an individual agent* or *ideolect*, which is the body of knowledge (sound system, individual lexicon, grammar) that an individual uses to

map form to meaning and meaning to form. It contains for example a specific inventory of sounds, a specific lexicon, a set of grammatical constructions, etc.

3. The *communal language*, which is the consensus that has arisen in a particular population on how to express meanings. This is obviously emergent from the activities of the individual agents and is not physically stored anywhere nor globally observable by the individuals in the population.

It is clear that language undergoes change at each of these levels. The biological capacity for language has somehow come into existence phylogenetically and is regenerated in every individual ontogenetically as part of a developmental process. The language inventory of an individual grows throughout childhood but then keeps changing as individuals adapt their inventories to the needs and communities in which they find themselves by a well documented alignment process [20]. The communal language is also constantly changing. New sounds pop up in languages, sometimes driven by mere fashion, or new aspects of sounds (like tones) may become exploited. New words enter all the time in the lexicon and the meanings of existing words change. New grammatical constructions enter and others disappear. These changes have been documented by historical linguists [22] and there is now clear evidence from sociolinguistic research that the change is much more rapid and ongoing than is usually assumed [27], particularly in multi-lingual settings. The idealised notion of a 'static uniform target language' has to be replaced by the notion that language has lots of internal variation and should be viewed as a complex adaptive system in constant flux [55]. As already argued by Hopper, grammar is forever changing and emergent [23].

We can furthermore make a distinction between the concrete elements routinely used in a language (the sound repertoire, the lexical items, the syntactic and semantic categories, the grammatical rules and constructions) and the meta-level structure of the language (further called the meta-system) which constrains how a language can be expanded if the need arises. This meta-system captures the sytematicity of the language. For example, a language might use a case system with nominative, accusative, dative, etc. and morphological case markings for the expression of event-argument structure of verbs (as in German or Latin), but it might also use grammatical relations (subject, direct object, indirect object) with word order and prepositions (as in English), or post-nominal particles (as in Japanese) for the same purpose [4]. Even within one of these approaches, there are still many further options, for example, a language might systematically adopt a nominative-accusative distinction (as in German) or an ergative-absolutive one (as in Basque). If a certain language has chosen one of these approaches and a new type of event needs to be expressed, the grammar should be expanded in the same 'style' as is already dominant in the language in order to make it easier to memorise and infer by analogy the intended meaning.

Just as we made a distinction between ideolect and communal language, a distinction can be made between:

1. The *individual meta-system*, which is the meta-system that a single individual maintains for his own language.

2. The *communal meta-system*, which is the meta-system underlying a communal language.

The individual and communal meta-systems also appear to change continuously, just as the language system itself, and these changes may be profound as they may lead to an overhaul of global systemic principles. For example, new word classes may appear (like articles which did not exist in Latin but then emerged in all derived Romance languages), the expression of certain syntactic features may shift (like the expression of past tense in English which shifted from morphological modulation (come/came, do/did, is/was) to the use of the -ed morpheme (as in dance/danced)), or grammatical systems may change (as happened in English which shifted from a case-marking system as in Latin to a word-order+preposition based system for the expression of event-argument structure [66]). So here again, we have to abandon the notion that there is a static meta-system in favor of a complex adaptive systems view.

In contrast, some linguists believe that there is a single communal meta-system for all human language communities consisting of a set of basic principles and parameters. This universal meta-grammar is usually called Universal Grammar (UG) [14] and is not assumed to undergo change. It then becomes conceivable to equate the universal meta-system with a biologically determined language faculty. The meta-grammar of a particular language (and the one known by an individual speaker of that language) then consists of a particular setting of the parameters of UG, and these parameter settings do undergo change [30].

In addition to linguistic capacities there are also the conceptual capabilities to produce or interpret the semantic structures expressed by language. These conceptual capabilities clearly also undergo change as well, and not just during childhood. As new tasks and domains are tackled, new categories or conceptualisations become relevant and others become obsolete. This expansion pushes the language to lexicalise or grammaticalise new conceptualisations while the conventionalised expression of a conceptualisation in turn helps to spread and maintain it in the population and helps individuals to align their conceptual repertoires so that they become more similar [20], [60]. Although the formation and coordination of new conceptualisations is crucial for understanding how language itself is possible, we will not develop it in the present paper.

3   Sociobiological versus Sociocultural Explanations

Broadly speaking, three approaches are being explored for synthesising the origins and evolution of linguistic structure: a sociobiological, a sociocultural, and a mixed biological/cultural approach.

Sociobiological explanations

The first approach has been pioneered by Hurford [5], who introduced the term sociobiological. In line with other nativist trends in linguistics [14], it relies on genetic coding and natural selection, and is hence closely related to evolutionary psychology [45]. Paraphrasing Hurford (o.c., p.194), the explanation structure of the sociobiological approach can be summarised as follows:

1. Individuals who are more successful communicators enjoy a selective advantage and are more likely to reproduce than individuals who are worse communicators.

2. If an innate strategy X for communication is superior to other conceivable strategies, its possessors tend to enjoy a reproductive advantage over others, thereby increasing the prevalence of this strategy in the next generation.

3. Therefore over an evolutionary timespan, strategy X displaces all rivals, and ends up being *the* strategy used for establishing and maintaining communication systems in the group.

The term strategy is construed broadly. It can be a strategy for acquiring the lexicon of a language, a strategy for using a particular word for expressing a certain meaning, a strategy for exploiting a particular cognitive mechanism for language, a strategy for using a certain grammatical construction, a meta-strategy for expanding the grammar of a language in a particular way, etc.

Hurford first applied this line of thinking to investigate the optimal strategy for the most basic stage of language in which agents name individual objects or situations. He shows (further confirmed by [47]) that a Saussurean strategy in which the associations between names and objects is bi-directional is optimal. Hence this strategy could in principle have become genetically innate by natural selection, assuming a strong feedback loop between communicative success and reproductive advantage.

The same line of argument can and has been applied for explaining other aspects of language. For example, to explain compositionality, there could have been a competition between individuals who use a compositional coding strategy and individuals who use a holistic strategy (single word sentences coding complex meanings). Assuming that a compositional coding strategy results in better communication systems and hence a reproductive advantage for those using it, the genes coding for such a strategy would progressively spread in the population so that it becomes part of the innate Language Acquisition Device, and this could explain why human languages are compositional [41].

The sociobiological approach can also be used to explain the origins and structure of communal languages, in which case the competing strategies are words or grammatical constructions for expressing certain meanings. One of the earliest demonstrations of this approach [72] uses a lexicon in the form of a bi-directional associative memory which is genetically coded in the sense that the weights of associations between words and their meanings is genetically inherited and changes to these weights arise by mutation and recombination. A particular linguistic choice, i.e. to use a certain word for a certain meaning, gives individuals in a particular population a reproductive advantage so that the gene coding this choice spreads in the population. Eventually all agents share the same winning genes and complete communicative success is reached. Several researchers have further explored this line (see e.g. [29]), including for the genetic evolution of simple grammars (see e.g. [12]).

Sociocultural explanations

The second approach is sociocultural. Its explanation structure is as follows:

1. Given a population where individuals have several possible strategies for negotiating and using communication systems.

2. Some strategies might be more effective than others and this impacts the expressive power of the language, the success that agents have in communication, and/or the effort that they require.

3. Agents try to maximise expressive power, optimize communicative success and minimise effort by choosing the most adapted strategy, therefore, over a cultural time-span, this strategy displaces all rivals, and ends up being *the* strategy by which communication systems are, at least temporarily, established and maintained in that specific population.

So there are two key differences between sociobiological and sociocultural models: (1) In a sociocultural model, selection does not go through fitness, reproductive success, and genetic coding, but through cultural choice and direct feedback on success or effort in communication, and (2) the creation of variation and new structure is not based on genetic mutation or recombination but on the invention and use of linguistic structure and its partial (possibly slightly differing) adoption by others.

This sociocultural explanation structure can be applied in a straightforward manner to tackle the origin of specific ideolects and communal languages. Choice for one strategy versus another (e.g. which word will be used to express a certain meaning or which grammatical construction will be used) is now not only based on effort or expressive power, but also on which choice is most frequent in the population as this determines the potential for success. Already from the mid-nineties (see e.g. [24], [50]) simulations started to appear which showed that the principle of self-organisation is effective to get a population to agree on a shared lexicon. Again, a bi-directional associative memory was used for the lexicon, but now the weights change based on success and failure in the communication and they are not genetically inherited. When agents choose the word-meaning pairs that had most success in the past, successful words will be used more and thus the positive feedback loop required for self-organisation is achieved. Lateral inhibition after a successful game and damping after an unsuccessful game have proven to be the most optimal strategy [43]. Very soon examples were shown where agents negotiated grammars using very similar mechanisms [1], [54]. The sociocultural approach could also be demonstrated for the evolution of perceptually grounded categories in co-evolution with a lexicon [60] and for the emergence of sound repertoires [**?**].

All these are examples where a communal language self-organises and continues to adapt under the forces of maximising communicative success and expressive power while minimising effort. But the same explanation structure has also been used with respect to the choice for a particular strategy. Thus whereas Nowak and collaborators [41] showed how the choice of holistic versus compositional coding could be the outcome of a socio-biological process in which the more optimal strategy becomes innate, the simulations reported by Kirby and collaborators [6], show that this choice need not be based on fitness of the individuals (as would be the case in a sociobiological model) but on the impact of the transmission bottleneck which brings a population towards choosing and sustaining a compositional strategy. In other words, the cognitive architecture of the agents supports a choice between compositional and non-compositional coding and agents choose for the compositional strategy. Another example of research along these lines is reported in [68]. It compares different strategies for grounded lexicon acquisition and shows that joint attention and corrective feedback are a better strategy than non-social forms of learning.

It is perhaps less obvious how to apply a sociocultural approach to explain the origins of the language faculty. If there is a single universal meta-system for language, it is conceivable that this system is biologically embodied as a highly specialised genetically determined 'language organ', as indeed some linguists have argued [14], [45] (figure 1 left). In that case the universal meta-grammar and the biologically acquired language faculty are two sides of the same coin and a sociobiological approach is the most logical explanation.

On the other hand, if the meta-systems of human languages undergo change, including true innovation, then a sociobiological explanation becomes less likely (or at least no longer necessary and sufficient), and a population can be seen as selecting from the set of all possible physiological and cognitive mechanisms normally available to human beings those that are adequate for supporting a specific way of creating a language. As many alternatives are possible, there is no longer a universally shared, genetically determined language organ but each language exploits different constellations of biological capacities (figure 1 right). A trivial example is the difference between a speech-based and a gesture-based language (i.e. a sign language). The first type has recruited vocal articulatory and auditory embodiments and processing whereas the second type has recruited bodily gestures and visual recognition. Another example is the difference between languages which almost exclusively rely on morphosyntax (as Australian Aboriginal languages [16]) and those almost exclusively relying on sequential ordering (as English). Quite different processing and learning mechanisms are involved in one or the other. There may even be differences among the individuals within the same language community, which would make the use of that language more robust. Like an amoeba, the required biological basis for a specific language adapts to exploit the available neural structures, even shifting from one hemisphere to another one if damage comes early enough [63].
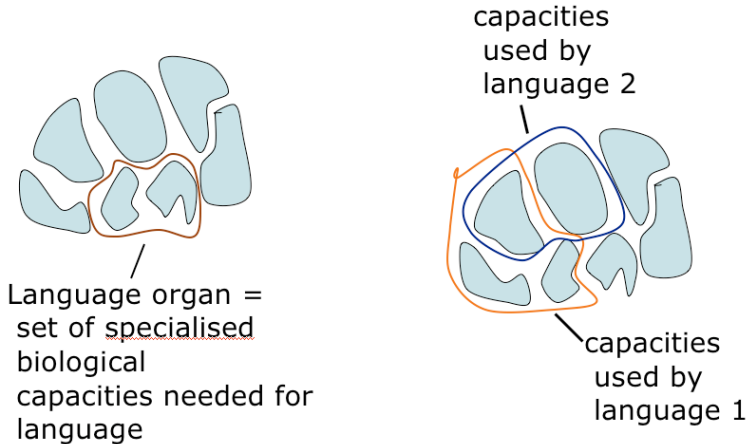

Figure 1: Two views on the relation between biological capacities and language. In the 'language instinct' view (left) highly specialised and genetically determined components make up the language faculty. In a cultural view (right), a particular language uses a particular constellation of generically available biological capacities.

Combining biological and cultural approaches

There are obvious parallels and interactions between sociobiological and sociocultural approaches. In fact, there are few researchers who believe that all of language has resulted from sociobiological evolution or alternatively that all of language has a sociocultural foundation. The selection criteria that are used in sociobiological explanations are often relevant to sociocultural explanations, and vice-versa. It is only the way these criteria are assumed to impact the evolving communication system that is different. Thus, if a particular style of grammar (say one exploiting word order instead of morphosyntax) is more effective for a population, sociobiologists would argue that this style becomes innate due to the reproductive success agents enjoy, whereas socioculturalists would argue that it is a (temporal) cultural consensus due to self-organisation.

Some experiments have studied the interaction between the two, i.e. how certain strategies, which have evolved in a sociocultural fashion, can become genetically engrained due to the Baldwin effect [37], [9]. This requires however that there is first a sociocultural evolution leading to sufficiently stable choices so that genetic evolution can catch up with them.

4 Signalling Systems versus Language Games

A second axis to structure research activities is orthogonal to the sociobiological/ sociocultural distinction and related to how communication itself is viewed, namely as pure information transmission or as inferential [49].

Following Tomasello [65], we define communication as an action whereby the speaker attempts to draw the listener's attention to some aspect of the world: an object, a situation, an action, a property or relation, etc. This is typically done to achieve a cooperative action in the world, but usually the desired action must be inferred. For example, the speaker could draw the attention of the hearer to an object with the implicit assumption that the hearer then hands the object to the speaker, or the speaker could draw attention to a potential action which he wants the hearer to perform. Speaker and hearer are considered to be able, directly or indirectly, to determine whether the communication was successful, for example because the hearer performs a consequential action. All this is already true for alarm calls: The animal emitting the call does not communicate whether and how to flee from a predator but draws attention to a predator of a particular kind, and he or she can afterwards easily observe whether his or her call had the desired effect. In most cases, communication is not a one-step interaction but participants repair each other's understanding by additional communications and actions. Note that a distinction must be made between pragmatic feedback, coming from subsequent action, and linguistic feedback, where agents directly correct each others' communications.

The first way in which evolution of communication systems can now be studied is within the tradition of Shannon's information theory. The information to be expressed (the message) is given in one way or another (for example as output of a perceptual system), and the task is to code and decode that information as efficiently and reliably as possible in terms of signals. A lot of simulation studies within this framework hence start from a set of possible messages and let agents evolve message-signal matrices (see e.g. [29]), possibly using syntactic structure to make the transmision more reliable or more compact or easier to learn (see e.g. [41]). The criterion for success in communication is

whether the information decoded from the signal by the receiver is identical to the information coded by the sender. Many simulations take a short-cut and compare directly the matrices developed by agents to compute the probability whether their communications are going to be successful [40]. This probability is then directly used to estimate the fitness of the agent in sociobiological models, thus avoiding the need for multi-agent simulations. The signalling viewpoint is most often adopted by those coming from the perspective of animal communication.

The second viewpoint is embedded in the tradition of philosophical investigations of communication [49] as well as cognitive linguistics [28] and emphasises the inferential nature of human communication. It is most often adopted by those with an interest in understanding the origins and evolution of human natural languages. The inferential approach starts from situated embodied interactions, often called language games. The interaction takes place in a shared setting in which the speaker wants to draw the attention of the hearer to some aspect of the world and thus invoke a possible action. The interaction requires two types of activities: (1) The speaker needs to conceptualise the world in a way that a distinctive description is found of the topic he wants to draw attention to. This is usually called the discrimination task. The hearer needs to use this distinctive description to determine what aspect of the world the hearer wants to draw attention to. This is the interpretation task. (2) The speaker needs to evoke the conceptualisation in the hearer by choosing a number of conventionalised symbols (this task is usually called production) and the hearer must use these symbols to reconstruct the intended conceptualisation (usually called parsing). A communication now succeeds when the hearer pays attention to the aspect of the world that the hearer originally has chosen, which is observable by the speaker through the hearer's subsequent action or to the hearer by corrections or repairs initiated by the speaker after his action.

A complete language game then combines these various steps is shown in figure 2. As a side effect of playing a game, speaker and hearer may expand their language or conceptual inventories, change weights to reflect usage frequencies, switch strategy, or recruit new types of cognitive processes, all aimed to achieve communicative success, maximise expressive power and minimise the effort involved in communication.
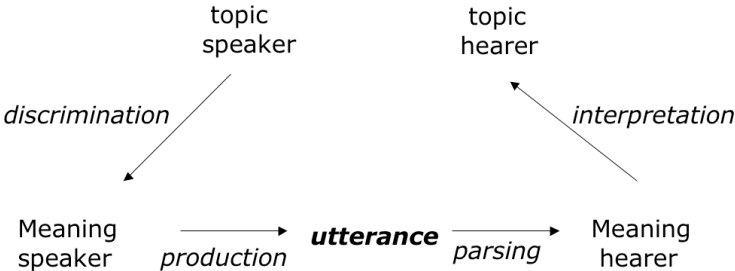

Figure 2: A language game requires discrimination and interpretation as well as production and parsing.

There are some important differences between inferential communication and information transmission. The first one is that speaker and hearer do not necessarily have to have the same meaning in order for the communication to succeed. Suppose there is a

small red ball and a big blue pyramid and the speaker says "bobado" to refer to the first object, with "bobado" meaning 'small'. Suppose the hearer parses "bobado" as meaning 'red'. Despite differences in meaning, the communication is still successful. Speaker and hearer will not even be aware that they used different meanings. If they live in a world where balls are always red and small, and where there are no other objects that are red or small except balls, then they will always succeed.

Second, the speaker's conceptualisation of the situation should be seen as a hint rather than a completely reliable specification, which is indeed what makes the communication inferential. Context and world knowledge has to be used to reconstruct the potential meanings or interpret them. This is the reason why natural language can tolerate ambiguity and why communication can be robust despite a very unreliable medium and divergent perception for the participants in the communication.

A third difference is that the inventories of agents do not need to be identical to have successful communication. It is enough that they are sufficiently coordinated to have successful communication within the environments that the agents find themselves. Indeed, given that lexicons, grammars and conceptual inventories are a function of the individual history of situated interactions with others, it would be very surprising if they are absolutely identical. Careful observations of human categories or naming behavior shows tremendous individual variation [70] and rapid alignment [20].

All this implies that it is no longer possible to simply compare coded and decoded meaning to determine communicative success, or compare the inventories of agents to determine the probability of their success in communication, as is done in signalling experiments, instead multi-agent simulations now become necessary, particularly if the environment is open-ended. Pragmatic feedback must be integrated in the interaction to determine communicative success.

5   Stages and Prerequisites for Communication Systems

We have already discussed two axes for structuring research in the origins and evolution of languagelike communication systems: the sociocultural/sociobiological axis which concerns the nature of the mechanisms responsable for the origin and propagation of linguistic structure, and the transmission/inferential communication axis which concerns the nature of the communication. We now turn to another axis for mapping out the research landscape, which is related to the complexity of both the meanings and the forms (or signals) used by the agents.

From a research point of view, it is obviously extremely useful to identify a set of evolutionary stages or milestones with respect to human language, because then simpler forms of language could be studied before tackling more complex ones and we could investigate transitions between stages. This approach is similar to current research in biology on the 'minimal cell' [31], and on transitions that would transform such a cell or collections of cells into more complex life forms [34]. From an engineering point of view, a division into stages is useful as well, because we can then already build applications with simpler systems without having to tackle the full complexity of human language.

It is not so obvious to identify different stages in the complexity of human-like communication, and there is definitely no consensus about it. All human languages are known to have a similar level of complexity and expressive power, so different stages are not so easily observable as in biological species. Bickerton [3] has proposed that there are two stages: a lexical stage, which he calls proto-language, similar to the language of 2-year olds or pidgins spoken in trading contexts, and a grammatical stage, which has the complexity of full human language. Jackendoff [25] paints a more complex picture with a dozen different stages. He includes for example a step in which a system of grammatical relationships conveys a set of semantic relations, a stage in which there is the emergence of a system of inflections, and so on. Jackendoff's stages are motivated by fossils that he identifies in existing human languages.

We consider in this paper an alternative set of seven stages, specifically motivated by research into the cognitive architecture and self-organisation of communication systems in embodied agents. Each stage is characterised by a particular level of complexity at the language side and a particular level of expressive power at the meaning side. Each stage is assumed to build further on the cognitive capacities achieved at earlier stages and requires a major 'breakthrough' or transition. The stages proposed here do not necessarily correspond to historical stages in human language evolution or child language acquisition but are considered to be useful plateaus for structuring the investigation and the engineering of embodied communicating agents.

To achieve even a minimal form of communication (particularly from the viewpoint of inferential communication), a number of prerequisites need to be satisfied:

1. The agents that will be involved in the communication need to be engaged in a cooperative task which usually implies that they are situated in the same environment to which they have access through their sensori-motor system. The environment and agent activities generate shared cooperative goals.

2. The agents need to have a medium in which signs can be constructed, i.e. a repertoire of sounds or gestures that they can mutually recognise and reproduce.

3. Agents need to be able to engage in turn-taking behavior. One needs to recognise that the other one wants to engage in communication and then they could both start a routinised series of interactions.

4. Agents need to be able to share attention, even without communication. Shared attention can be achieved in many ways: by pointing, by eye gaze following, by bodily movements, by recognising actions of the speaker and predicting what he will do next, etc. Shared attention is necessary to drastically restrict the set of possible aspects of the world the speaker may draw attention to and to guess the set of possible meanings used by the speaker in case the listener does not know the conventions that were used.

Clearly each of these prerequisites is highly challenging and profound work has been going on in many corners of cognitive science to understand them. But this will however not be discussed further in this paper.

6 Stages in Lexical Communication Systems

The first three stages concern lexical communication systems because they use individual meaning-carrying units without syntax:

Stage	Meaning	Form	Breakthrough
Stage I	Individuals	(proper) Names	Convergence on convention
Stage II	Single Categories	Single Words	Co-evolution language/meaning
Stage III	Multiple Categories	Multiple Words	Compositionality

I The Naming Game

In stage I, the meaning M consists of an individual object and the utterance S to a single word or sign naming this object. This corresponds to proper names in natural language, such as "John" refering to the unique object named John. Language games at this stage are also known as Naming Games. The major breakthrough required is that the population is able to negotiate a shared set of linguistic conventions.

There are three types of functions that agents need to build up and maintain a lexicon of form-meaning pairs in order to reach shared conventions:

Invention When a new name is needed, the speaker should be able to generate a new word from scratch and associate that in his memory with the object he wants to draw attention to.

Adoption When the hearer encounters a new name and is able to guess the referent after additional interaction or joint attention, he should be able to associate this new word in his memory with the object.

Alignment Agents should update the strength of their association based on success in the game. When a particular association was successful, the strength is increased and that of competitors decreased (lateral inhibition). When an association was not successful its strength is decreased (damping). This implements a reinforcement learning approach which has been widely studied in the connectionist literature.

There is already a well-established literature on the Naming Game (starting from [50], [43]) and many computer simulations have shown beyond doubt that the lateral inhibition dynamics is effective (see for example figure 3 from [61]). Although the cognitive architecture of agents at this stage appears relatively straightforward, it is far from trivial to show in a theoretical way that a shared vocabulary emerges in the population given a particular set of microscopic behaviors. Progress is being made using techniques from complex systems science but theoretical proofs are still forthcoming. The question is similar to multi-agent decision problems in economics (for example the opinion dynamics discussed in [71]) or more generally, the emergence of global coordinated state based on local interactions as studied in statistical physics.

Many of the computational components required for stage I can be easily implemented using neural networks. For example, agents could use a classical feed forward neural network with a winner-take-all strategy for object recognition (for the discrimination
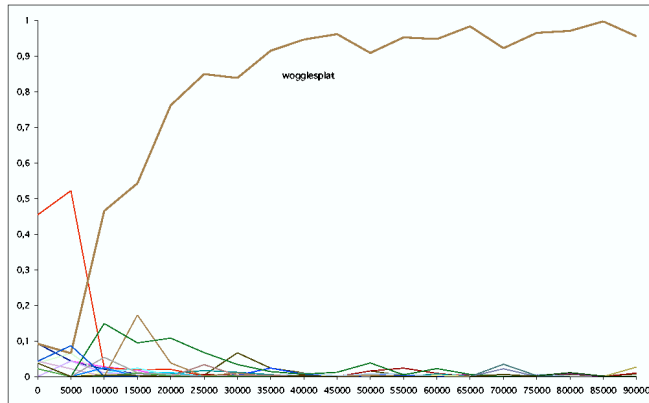
Figure 3: The graph plots the frequence with which a certain word is used for expressing a particular meaning. There is a winner-take-all effect due to self-organisation.

game). The associative memory for the lexicon is also relatively easy to implement in neural networks, for example using a bidirectional associative memory ([26]) which is itself an extension of the Hopfield network. The strength of associations is then implemented as weights of relations in the network and a winner-take-all process decides which name to use for a given object and which object to use for a given name. Instead of using only Hebbian learning, lateral inhibition has to be added as pioneered in Kohonen networks.

II. Coordinating Categories through Names

In Stage II, the meanings M consist of a single category that identifies an individual object in a particular context. The utterance S names such a category. This corresponds to the use of words like "table" or "red" to refer to a table or a red object. To reach this stage, agents must be able to categorise objects in order to find a discriminative category, to identify an object based on being given a discriminating category, and to build up a repertoire of such categories, adequate for playing discrimination and interpretation games. To do this, three functions are required:

Invention When a new category is needed by the speaker (because no distinctive categories exist to discriminate the referent from other objects in the context), he should be able to generate a new category and add it to the repertoire.

Adoption When a new category is needed by the hearer (because the language game failed and the hearer should be able to generate a new category using the same or similar mechanisms as the speaker.

Alignment Agents keep track of the success of their categories in the language game and prefer those that have been shown to be successful in the past.

To express these categories, agents must maintain a two-way associative memory similar to Stage I. Instead of associating names to uniquely identified objects, agents now associate words (or signs) to categories.

Computational mechanisms capable to achieve these functions have been demonstrated and studied since the mid-nineties. Many possible ways exist to implement category formation for discrimination and interpretation, ranging from neural network (e.g. using radial basis function networks [60]) to symbolic approaches (e.g. discrimination trees [51] or prototypes [62]). It has also been demonstrated that output of discrimination can be fed into coding processes of the same sort as used in the Naming Game (Stage I) and that the same sort of decoding processes as in Stage I can be used. Moreover by the proper coupling of the two games, it has been shown that agents are able to coordinate their repertoires of perceptually grounded categories, even if these categories were not given innately nor centrally coordinated. The coordination takes place entirely through language (see figure 4 taken from [60]). So implementing embodied communicating agents for Stage II is at the moment well mastered and can already be the basis of non-trivial applications.
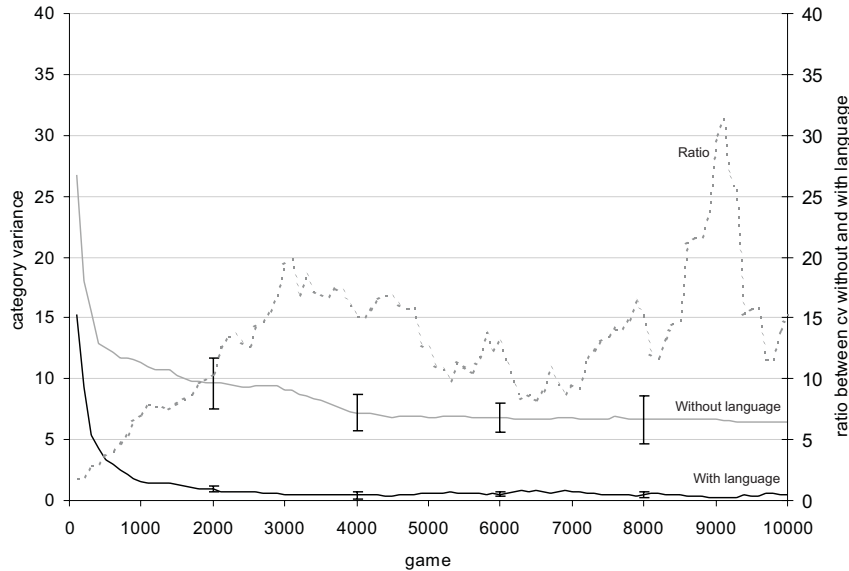

Figure 4: The graph plots the cumulative category variance between the ontologies in a population of agents with and without language. Agents play discrimination and interpretation games and construct an ontology as a side effect. When there is no coordination through language, the ontologies do not converge, otherwise they do.

III. Compositionality

A communication system is compositional if an utterance can contain several parts (signals or words) and the meaning of the total is a combination of the meanings of the parts. The simplest form of compositionality is one where a sentence has multiple words

and each word expresses a category or combination of categories. So this stage is quite naturally triggered by a need to express multiple categories which arises when the domain of discourse has sufficient structure so that categorisation in multiple conceptual spaces is more efficient than categorisation in a single space. Although the need to express multiple categories (such as 'small', 'red' and 'ball') introduces the potential for using multiple words, and thus a compositional coding, that is not necessarily so. It is also possible to use the same strategy as in stage II and use holistic coding. For example, a single word to express the conjunction 'small', 'red', 'ball' could be chosen as opposed to three different words.

To reach this stage, agents require more complex computational mechanisms that are able to cover a set of categories using one or more words, or to reconstruct a combination of categories using multiple words. The semiotic dynamics becomes more complex compared to stage II, as different solutions, some more holistic and others more compositional, compete for dominance in the population [67]. It has been shown that a strategy of using multiple words is beneficial (in the sense that it leads to higher communicative success and less effort) when the domain of discourse has enough regular structure [36]. As mentioned earlier, some researchers have focused on the issue how compositional systems might come to dominate in order to overcome the learning bottleneck [6].

7 Stages in Grammatical Communication

The three stages discussed so far all concern lexical languages, without any form of syntax. The mechanisms needed for conceptualisation/interpretation and for parsing/production at each of these stages are now reasonably well understood. There is not a single solution but multiple strategies are possible, particularly for concept acquisition. Research beyond these stages is less advanced, although progress clearly has been made and important breakthroughs can be expected the coming years.

The first question is why a communication system would emerge that exploits syntax. Opinion differs whether one takes a Shannon-style information theoretic view or whether one adopts an inferential view of communication. In the first view, syntax arises to optimise transmission. Indeed, if sender and receiver agree on the use of syntactic structure then they have to transmit less information in the message themselves. The transmission can become more compact and more reliable. This argument has been made by several researchers [41], [19] using mathematical proofs and computer simulations. Another argument, but still from an information-theoretic view, focuses on the learning bottleneck, caused by a lack of learning data for a new generation of learners and generally accepted in linguistics as a 'poverty of the stimulus'. If there is such a learning bottleneck, then a communication system with syntax is more beneficial because it can be learned more reliably with fewer data [6], [39].

The inferential view on communication and the language game framework that explores it, is proposing a quite different hypothesis concerning the nature and origin of grammar, namely that it arises from the need to make the conceptualisation process richer and the interpretation process more reliable [59]. This line adopts the 'cognitive linguistics' view that syntax is not arbitrary or a mere structuring device but rather that it carries meaning, i.e. that syntax becomes grammar in the sense of a carrier of meaning

(see figure 5). Consider for example the multi-word sentence, "red ball under small box". Let us assume that there is no syntax yet, so the sentence is equivalent to "under ball box red small" or "small box ball red under". Based on the meaning of the words only, the sentence conveys that there is a ball, that there is something red, that there is a box, that there is a spatial relation between two objects, etc. This translates to the following predicate-calculus expression (with question marks indicating the variables):

red(?x), ball(?y), under(?a,?b), small(?c), box(?d)

Without syntax, the speaker does not communicate that "red" and "ball" are about the same object (i.e. that ?x and ?y are to be bound to the same object), that "small" and "box" are both about another one (i.e. that ?c = ?d), and that the ball is under the box and not the box under the ball (i.e. that ?x = ?a and ?b = ?c). By conveying this additional information through syntax, the speaker would not only avoid misinterpretations but also reduce the computational complexity of the semantic interpretation process, which is exponential w.r.t. the number of variables. One way to link different lexical items and their meanings is by combining words into patterns. Thus the co-occurrence of "red ball" can be taken to mean that red and ball are about the same object and should therefore be decoded as 'red(?x), ball(?x)'. There are other syntactic means that could be used, for example both words could be tagged with an additional affix specifying that they are about the same thing, as in "red-ba ball-ba", giving rise to a kind of morpho-syntax.

Research to explore the origins of grammar from the viewpoint of optimising inferential communciation is stil in its infancy. It appears useful to distinguish four different stages:

Stage	Meaning	Breakthrough
Stage IV	Multiple Objects + Predicates	Constructions
Stage V	id.	Meta-grammar
Stage VI	Second Order predicates	Second order expression
Stage VII	Meta-level	Level formation

Stage IV involves the use of syntactic patterns for linking meanings of individual words. In order to achieve it, agents need considerably more complex cognitive functions compared to lexical stages: They need to be able to detect the need for the introduction of syntax or be able to interpret a sentence even if no syntactic information is available. They need to be able to recognise and reproduce the syntactic or morpho-syntactic patterns that are used in a particular language, and invent syntactic conventions or adopt them from others. The same alignment dynamics as in earlier stages is required to align the different syntactic conventions that agents invent, so that the population converges on a shared syntax.

The first significant experiments at this stage were carried out by Batali [1] who constructed a simulation where agents are capable to construct sequential patterns that link a set of predicate-argument structures and are able to combine patterns by an appropriate mapping of the variables. Batali used an exemplar-based learning technique for retrieving form-meaning associations that most closely match the meaning to be expressed (in production) or the form (in parsing). The same sort of reinforcement learning is used as earlier discussed for Stage I. Agents align patterns because they tend to use the ones which occur most frequently in the population. Batali used symbolic representations and

learning and implementing them in a realistic neural embodiment is much more difficult. There are various types of networks (such as recurrent neural networks [18]) which are able to learn syntactic patterns and they are therefore a potential basis for implementing these experiments as well.

Stage V moves from the ad hoc use of syntactic patterns to more systematicity, through an intermediary layer of syntactic and semantic categories intervening between form and meaning, so that the linking takes place at a more abstract level and the communication system becomes much more regular in structure. This is indeed what human natural languages do. Typical examples of syntactic categories are noun, verb, nominative, masculine, past, perfect, etc. They are syntactically marked with word order, affixes, intonation, stress, etc. and are used to define or recognise abstract syntactic patterns like SVO. Typical examples of semantic categories are agent, beneficiary, source, cause-transfer, state, etc. They are used to define or recognise semantic frames which reconceptualise meanings in a more abstract way. Grammatical constructions then link semantic frames to syntactic patterns. For example, the Cause-transfer links a semantic frame involving an agent, patient and target to a syntactic SVOtoO pattern (subject-verb-directobject-to-object,) [21] (see figure 5). A specific verb, like slides, is an instantiation of this construction.
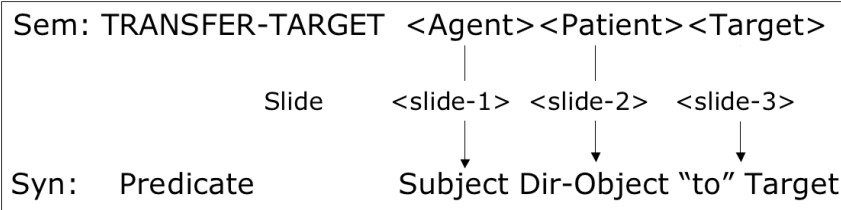
```
Sem: TRANSFER-TARGET  <Agent><Patient><Target>
                            |         |
              Slide     <slide-1>  <slide-2>  <slide-3>
                            ↓         ↓         ↓
Syn:    Predicate         Subject Dir-Object "to" Target
```

Figure 5: A construction relates a syntactic pattern such as Subject+Predicate+DirectObject+PrepObject with a semantic frame such as TRANSFER-TO-TARGET+Agent+Patient+Target.

The introduction of an additional layer requires a more sophisticated computational framework to implement grammatical constructions and the syntactic and semantic categorisations that they require. The meta-grammar has to become explicit as well so that the adoption and expansion of syntactic form to reflect new syntactic categories is systematised. At the moment there is growing research into the computational implementation of construction grammars [2], [11], [58], some of it with neural inspiration [17], and on the acquisition and invention of constructions [13] strongly inspired by research on child language acquisition [10] and research in grammaticalisation phenomena [22]. All this work currently forms the basis of the first construction grammar based experiments in grammar emergence (such as [58]) but much remains to be done.

There are still two further stages that can be identified compared to human natural languages: Stage VI is triggered when the need arises to include predicates that modify other predicates, i.e. second order (or even higher order) predicates. For example the adverb "very" in "very big ball" is modifying the category expressed by the adjective "big". It is not a predicate that ranges over objects in the domain of discourse directly. The known arsenal of logical operators, quantifiers and connectives now becomes part of the expressive power of possible semantic structures. When this higher order usage of predicates needs to be communicated explicitly, it is an additional reason why grammar

becomes crucial and necessary. Although there is significant amount of research in (logical) semantics and linguistics beyond first order logic (see e.g. research into Montague Grammar [44]) there are so far no computationally tested theories how such second order predicates could arise nor how the grammar could arise to express them.

Finally Stage VII arises when a language becomes its own meta-language, i.e. when it becomes possible to state information about meaning or structure in terms of the language itself, and thus allow for a much more rapid cultural spreading. To implement the necessary computational functions for the latter stages in realistic neural networks appears at the moment quite beyond the state of the art.

8 Conclusions

This paper discussed a number of issues that have come up in attempts to synthesise how a population of agents might arrive at a shared communication system with properties that are similar to human languages. We first identified the main linguistic and conceptual aspects that need to be tackled and the distinction between sociobiological and sociocultural approaches for the synthesis of structure and the transition between stages. An orthogonal distinction was also made between signalling and inferential communication which leads to different types of experiments and sources of explanation. Next we looked at a number of possible stages. Each stage is characterised by a particular level of complexity at the meaning side and a particular level of complexity at the form side. Each stage also requires an increased set of more complex strategies for playing discrimination and interpretation games, including the formation of the ontologies required to play those games, as well as more powerful strategies for communicating these meanings in coding and decoding. Although there are quite solid results for the first stages, much remains to be done for reaching the higher stages, particularly when a system of syntactic and semantic categories underlies the mapping from meaning to form as is clearly the case in natural languages.

9 Acknowledgement

References

[1] Batali, J. (2002) The Negotiation and Acquisition of Recursive Grammars as a Result of Competition Among Exemplars. In: [8] p. 111-172.

[2] Bergen, B.K. and N.C. Chang. 2003. *Embodied Construction Grammar in Simulation-Based Language Understanding.* TR 02-004, ICSI, Berkeley.

[3] Bickerton, D. (1990) Language and Species. University of Chicago Press, Chicago.

[4] Blake, B.J. (1994) Case. Cambridge University Press, Cambridge.

[5] Hurford, J. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua, 77(2):187–222.

[6] Brighton, H. and S. Kirby (2001) The survival of the smallest: stability conditions for the cultural evolution of compositional language. In J. Kelemen and P. Sosik (Eds.) Advances in Artificial Life. pp592–601 Berlin: Springer-Verlag

[7] de Boer, B. (2000) Self organization in vowel systems. Journal of Phonetics 28 (4), pp. 441465

[8] Briscoe, T. (ed.) Linguistic Evolution through Language Acquisition: Formal and Computational Models. Cambridge University Press, Cambridge, UK.

[9] Briscoe, T. (2002) Grammatical Acquisition and Linguistic Selection. In: [8] p. 255-300.

[10] Brooks, P. and M. Tomasello (1999) How children constrain their argument structure constructions. Language, 75(4), pp. 720-738.

[11] Bryant, J. (2004) Scalable Construction-Based Parsing and Semantic Analysis. In Proceedings of the Second International Workshop on Scalable Natural Language Understanding. Boston 2004

[12] Cangelosi, A. and D. Parisi (eds.) (2001) Simulating the Evolution of Language. Springer-Verlag, Berlin.

[13] Chang, N.C. and T.V. Maia. (2003) *Learning Grammatical Constructions* In Proceedings of 2001 AAAI Symposium on Learning Grounded Representations.

[14] Chomsky, N. (1981) Lectures on Government and Binding. Foris, Dordrecht, 1981.

[15] Deacon, T. (1997) The Symbolic Species: The Co-Evolution of Language and the Brain W. W. Norton and Co: New York and London.

[16] Dixon, R. (1996) Australian languages. Cambridge Language Surveys. Cambridge UK.

[17] Dominey, P. (2003) Learning Grammatical Constructions in a Miniature Language from Narrated Video Events. Proceedings 25th Annual meeting of the Cognitive Society, Boston.

[18] Elman, J. (1991) Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning, 7(2/3):195–226, 1991.

[19] Ferrer-i-Cancho, R. and Sole, R. V. (2003) Least effort and the origins of scaling in human language. PNAS, 100:788–791.

[20] Garrod, S., and Anderson, A. (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition, 27, 181-218.

[21] Goldberg, A.E. (1995) Constructions.: A Construction Grammar Approach to Argument Structure. University of Chicago Press, Chicago

[22] Heine, (1997) The cognitive foundations of grammar. Oxford University Press, Oxford.

[23] Hopper, (1991) Emergent Grammar. In: Traugott, E. and Heine, B. (1991) Approaches to Grammaticalization. Volume I and II. John Benjamins Publishing Company, Amsterdam, 1991.

[24] Hutchins, E. and B. Hazlehurst (1995) How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte (eds.) Artificial societies: The computer simulation of social life (p. 157-189). UCL Press, London.

[25] Jackendoff, R. (1999) Possible Stages in the Evolution of the Language Capacity. Trends in Cognitive Science. Vol 3. 7. p. 272-279.

[26] Kosko, B. (1988) Bidirectional Associative Memories. IEEE Transactions on Systems, Man, and Cybernetics, 18, pp. 49-60, 1988

[27] Labov, W. (1994) Principles of Linguistic Change. Volume 1: Internal Factors. Oxford: Basil Blackwell, 1994,

[28] Langacker, R.W. (1991) Foundations of cognitive grammar. Stanford University Press, Stanford Ca.

[29] Levin, M. (1995) The evolution of understanding: A genetic algorithm model of the evolution of communication. Biosystems. Vol 36. p. 167-178.

[30] Lightfoot, David W. 1991. How to set parameters: Arguments from language change. Cambridge, Mass.: MIT Press.

[31] Luisi,, P.L. T. Oberholzer Origin of life on Earth: Molecular Biology as an Approach to the Minimal Cell In: The Bridge between the Big Bang and Biology, F. Giovanelli (ed.), CNR Pres., 345-355 2001

[32] von der Malsburg, C. (1995) Binding in Models of Perception and Brain Function. Current Opinion in Neurobiology 5:pp.520-526

[33] Marr, D. (1982) Vision. Freeman Publishers, San Francisco.

[34] Maynard Smith, J. and E. Szathmary (1995) The major transitions in evolution. W. H. Freeman, San Francisco

[35] Mufwene, S. (2001) Competition and Selection in Language Evolution. Selection 3 (2002) 1, 4556

[36] Neubauer, N. (2003) Emergence in a multiagent simulation of communicative behavior. PICS, Publication Series of the Institute of Cognitive Science, vol. 11, 2004, Department of Cognitive Science, University of Osnabrck.

[37] Munroe, S. and Cangelosi, A. (2002) Learning and the evolution of language: The role of cultural variation and learning costs in the Baldwin Effect. Artificial Life, 8(4):311–339.

[38] Niyogi, P. and Berwick, R. C. (1995) The Logical Problem of Language Change. Technical report, AI Lab, MIT

[39] P. Nyogi, R. Berwick, Evolutionary consequences of language learning. *Linguistics and Philosophy* **20**, 697-719. (1997).

[40] Nowak, M.A. and D.C. Krakauer (1999) The evolution of language. Proc.Natl.Acad.Scie. USA. Vol 96, pp. 8028-8033, July 1999.

[41] Nowak, M.A., J.B. Plotkin, and V. Jansen (2000) The Evolution of Syntactic Communication. Nature (404), p. 495-498.

[42] Nowak, M. A., N.L Komarova and P. Niyogi. (2001). Evolution of Universal Grammar. Science. 291: 114-117

[43] Oliphant, M. (1997) Communication as altruistic behavior. Ph.D. Dissertation. UCSD, San Diego.

[44] Partee, Barbara H. (2003) Compositionality in Formal Semantics: Selected Papers of Barbara Partee . Oxford: Blackwell Publishers.

[45] Pinker, S. (1994) The Language Instinct. Penguin Books. London.

[46] Prigogine, I. and G. Nicolis (1977) Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations J. Wiley and Sons, New York 1977

[47] Smith, K. (2002). The cultural evolution of communication in a population of neural networks. Connection Science 14(1), p65-8

[48] Smith, K., Kirby, S., and Brighton, H. (2003). Iterated Learning: a framework for the emergence of language. Artificial Life, 9(4):371–386.

[49] Sperber, D. and Wilson, D. (1995) Relevance communication and cognition Blackwell Pub. London.

[50] Steels, L. (1996) Emergent Adaptive Lexicons. In: Maes, P. (ed.) (1996) *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, The MIT Press, Cambridge Ma.

[51] Steels, L. (1996c) Perceptually Grounded Meaning Creation. In Tokoro, M., editor,Proceedings of the International Conference on Multi-Agent Systems, Cambridge, MA, 1996. The MIT Press.

[52] Steels, L. (1997) The Synthetic Modeling of Language Origins. Evolution of Communication, vol. 1, nr. 1, pp. 1-34, Amsterdam: John Benjamins Publishing Company.

[53] Steels, L. (1998) The Origins of Ontologies and Communication Conventions in Multi-Agent Systems. Autonomous Agents and Multi-Agent Systems, vol. 1, no. 1, Oct. 1998, pp. 169194.

[54] Steels, L. (1998b) The origins of syntax in visually grounded robotic agents.. Artificial Intelligence, 103:1-24 1998.

[55] Steels, L. (2000) Language as a complex adaptive system. Lecture Notes in Computer Science. Parallel Problem Solving from Nature - PPSN-VI. Volume Editor(s) : Schoenauer et. al., Springer-Verlag, Berlin.

[56] Steels, L. (2001) Language games for autonomous robots , IEEE Intelligent Systems 16 (5) (2001)

[57] Steels, L. (2003) Evolving grounded communication for robots. Trends in Cognitive Science. Volume 7, Issue 7, July 2003 , pp. 308-312.

[58] Steels, L. (2004a) Constructivist Development of Grounded Construction Grammars Scott, D., Daelemans, W. and Walker M. (eds) (2004) Proceedings Annual Meeting Association for Computational Linguistic Conference. Barcelona. p. 9-19.

[59] Steels, L. (2005) What triggers the emergence of grammar? Cangelosi, A. and C. Nehaniv (2005) (eds.) Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication. University of Herfortshire.

[60] Steels, L. and T. Belpaeme (2005) Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. Behavioral and Brain Science. To appear.

[61] Steels, L. and McIntyre, A. (1999) Spatially Distributed Naming Games. In: Advances in Complex Systems, vol. 1, nr. 4, pp. 301-323, Paris: Hermes Science Publications.

[62] Steels, L. and P. Vogt (1997) Grounding adaptive language games in robotic agents. In Harvey, P. and P. Husbands (eds.) (1997) Proceedings of the 4th European Conference on Artificial Life 97.

[63] Szathmary, E. (2001) Origin of the Human Language Faculty: the Language Amoeba Hypothesis". In: Trabant, J. , and S. Ward, ed. (2001) New Essays on the Origins of Language. Mouton de Gruyter.

[64] Talmy, L. (2000) Toward a Cognitive Semantics: Concept Structuring Systems (Language, Speech, and Communication) The MIT Press, Cambridge Ma.

[65] Tomasello, M. (1999). The cultural origins of human cognition. Harvard University Press, Cambridge, MA.

[66] Van Kemenade, A (1987) Syntactic Case and Morphological Case in the History of English. Forist Publications, Dordrecht.

[67] Van Looveren, J. (1999) Multiple word naming games. In Proceedings of the 11th Belgium-Netherlands Conference on Artificial Intelligence. Universiteit Maastricht, Maastricht, the Netherlands, 1999.

[68] Vogt, P. and Coumans, H. (2003) Investigating social interaction strategies for bootstrapping lexicon development Journal of Artificial Societies and Social Simulation 6(1)

[69] Wagner, K., J.A. Reggia, J. Uriagereka, G.S. Wilkinson (2003) Progress in the Simulation of Emergent Communication and Language. Adaptive Behavior. Volume 11 (1): 37-69.

[70] Webster, M. A. and Kay, P. (2005) Individual and population differences in focal colors. In: The anthropology of color, ed. R. L. MacLaury, G. Paramei and D. Dedrick. Benjamins.

[71] Weisbuch, G., G. Deffuant , F. Amblard and J-P Nadal (2003) Interacting Agents and Continuous Opinions Dynamics. Santa Fe Institute Working Paper 01-11-072

[72] Werner, G. and M. Dyer (1991) "Evolution of Communication in Artificial Organisms", in Langton, C. et al. (ed.) Artificial Life II. Addison-Wesley Pub. Co. Redwood City, p. 659-687