THE ORIGINS OF LANGUAGE – DARWIN'S UNSOLVED MYSTERY

LUC STEELS

1. Introduction

Linguistics in the second half of the 20th century was dominated by the framework of generative grammar, which was proposed by the American linguist Noam Chomsky in the nineteen-fifties (Chomsky 1965). Mainly under the impetus of the Swiss linguist Ferdinand de Saussure, linguists had focused already from the beginning of the 20th century on trying to capture "the system" that underlies a given language. They had been busy identifying the sound systems used by various human languages: the vowels, the consonants, the constraints on syllables and word structure, and the possible intonation and stress patterns. They had been trying to catalogue and systematize the set of possible syntactic structures that may be used by the grammar of a language and to find the underlying structures of the vocabulary of a language, primarily by identifying the primitive meaning components common to a large range of words or the semantic constraints that a verb may impose on its objects. All of this is extremely useful, particularly if you want to learn a foreign language or get an insight in how your own native language works.

In the second half of the 20th century, Chomsky pushed this structuralist research program to a new level in three ways. First, he proposed that the system underlying a particular human language could be captured in terms of a formal calculus, and he developed a formal framework known as generative grammar to do this. A generative grammar defines the set of all possible sentences of a language in terms of a set of rules to generate them. It is like defining the set of prime numbers by giving an algorithm that generates one prime number after another. This gave linguists a mathematical tool for developing objective, precise definitions of a language and they enthusiastically threw

Tuesday Colloquium held at the Wissenschaftskolleg on 12 May 2009.

themselves at the task. They were thus able to follow in the footsteps of logicians, who had already a longer tradition, started by Russell and Whitehead in their Principia Mathematica, of formalising meaning in terms of a clean, explicit calculus so that the process of reasoning could be defined in a way similar to that for numerical calculation or algebra.

Second, Chomsky proposed that this generative grammar framework is also a theory of the information processing that the brains of human speakers and hearers carry out when they parse and produce sentences. A similar move was made in the late nineteen-fifties by John McCarthy (1958), who had proposed that the logical calculus designed to formalise the set of possible meanings that could be expressed by human language, could also be the basis of operational reasoning systems that would automatically derive proofs for any proposition that one might want to examine. Shifting from a generative framework to an operational framework is not as obvious as it may sound. Very often it is highly non-trivial to go from a procedure to generate the elements of a set to a procedure for recognising or producing those elements in order to satisfy a particular function. For example, it is easy to generate the set of all squares of the natural numbers but it is far from easy to determine what the square root is of 232400783.

Chomsky made a third incredibly bold proposal. He argued that the core of the calculus circumscribing a language is universally shared by all human languages and that this Universal Grammar or "linguistic Bauplan" had been fixed at the dawn of humanity as part of the set of genes that make us humans unique. From this perspective, different languages (Hindi, English, Bantu, Hungarian) vary only in terms of their vocabulary and in terms of the choices they have made within the narrow options allowed by Universal Grammar. The basic structural patterns (for example that a normal declarative sentence is made up of a subject, a verb and an object) and the categories to define these patterns (for example the distinction between noun and verb, subject and object, nominative and dative, count noun and mass noun) are all considered to be universal and innate. So Chomsky's theory is not only a linguistic theory, because it provides a framework for describing languages, and a psychological theory, because it claims to specify the nature of human language processing, but also a biological theory, because it gives a strong hypothesis about what structures the brain innately possesses to perform language processing and acquire language. No wonder that this encompassing vision had such a big impact on the many scientific disciplines interested in language.

Although Chomsky has shied away from addressing the fascinating question of the origins and evolution of language, many scholars, particularly Steven Pinker (1994), have

argued that if there is a highly specialised organ in the brain for language, then it must have evolved like any other organ, i. e., through genetic evolution and natural selection. Very similar claims have been made for the conceptual apparatus that is needed to formulate and apply meanings to reality, particularly by Jerry Fodor (1983). The conceptual-intentional system is considered to be based within another "brain organ", which is genetically equipped with the innate concepts available for conceptualising and structuring the world.

No one denies that Chomsky's revolutionary proposals have been fruitful for linguistics as a science. They also have had an important impact on information technology, because computer scientists and AI researchers have seriously tried to operationalise generative grammar and they have tried to build various applications for information retrieval, automated language translation, human-robot interaction, text correction, etc. But half a century later, the limitations of the paradigm have become very apparent.

The first fundamental problem is that there is no clear-cut static system, uniformly known and perfectly used by all speakers of a language community. Observations of natural dialogue and language use show immediately that there is huge variation and that all elements of a language undergo constant change by the individual actions of language users, even in the course of a single dialogue. New sounds get into a language or the sounds normally made by a speaker get modulated. Whether you like it or not, you are influenced by the way your dialogue partner pronounces words, and when you speak a lot with people from another dialectal group your own speech system starts to change. We are social chameleons who want to be and behave like others.

Other aspects of language behave likewise. New words pop up all the time, simply because language users need to express a never-ending stream of new meanings. The meaning of existing words is constantly being stretched and expanded to handle new situations. New grammatical constructions arise and become fashionable for a while, and existing words or constructions are coerced into new uses. New conceptualisations of reality arise as we discover more about our world and ourselves or create virtual worlds, such as the Internet and the World Wide Web. Careful analysis of natural dialogue by psycholinguists like Garrod and Pickering (2006) has shown that the rapid adaptation and implicit negotiation of language happens even at the conceptual level. Two partners might start out with quite different concepts of the colour "mauve" for example, but then gradually coordinate their colour prototypes and settle on a common understanding as the dialogue takes its course. Often these inventions, variations and adaptations do not

survive beyond their short-term usage within a particular dialogue, but some of them do and then they might propagate further, like viruses. Propagation can go extremely fast. It is just unbelievable how Internet-related terms like spam, email, chatting, browsing, uploading, blog, tagging, phishing, etc. have spread worldwide within the time-span of just a few years. When language innovations and adaptations accumulate, layer upon layer, they lead to the long-term observable language change that you see when comparing Middle English and modern English, or classical Latin and Italian.

Of course there is systematicity in language use, both in terms of the ideolectal habits of a single speaker and at the communal level of a speech community, otherwise understanding among individuals would be impossible and language evolution could not be cumulative. But there is a big difference between a clear static system that can be captured by a formal calculus and systematic trends that are temporary and always on the move. The great strength of human languages is precisely their open-ended, fluid character, so that they can adapt extremely quickly to cope with the never-ending stream of novel meanings that need to be expressed and the changing social functions and evolving social strata of speakers. Consequently it has turned out to be very difficult to press natural languages into a formal calculus, which explains why it is still not possible to build artificial natural language systems that can interact fluently with human users about the world.

What about Universal Grammar? Here a second fundamental problem has come up. Linguistic typologists, like Martin Haspelmath (2007), who are in the business of comparing different languages on the basis of real language data from a wide variety of languages, have come to the conclusion that linguistic categories (such as noun, dative, agent, etc.) are not universal nor uniform across a language community, let alone that all possible patterns of usage can be pressed into a fixed enumerable set with a finite set of parameters (Evans and Levinson 2010).

This does not mean that there are no universal tendencies in human languages. The vowel systems of the world exhibit certain regularities with some constellations more common than others. Which colour categories are lexicalised as basic colour terms follows certain trends among human languages, even if a specific language may still use a set entirely different from this trend. Human languages are not only compositional in the sense that they use multiple words to convey complex meaning as opposed to a single holistic sign, but also in that they all use elaborate systems of syntactic and semantic categories to map meaning to form. All human languages have a way to express spatial relations from a certain vantage point, as in “the car behind the tree”, or “the block to your left”. So we

find numerous regularities, but is this because these properties are determined by an innate language organ or conceptual-intentional system, or is it because these solutions are the natural outcome of the processes by which humans invent, adapt, and negotiate shared communication systems with the embodiment and the cognitive functions also available for other tasks?

If universal trends in language are an emergent side effect and if languages are in constant flux with grammar being adaptive rather than static and homogenously shared among the members of the population, then how should we go about studying language? What implications do these observations have for theories of human language processing? Where is the invisible hand that ensures there is still enough systematicity and coordination between different speakers so that they have the high success rate in communication that we tend to see? How can we explain the undeniable trends in the grammars, vocabularies and conceptualisations used by human languages?

These puzzles led me in the mid-nineties to develop a new paradigm for the study of language and thus lay the foundation of evolutionary linguistics (Steels 1997). In the remainder of this essay, I summarize some of the main ideas and how we have been trying to work them out and substantiate them.

2. Language Games

The first idea is to view a population of language users as a complex adaptive system like an ant colony. Although complex adaptive systems can be studied by formulating equations that govern the aggregate behaviour of the population, computer scientists pioneered in the early eighties an "agent-based modelling" approach, which has since been applied to hundreds of phenomena in biology, economics, sociology and other fields in which complex adaptive systems are studied. Basically, you consider each component of the system to be an agent if it has an internal state and then you define a script that specifies how the agent should react to events in a model of the world. The phenomena of interest should then emerge in the simulation through the interactions of the agents. You can then vary parameters or vary the architecture of the agents in order to test the degree to which the model captures the necessary and sufficient properties of an agent as needed to give rise to the phenomena one is trying to understand.

Agent-based models relevant for understanding emergent communication naturally take "language agents" as their primitive elements. Such agents would need all the neces-

sary machinery for perceiving and acting in the world, conceptualising what to say, interpreting what has been said, and parsing and producing sentences. However the agents should not have any innate set of concepts, no innate knowledge of the world, no innate Universal Grammar and no specific knowledge of the language used by the other agents. Indeed we want to understand how agents are themselves able to bootstrap ontologies and communication systems without human assistance and without an existing human-invented system. Because we want to understand how language can be about the world, the agents should ideally be physically embodied as opposed to being agents in a virtual world, and this implies that we have to use robots for the experiments.

What kind of interaction patterns should we set up for such language-agent experiments? Here is the second key idea: we should focus on language games, thus revitalising a suggestion already made by the philosopher Ludwig Wittgenstein. Wittgenstein (1953) was one of the first to see the restrictions of studying only the semantics of isolated sentences devoid of context and use. He realised that language interacts strongly with conceptualisation and communicative function. He suggested that a word is like a chess piece in the sense that its meaning comes from the role it plays in an overall system. And he argued that, without studying language games and how they become set up and evolve, you cannot understand the origins of meanings or how meanings become expressed in language.

Human language games fail quite often – more than we tend to believe – but this is seldom catastrophic. The context helps a lot to restrict the set of intended meanings and failed communications can usually be repaired easily. In fact, the creation of new meanings and the acquisition and negotiation of linguistic conventions is an integral part of human language games. Human language games are fluid. The turn-taking, conceptualisations of reality and linguistic conventions used are not entirely fixed. In a game of chess, the role of each piece is strictly defined in advance and the shape of each piece is settled in the beginning of a game. Players would heavily protest if a knight suddenly turns out to be the king or if the rules for moving a pawn suddenly change. This is not the case in human language games. New conventions emerge, possibly shared only by a small group initially, but they are quickly picked up by others and propagate through consecutive situated embodied language games because of the pragmatic feedback and repair strategies of the partners. We see similar phenomena arise in the text messaging for mobile phones, which generate whole new language variants unknown in existing languages.

Agent-based modelling and language games provide a framework with which we can study language and meaning as complex adaptive systems. In my group we are now doing this in a systematic way, using the following methodological framework:

a) Select a subsystem of language and meaning for which you want to understand how it could form and remain adaptive in a population of distributed agents. For example, colour categories, spatial prepositions, case grammar for expressing the role of participants in events, body posture language and its metaphorical extensions to space (as in "the bottle on the table"), a grammatical system for marking tense, aspect and modality, a system of determiners, etc.
b) Define a language game within a contextual setting to understand the ecological or functional significance of the language subsystem. An experimental set up may consist for example of a physical environment in which one agent has to draw the attention of another agent to some object in the world, or an action game in which one agent asks another agent to make a certain movement.
c) Reverse engineer a natural system found in an existing human natural language, which means: reconstruct the ontology, the lexicon and the grammar and operationalise the necessary comprehension and production processes to show that the reverse-engineered system is adequate for comprehending and producing the utterances needed in the language game. Examples of natural systems could be: tense in French, aspect in Russian, colour terms in Spanish, articles in English, body posture expressions in Dutch, role marking in Japanese. By first reverse engineering a natural system, we demonstrate that research results will have empirical validity. Note that not only the linguistic parts, but also perception, action, conceptualisation and semantic interpretation have to be operationalised.
d) Reverse engineer the learning strategy, so that an artificial agent can acquire the natural system through situated embodied interactions with other agents who already master the natural system. For example, a group of robots is programmed with the Spanish colour terms and a new robot then has to acquire this vocabulary and its underlying colour categories by interacting with them. This will demonstrate that the reverse-engineered learning strategy is adequate to acquire the relevant area of natural language.
e) Reverse engineer the invention, adaptation and alignment strategies that artificial agents need to self-organise a symbolic communication system from scratch and to remain adaptive even if the environment changes. This symbolic communication will

be similar but not identical to the lexical and grammatical systems found in human languages, because it uses a similar strategy. Increase in communicative success is the ultimate criterion to evaluate whether the experiment was successful.

f) Finally investigate how the strategies themselves may arise and how they propagate and remain as part of the strategic toolkit of the language based on the average communicative success that users of the strategy have had in the population. For example, how a strategy for expressing aspect may arise in a language and remain in widespread use.

This methodology leads to clear, repeatable experiments, and the next section describes very briefly a concrete example.

3. The Emergence of Colour Terms and Colour Categories

Let us examine an experiment for the emergence of colour terms and colour categories, reported earlier in Steels and Belpaeme (2005). We have done much more complex experiments in our group, but their discussion falls outside of the scope of the present essay.

3.1. Identify a Language Subsystem

Most languages have a vocabulary of basic colour terms. Anthropologists have studied these intensively, often by asking human subjects to name Munsell colour chips or which chip they consider the most representative example of each basic colour term in their language (Berlin and Kay 1969). For example, the eleven basic colour terms in English are: black, white, red, green, blue, yellow, pink, purple, brown, orange and grey. There are known to be universal trends in the basic colour terms of human languages but there is also a lot of variation, not only in terms of the names that are chosen for colours but also in terms of the perceptually grounded colour categories that they express. As evolutionary linguists, we want to understand why and how such a system of basic colour terms could arise.

Several models of the processing needed for basic colour terms have already been discussed in the literature, starting with Rosch (1975). Most of them centre on the notion of a 3-dimensional colour space formed by the red-green and yellow-blue opponent channels and the brightness channel. Colour prototypes that are the most typical example of a colour are mapped as points in this colour space and colour categorisation can be achieved

using a nearest-neighbour computation: a sample to be categorized is compared to all prototypes in the inventory and the prototype that is nearest to the sample is regarded as identifying the matching category. This kind of information processing can be operationalised easily by neural networks (for example radial basis function networks) or by a straightforward computational implementation of nearest-neighbour computation.

When we restrict ourselves to basic colour terms, the linguistic component is straightforward as well. It consists of a bi-directional associative memory that associates colour categories (or more precisely their prototypes) with colour words. When the speaker needs to name a colour, he should first find the nearest colour prototype and then look up the colour term in his lexicon associated with this prototype. When the listener wants to understand a colour term, he looks up which prototype corresponds to this name in his own lexicon and then searches in the context for an object that matches the closest to the prototype that was communicated.

There has been a lot of empirical psychological research to find out what colour prototypes the speakers of a given language employ, and these data can directly be plugged into an operational model of colour comprehension and production processes. For example, Lillo and colleagues (2007) have used the CIE L*u*v colour space (where L is the brightness dimension and u and v are dimensions modelling the human opponent channels) to identify the prototypes that Spanish speakers associate with the eleven basic colour categories in Spanish: blanco, negro, rojo, verde, amarillo, azul, marrón, rosa, naranja, morado, gris. For example, the prototype for "verde" (green) is located at the point <L=44.85,u=38.42,v=29.15> in the L*u*v colour space. These values can be used to simulate in a realistic fashion the colour naming of Spanish speakers.

3.2. Identify the Function in Communication by the Design of a Language Game

An inventory of perceptually grounded colour prototypes for Spanish and a lexicon associating these colour prototypes with their Spanish names constitutes the basic colour language subsystem of Spanish, so we can now turn to the second stage of the methodology and address the question: What is this language subsystem for? If colour terms have no purpose whatsoever in human communication we would not expect to find them in human languages. There are in fact several functions. Here I will just focus on one, namely reference: A speaker can use basic colours to draw the attention of the hearer to an object in the world by naming its distinctive basic colour. This kind of interaction can be cap-

tured quite succinctly in a language game, which I have called the Colour Naming Game:

The Colour Naming Game assumes an open-ended set of possible contexts consisting of objects of different colours. To play the game, two agents are randomly chosen from the population. One of them takes on the role of speaker and the other that of hearer. They then go through the following interaction:

a) The speaker chooses one of the objects in the context as topic.
b) The speaker categorizes the colour of the topic based on his internal inventory of colour prototypes.
c) The speaker looks up in his lexicon the colour term associated with the prototype and transmits this to the hearer.
d) The hearer looks up the prototype associated with this colour term in his own lexicon.
e) He then selects the object whose colour matches the closest with the prototype and points to this object.
f) The speaker checks whether the object pointed at is the one he originally chose. If that is the case he signals success.
g) If it is not the case he signals failure and points to the correct object.

The Colour Naming Game can be played with contexts consisting of Munsell chips so that we can approach anthropological test conditions. But they can also be played with real world objects, and indeed we have already done experiments in which the language game is implemented using autonomous robots playing the game about the colourful objects they encounter in the room.

Figure 2 shows the result of a language game experiment in which two agents were endowed with the basic Spanish colour language subsystem discussed earlier. Each game involves a randomly assembled set of Munsell chips. One Munsell chip is chosen as topic by the speaker and its colour named. The listener needs to guess which chip was intended and the game is a success if the listener was able to do this. We see that the agents are not always successful (average success rate is 90 %) because in some cases the colour of the topic chosen by the speaker is so close to that of another chip that they can no longer be distinguished by basic colour prototypes. This would be the case if the second or the fourth chip in the context shown as inset in Figure 2, because they are both close to the prototype for "gris" (grey). In that case a more sophisticated colour expression (such as

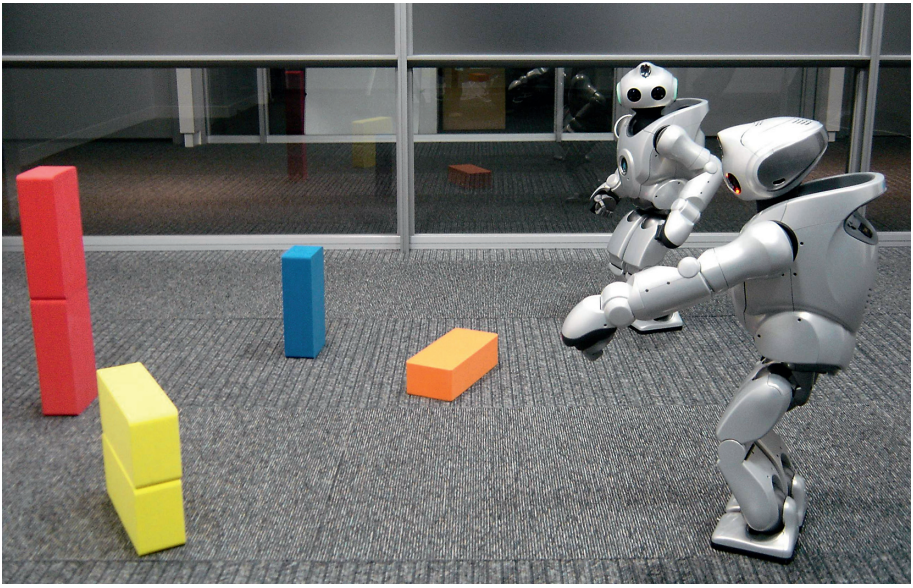

Fig. 1: It is possible to carry out language game experiments with autonomous robots, which implies that the script of the game, the perception and action, the parsing and production and the learning, expansion and alignment functions of language strategies are completely operationalised.

"slightly light grey") would need to be utilized that goes beyond the language strategy under investigation here. On the other hand, if the first chip is chosen, it matches distinctly with the prototype for "amarillo" (yellow) and a successful game is possible.

3.3. Understand How the Language Subsystem Gets Built, Given a Strategy

Figure 2 shows that stage 1 and 2 could be successfully concluded: We were able to operationalise the language subsystem of interest (the Spanish basic colour terms) and show its utility in the context of a particular language game (the Colour Naming Game). Next we investigate what kind of language strategy is able to learn and bootstrap such a colour system.

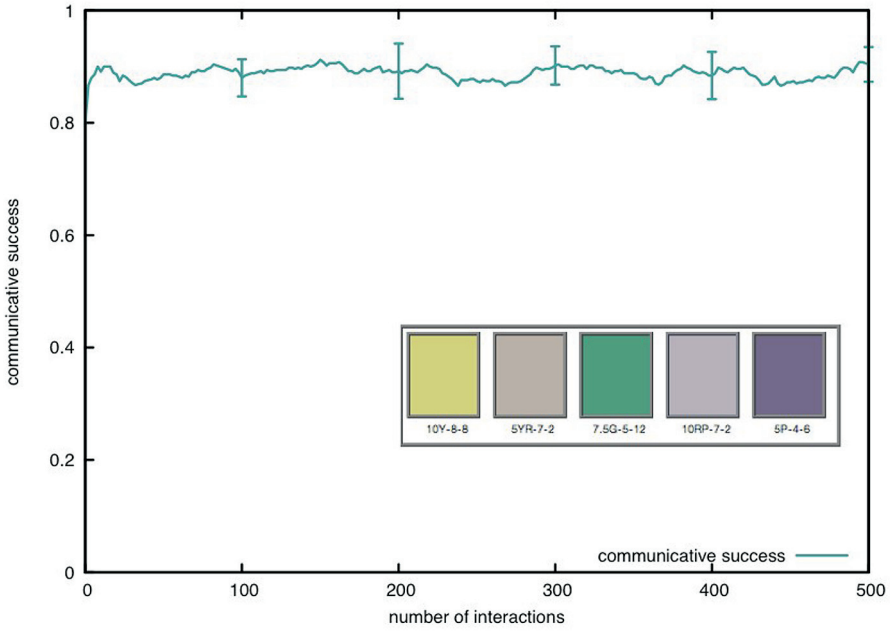

Fig. 2: Graph showing the results of 2 simulated Spanish speakers playing a series of 500 Colour Naming Games (x-axis) establishing the base-line performance for this micro-ecology. The running average of communicative success is shown on the y-axis. The inset shows a typical example context. About 90 % success is reached with the Spanish basic colour terms.

The first step is to identify and operationalise functions for learning the basic colour prototypes used in Spanish, and for acquiring the words for these prototypes. Learning colour words is relatively straightforward once the colour categories have been mastered: When the listener encounters a word not heard before or when he uses a word in a wrong way (from the viewpoint of the speaker), the game is a failure and the speaker points to the topic he chose. Because the listener can already categorise the colour of the topic, he can infer the possible meaning of the uttered word and store that in his lexicon. The association is only an hypothesis that has to be confirmed by further interactions. Agents must therefore maintain a score between words and meanings. They should prefer to use

the association with the largest score because it is their best hypotheses so far. If there is a successful game, then the score of the association that was used is increased and the competing associations (other words with the same meaning for the speaker or other meanings for the same word for the hearer) are decreased. If there is an unsuccessful game, the score of the used association is decreased. This lateral inhibition dynamic has now been widely employed and studied as an adequate vehicle for modelling how a population settles a convention. It can be operationalised using neural networks (such as bi-directional associative memories) or straightforward computational implementations.

How are colour prototypes learned? When the listener encounters a new word and gets feedback on the topic from the speaker, he categorizes the topic himself to guess the possible meaning. But it is possible that there is no distinctive category yet. In that case, the listener should introduce a new prototype, using the sample that acted as topic as a seed. As in the case of words, the new prototype is only an hypothesis that needs to be confirmed by further interactions. Agents must therefore maintain a score on the utility of a prototype and they should shift the prototype in the face of new evidence. For example, they should shift the prototype that was successful in the language game slightly in the direction of the topic. When agents keep doing that over a series of games, their prototypes will not only become more similar, they will also become and remain adaptive to the situations they effectively encounter in their world. The information processing required to operationalise this learning strategy can be achieved with neural networks (for example radial basis function networks) or with a straightforward computational implementation of the same functions.

Figure 3 shows the results of an experiment to test both of these learning mechanisms. It is an experiment involving two artificial agents. One is acting as tutor and has been programmed with the Spanish colour language subsystem (from stage 1). The other is a novice. He starts without an inventory of perceptually grounded categories and without a lexicon for expressing them. The novice has been programmed with the learning strategy discussed above. We see that the learning strategy is entirely effective. The novice quickly reaches the same level of performance as the tutor.

We next turn to the alignment and expansion strategies. In this particular case, the alignment strategy is already part of the learning strategy discussed earlier. After every game, speaker and hearer adjust the scores of the associations in their lexicon and they shift prototypes and keep track of their utility. Consequently their language subsystems become more and more aligned. This is shown with an additional graph in Figure 3 (in-

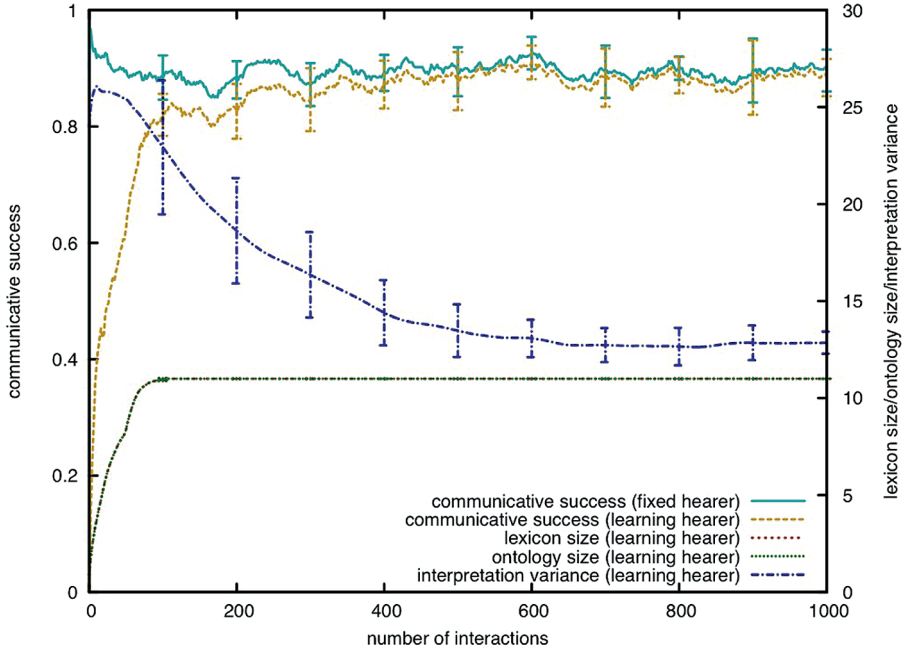

Fig. 3: Graph showing baseline communicative success (left y-axis) with 2 simulated Spanish speakers as in Figure 2 and a sequence of language games (y-axis) between a novice and a tutor, using the same contexts as for the baseline. We see that the novice progressively reaches the same level of communicative success as the base line. The same number of eleven basic colour terms as the simulated Spanish speakers (right y-axis), and interpretation variance drops showing that the categories of the agents become similar.

terpretation variance), which displays the average distance between the colour prototypes of novice and tutor. The distance (measured in the L*u*v space) never becomes 0.0 but is small enough to support successful communication.

The expansion strategy needed by the speaker is reminiscent of the learning strategy used by the listener. When the speaker is unable to find a distinctive prototype, for example because several samples including the topic are equidistant from the same prototype, then the speaker should introduce a new prototype by taking the topic as initial seed. And

when the speaker has no word yet to name a distinctive prototype, he can invent a new name, for example by choosing a random combination of syllables, and adding a new association to his lexicon. Once introduced, the learning strategy and the lateral inhibition dynamics of the naming game do their work and the invention potentially spreads in the population.

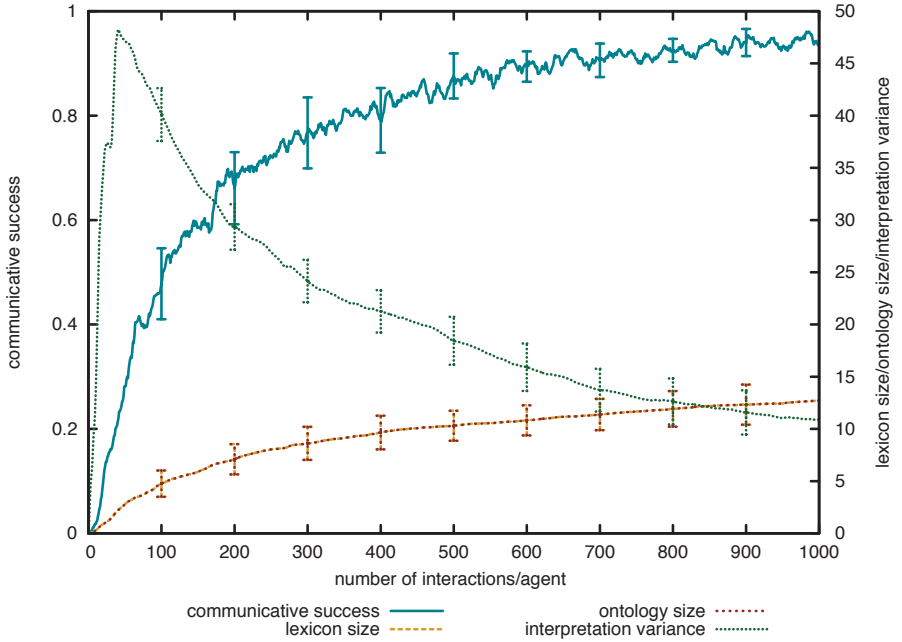

Fig. 4: Experiment in which a population of 5 agents self-organises a colour language system from scratch, given a strategy. The graph shows communicative success (left y-axis), average lexicon size (right y-axis) and interpretation variance. After about 500 language games (per agent), the agents reach a similar level of communicative success as the (simulated) Spanish speakers do.

Figure 4 shows that all this works beautifully. A population of five agents starts from scratch with zero communicative success but rapidly reaches a success rate comparable to

the Spanish basic colour system. There is no fixed limit on the number of basic colour prototypes and so the agents keep refining their ontologies and keep inventing new words, so that they actually reach a higher level of success than the Spanish system, which has only 11 basic colour terms. The most common words in one run of the experiment are: vamasi (greenish), fidate (brownish), bamoru (black), bamova (bluish grey), riveke (purple) and kenafo (brownish grey). Every time the experiment is run, another colour language, including another set of perceptually grounded colour categories, emerges.

Figure 5 shows the colour prototypes for these words after 1000 games (left) and after 2500 games (right). We see that the prototypes of the 5 agents start to look more and more similar. This emerging coherence is remarkable because there is no central supervisor or controlling authority, no prior knowledge of the categories or lexicon and no telepathic relation between the agents. The colour categories in this artificial language are not identical to those of Spanish speakers and that cannot be expected. In fact, it is still an open question what additional constraints need to be imposed on the micro-ecology of the agents or their perceptual and cognitive apparatus in order to see colour language subsystems that exhibit the kinds of trends seen in human colour language subsystems, but at least we now have a very clear framework to investigate this.

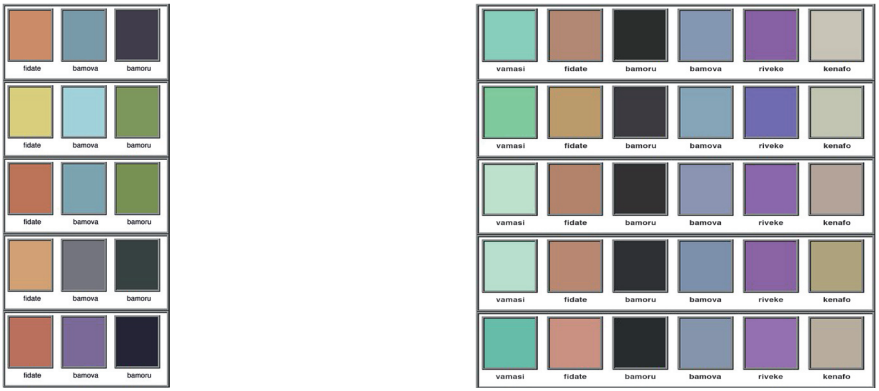

Fig. 5 a, b: Basic colour prototypes of each agent (from top to bottom) for the most dominant words. Left: after 1000 games. Right: after 2500 games. The alignment strategy causes not only the words but also the prototypes to become similar, and this will increase the chance of communicative success.

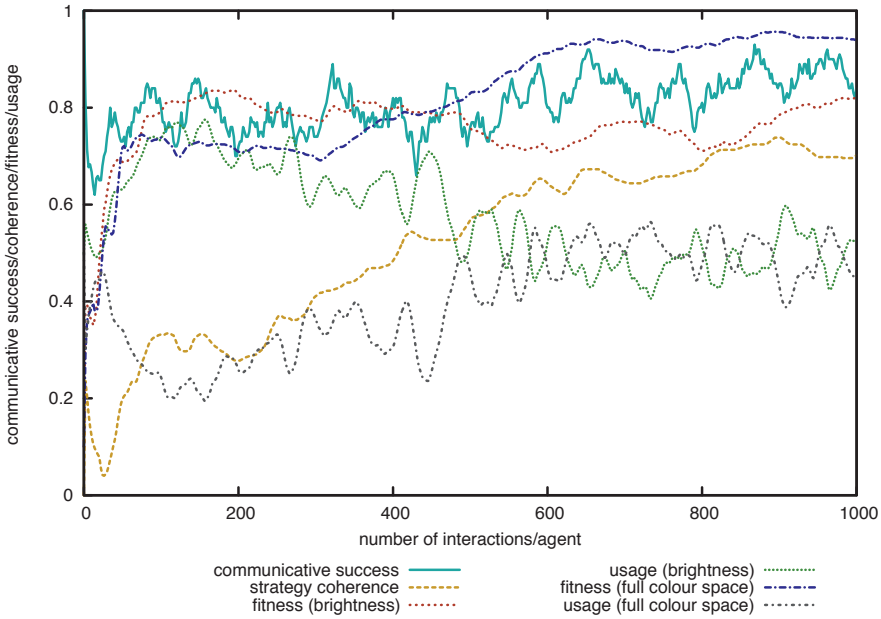

Fig. 6: This figure shows the results of an experiment in which a brightness-based strategy competes with a full-colour strategy in a population of agents. Communicative success hovers around 85 % even though there is a shift in which brightness first dominates and is then overtaken (after about 410 games per agent) by the full-colour strategy. The graphs show the communicative fitness (the running average tracked by each agent) as well as the frequency of choice for each strategy. The evolution of strategy coherence, which is the degree to which the different agents share the same language strategy, is shown as well. It starts from low values to reach 70 %.

3.4. Understand Where a Strategy Comes from

The final type of experiment investigates the rise and competition of language strategies. Different language strategies for similar ecological conditions and communicative goals are made available to the agents in a population at the start of the experiment, and the question is whether and how they will be able to select the strategy that is most adapted to their micro-ecology. If the selectionist logic has been captured properly, we should see

that long-term communicative success with language subsystems built by one strategy should give that strategy a selectionist advantage so that it comes to dominate in the population. An example of such an experiment for the domain of colour is reported in Bleys and Steels (2009), using the brightness-based and full-colour language strategies discussed earlier. Results are shown in Figure 6. We see that in a first phase the brightness-based strategy is winning, in the sense that it was used more often by the agents to invent or interpret new words. Progressively the full-colour space strategy overtakes this initial advantage. Interestingly, the same word can temporarily be used by one agent using the brightness-based strategy and by another agent using the full-colour strategy (as happened in the evolution of English with the word "yellow" for example).

4. Evolutionary Linguistics and Evolutionary Biology

For the past decade, I have been doing experiments of this kind with my collaborators for different domains of the lexicon and grammar (Steels 2004). We also now have mathematical models of how conventions can become shared (Loreto and Steels 2007) and powerful tools for setting up and running experiments (Steels and De Beule 2007). The question is now whether we can learn anything from all this with respect to the fascinating question of the origins and evolution of human languages. I discovered at the Wissenschaftskolleg that our approach to evolutionary linguistics is quite similar to the methodology adopted in evolutionary biology and consequently that the selectionist explanatory framework introduced by Darwin can in fact also be applied to language, except that we have to map it from the biological to the cultural domain.

What exactly is the methodology adopted by evolutionary biologists? Roughly speaking there are the following steps:

a) Select and describe the phenotypic trait of interest, for example, the colour of butterfly wings, patterns on the fin of a fish, lungs in vertebrates, songs of finches. The trait may be chosen because it is particularly relevant for understanding the evolution of a species or because it may shed light on evolutionary processes in general.
b) Understand the ecological or functional significance of the trait. This is done by looking at the role of the trait within the functioning of the organism (for example oxygen supply) or within the behaviours and interactions the organism has within the ecosystem in which it attempts to survive and reproduce (for example, colour may play a role in attracting a mate).

c) Understand how the trait becomes established. For physiological traits, this is through a combination of genetic and developmental processes. For behavioural traits, like strategies for catching prey, this is through neuronal growth processes and learning.
d) Understand how the trait may have appeared in evolution. This amounts to figuring out when and where in evolution the genetic basis for the trait appeared and what kind of genomic changes might have taken place.
e) Show that the trait has a selective advantage. This is achieved by comparing the effect of having or not having the trait on the fitness of individuals in the ecosystem, for example by investigating how different variants in a population are selectively able to survive in different circumstances.

Once all these points have been clarified, the Darwinian selectionist logic provides the explanatory glue: When the trait has a selective advantage, the relevant genes will proliferate in the gene pool and be preserved in subsequent generations. Thanks to heredity, the trait can also be further refined and used as a building block for more complex traits.

Examples of how this methodology is being applied can be found abundantly in biological journals. Here is one example presented at the Wissenschaftskolleg by Axel Meyer: the explanation for the "eggspots" on the haplochromines, certain types of cichlid fish (Salzburger, et al. 2005). These branches of cichlid fish have drawn the attention of evolutionary biologists because they have very rapidly diversified within the lakes of East Africa into several hundred species and therefore seem to defy the normal pace of evolution. One distinctive phenotypic trait of haplochromines is circular eggspots on the anal fin in males. They are called eggspots because they look like the eggs that female cichlids produce. Why are these spots there and how did they contribute to successful speciation?

The first step is to understand the ecological or functional significance of this trait. We first need to know that these cichlid fish have developed maternal mouth-brooding of eggs. Females lay eggs and then suck them up again and continue to brood the eggs in their mouth. This gives a selective advantage because it helps to protect the eggs against predators. The male needs to fertilize these eggs. By presenting his anal fin, which contains the eggspots, the female "believes" there are more eggs to be sucked up, and thus receives the male's sperm, which fertilizes the eggs already in her mouth. So this explains the ecological function of the egg spots.

Next, the complete genetics and developmental process were established that show how these eggspots are laid down in a process of pattern formation under genetic control (Salzburger, et al. 2007). And once the genetic basis was known, it could be reconstructed

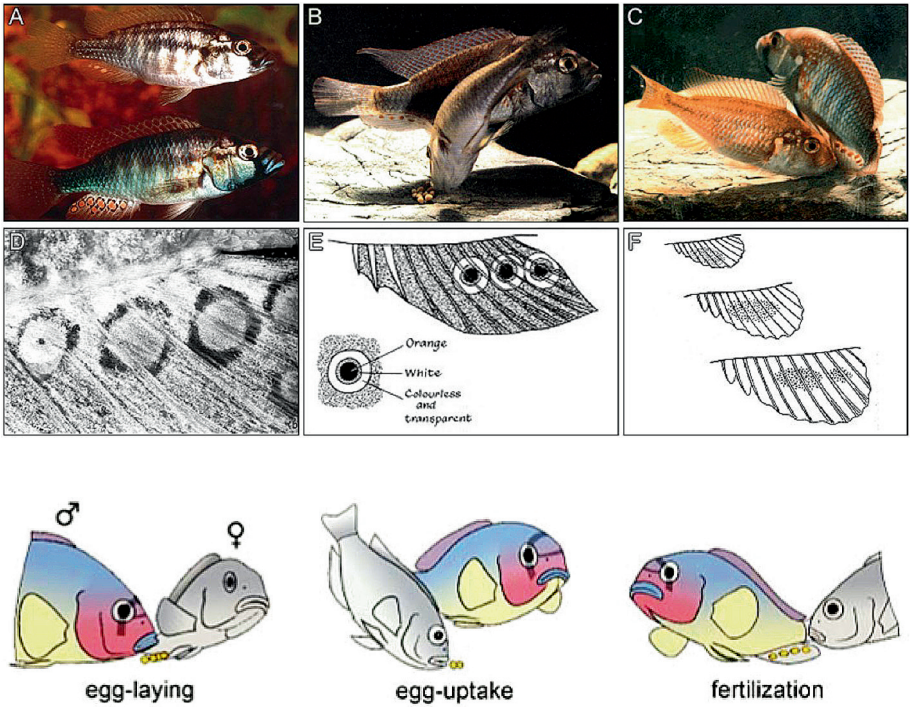

Fig. 7 a, b: Males of one species of cichlid fish (figure A) have spots on their anal fin that look like eggs (figures D and E). Females suck up eggs and brood them in their mouth (figure B). Because the eggspots act as dummy eggs, females are enticed to suck them up but receive the male's sperm instead (figure C).

that these eggspots appeared once in the ancestral line of this cichlid branch, although some other species lost it again, for example because they invaded deep-water habitats that were too dark for the eggspots to play a role, and that the genetic change involved a "co-option" or exaptation of an already existing genetic mechanism inducing pearly spots on anal or other fins.

The reproduction strategy based on mouth-breeding has a clear selective advantage for the male because he can be much more sure that he is the one fertilising the eggs.

There is also a selective advantage for the female because she is surer to be fertilized after courtship by the male she had chosen. Consequently the Darwinian selectionist loop can be closed. The selective advantage will cause the relevant eggspots genes to proliferate and we thus get an evolutionary explanation for this phenotypic trait.

Here is then a possible mapping from biology to linguistics (summarized in the table below). The analogue of a phenotypic feature is a language subsystem. A language subsystem determines certain features of an utterance, which then gives success or failure to the speaker or the hearer. Phenotypic features are built by the interaction between genes or gene networks and the environment. So the analogue of a gene or gene network is a language strategy, because just like the genes, language strategies build concrete language subsystems in interaction with the cognitive, social and linguistic environment. The equivalent of the utility of a phenotypic feature is communicative success of the utterances in which a language subsystem has participated and hence the equivalent of fitness is the rate of communicative fitness of the language strategy used to help build a particular language subsystem.

biological	linguistic
genes	language strategies
phenotypic traits	language subsystems
featurese of behavior	features of utterances
behavioural success	communicative success
biological fitness	communicative fitness

What does this analogy tell us? First of all, it helps us to see that language evolution could be based on the same principle as biological evolution, namely selectionism, but now applied at the cultural level rather than the genetic level. Communicative success steers both the choices that are made to work out the details of a language subsystem, for example which colour terms or colour categories are adopted, and the choices to adopt a certain language strategy or not, for example to use a brightness-based or hue-based strategy. Communicative success includes not only adequate expressive power. Success also depends on whether the linguistic conventions and conceptualisations used by the speaker are shared by the hearer. Hence the use of communicative success as a selectionist criterion will drive the population towards a shared language system, as indeed we observe in the simulations.

Needless to say that it is important to keep not only the analogies but also the differences in mind. There is clearly heredity involved in language in the sense that language subsystems and language strategies are preserved through the memories of the individuals using them. But there is no direct physical copying going on of language strategies between the brains of individuals. There is no telepathy. Every individual has to the strategies necessary to deal with the language of his community and then enact them to acquire the specific conceptualisations and conventions that are in common use. Hence, the innovation and variation in language strategies cannot arise from errors in physical copying or from the recombination of strategies from the two parents. They are due to the fact that the constructions or reconstructions of strategies carried out by different individuals are always a matter of guessing and trying. Nevertheless, the analogy still remains useful because we are dealing in both cases with a selectionist system.

A second important insight that we get from these experiments and analogies is that it is very unlikely that we will be able to find a single language gene or a few language genes that shape a highly specialised modular "language organ" in the brain. Language appears to require a large number of cognitive functions that are also useful for other tasks. They are recruited for the task of symbolic communication (Steels 2007). This means that the defining neurobiological characteristic of Homo sapiens does not come from the evolution of highly specialised modules, as often advocated by evolutionary psychologists, but rather in its ability to flexibly configure a rich set of cognitive functions. A key open question for evolutionary biology is when and how such high plasticity developed in the ancestors of our species.

Acknowledgement

The ideas expressed in this paper have been greatly aided by the presentations and discussions I was able to have at the Wissenschaftskolleg with the evolutionary biologists: Jeffrey Feder, James Mallet, Axel Meyer, Patrik Nosil and Robert Trivers, as well as with my colleagues in the "Understanding the Brain group": Holk Cruse, Thomas Metzinger, Srini Narayanan and Rafael Núñez.

The research described in this paper involves teams from the University of Brussels (VUB AI Laboratory) and the Sony Computer Science Laboratory in Paris. I am particularly indebted to Joris Bleys for cooperation in the color naming game experiments. I also

thank the entire staff of the Wissenschaftskolleg for their incredible dedication to making this institution so remarkably effective in fostering scientific creativity and reflection.

References

Berlin, D. and P. Kay. 1969. *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Bleys, J. and L. Steels. 2009. "Linguistic selection of language strategies. A case study for colour." *Proceedings of the European Conference on Artificial Life*.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Evans, N. and S. Levinson. 2010. "The myth of language universals: Language diversity and its importance for cognitive science." *Behavioral and Brain Sciences*. In press.

Fodor, J. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Haspelmath, M. 2007. "Pre-established categories don't exist: consequences for language description and typology." *Linguistic Typology* 11, 1: 119–132.

Lillo, J., H. Moreira, I. Vitini, and J. Martin. 2007. "Locating basic Spanish color categories in CIE L*u*v* space: identification, lightness segregation and correspondance with English equivalents." *Psicologica* 28: 21–54.

Loreto, V. and L. Steels. 2007. "Emergence of language." *Nature Physics* 3, 11: 758–760.

McCarthy, J. 1958. "Programs with common sense." *Proceedings of the Teddington Conference on the Mechanization of Thought Processes,* December 1958.

Pickering, M. J. and S. Garrod. 2006. "Alignment as the basis of successful communication." *Research on Language and Computation* 4: 203–228.

Pinker, S. 1994. *The Language Instinct*. New York, Basic Books.

Rosch, E. 1975. "Cognitive representations of semantic categories." *Journal of Experimental Psychology* 104, 3: 192–233.

Salzburger, W., T. Mack, E. Verheyen, and A. Meyer. 2005. "Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes." *BMC Evolutionary Biology* 5, 17.

Salzburger, W., I. Braasch, and A. Meyer. 2007. "Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes." *BMC Biology* 5, 51.

Steels, L. 1997. "The synthetic modeling of language origins." *Evolution of Communication Journal* 1, 1: 1–34.

Steels, L. 2004. "Constructivist development of grounded construction grammars." In *Proceedings Annual Meeting of Association for Computational Linguistics Conference*, edited by D. Scott, W. Daelemans, and M. Walker, 9–16. Barcelona: ACL.

Steels, L. 2008. "The recruitment theory of language origins." In *Emergence of Communication and Language*, edited by C. Lyon, C. Nehaniv, and A. Cangelosi, 129–151. Berlin: Springer.

Steels, L. and T. Belpaeme. 2005. "Coordinating perceptually grounded categories through language. A case study for colour." Target article. *Behaviorial and Brain Sciences* 24, 6.

Steels, L. and J. Bleys. 2009. "Linguistic selection of language strategies. A case study for colour." *European Conference on Artificial Life*.

Steels, L. and J. De Beule. 2007. "Unify and merge in fluid construction grammar." In *Proceedings of EELC III. Lecture Notes in Computer Science*, edited by P. Vogt et al. Berlin: Springer.

Steels, L. 2007. "Is symbolic inheritance similar to genetic inheritance?" *Behavioral and Brain Sciences* 2007.

Steels, L. 2007. "The recruitment theory of language origins." In *Emergence of Communication and Language*, edited by C. Lyon, C. Nehaniv, and A. Cangelosi, 129–151. Berlin: Springer.

Wittgenstein, L. 1953. *Philosophical Investigations*. London: Blackwell Publishing.