Social and Cultural Learning in the Evolution of Human Communication

Luc Steels
SONY Computer Science Laboratory
6 Rue Amyot, Paris
University of Brussels VUB AI Lab
steels@arti.vub.ac.be

Draft Version

2004

1 Evolutionary Linguistics

In order to understand how human languages could have emerged and continue to evolve, we need above all explanations for the enormous increase in complexity compared to animal communication systems. This increase has taken place for all aspects of language:

1. Form: The repertoire of speech sounds used in human language is extraordinarily complex. It relies on an articulatory apparatus which needs to be controlled very fast and at a very fine-grained level. It requires the real-time processing of structured sounds despite noise and individual variation.

2. Meaning: An intricate system of conceptualisation underlies language (Langacker,1987). This system consists of a way to organise the world into different objects and events, a way to categorise them, and a way to introduce structure from the viewpoint of the speaker and listener. For example, the phrase "The car is behind the tree", implies that the speaker and listener view themselves as positioned on a line which goes from themselves to the tree and then to the car. Human conceptualisation is not fixed but expands and adapts to the need of the community. It is partly grounded in the real world through a sensori-motor apparatus and partly abstract based on purely symbolic relations between abstract concepts.

3. Lexicon: The lexicons of human languages are very large (an educated person uses on average 60,000 words and understands 100,000 of them).

Moreover they are open-ended. New words or new meanings are created even as conversations take place (Clark and Brennan,1991). Natural lexicons exhibit homonyms, synonyms and polysemy, which make communication ambiguous and very difficult to learn.

4. Grammar: Human languages use a wide variety of means to indicate the structure and function of words in a phrase (word order, intonation, tone, stress, morphological marking, word form variation, etc.). Using grammar requires a complex planning process from the part of the speaker and a complex plan interpretation process from the part of the hearer.

5. Pragmatics: Human language is used in many different contexts and for many different purposes. Each context induces different registers of language and different dialog patterns. New contexts arise continuously and hence human language conventions need to adapt to remain flexible in achieving pragmatic goals.

There is probably not going to be a single simple mechanism to explain this incredible rise in complexity for all these various aspects, and there is not a single approach to study it.

Generally speaking, three types of questions have been asked, giving rise to three different fields of inquiry:

- Linguists working in the area of historical linguistics have asked the question what changes have taken place in human language. They have amassed a remarkable set of facts documenting language change at all levels and have attempted to organise this in language typologies and laws of change (see e.g. (Traugott and Heine, 1991), or (Vogel and Comrie, 2000).

- Anthropologists and biologists have asked the question why human language might have evolved. Particularly the field of cultural anthropology has tried to identify changes in the ecology or the social organisation of early hominids and relate that to the need for a more complex form of communication (see e.g. (Dunbar 1994)).

- A third type of question is how these developments took place, in other words what are the causal mechanisms both at the level of individuals (their cognitive and bodily apparatus) and at the level of the group (their interactions). This question is more recent and has been asked by a variety of people, ranging from cognitive scientists, psychologists, linguists and A.I. researchers.

Evolutionary linguistics, by analogy with evolutionary biology, is the field of study that attacks these three questions, particularly thethird one. This paper is intended as a contribution to the emerging field of evolutionary linguistics.

The methodology of evolutionary linguistics is reminiscent of that of theoretical evolutionary biology, namely formal modelling and experiments with

artificial systems. Such an approach makes it possible to examine the consequences of certain hypotheses and show with computer simulations or through mathematical analysis whether genes will spread in a population given certain assumptions (Maynard Smith, 1975).

In a similar spirit, I have been doing with my collaborators and students a variety of experiments with robots that engage in various forms of communication with language-like features. The performance of 'artificial linguistic agents' or robots in these experiments falls far short of human capabilities. But realism is not the point. The main goal is a precise and objective exploration of what factors could intervene in the origins and evolution of human communication systems (Steels, 2001).

2 Issues and Hypotheses

2.1 Universals and particulars

There has been an ongoing debate in linguistics between those emphasising the universal character of language (Chomsky, 1975) and those emphasising the uniqueness of each language. The latter group includes historical linguists (Vogel and Comrie,2000) and researchers carrying out empirical studies of language and language use (Labov, 1994).

At the level of sound systems, it has been pointed out that every possible sound that can be made by the human vocal tract is a potential element of a human language repertoire (Ladefoged and Maddieson,1995). We therefore observe a bewildering variety, and speakers of one language are in general not even capable to perceive the subtle sounds of another language or reproduce them, unless they have started at a very early age. The tones in Chinese or the clicks in K!ong are some obvious examples, but so are the v/b sounds in Spanish which are subtly different from those used in English.

Nevertheless there are clear universal tendencies as well. For example, in the case of vowels, it is known that there is a progressive complexity in vowel systems (from 3 to 4, to 5, etc.) but that there is a very specific probability distribution for each size (Schwartz, et.al., 1997). Similar universal tendencies can be found for other aspects of language, such as syllable structure (Venneman, 1992). It is also well established that there is a constant evolution in the sound systems of human languages, sometimes going very rapidly, and that this evolution also exhibits universal tendencies (Labov, 1994).

Many similarities in the conceptualisations underlying different languages have been postulated (Wierzbicka, 1992). For example, the distinction between objects (things, people) on the one hand and events on the other appears universal. Similarly many categorial dimensions like space, time, aspect, countability, kinship relations, etc. are lexicalised in almost all languages of the world, even though there may be profound differences in how this is done. For example, some languages lexicalise kinship relations as nouns (like English: father) and others as verbs (Evans, 2000).

But there are profound differences in the way different languages conceptualise reality (Talmy (2000), Bowerman and Levinson (2001)). And this may impact other cognitive behavior, like memory tests (Davidoff, et.al. 1999). For example, the conceptualisation of the position of the car in "the car is behind the tree" is just the opposite in most African languages. The front of the tree is viewed as being in the same direction as the face of the speaker and hence the car is conceptualised as in front of the tree as opposed to behind the tree (Heine).

The lexicons of languages differ profoundly as well because they use different word forms, even though it has been argued that there is a common core of words which is shared by all languages, pointing to a possible common 'Ursprache' (Ruhlen, 1994). At the same time there are also profound differences in terms of which aspectsof reality are lexicalised. For example, in Japanese the word "-san" (mister or miss/misses) is neutral with respect to male or female, whereas English forces us to make a gender distinction and for females a distinction based on marital status. The lexicon of a language is clearly in constant rapid evolution.

Finally, many linguists in the Chomskyan tradition of generative grammar have argued that all languages are variations on the same basic pattern, that of universal grammar. They have tried to capture the universality as a set of principles and the variation as a set of parameters that have to be set in particular languages (Chomsky and Lasnik, 1993). Based on empirical evidence, others have argued that natural languages are in fact profoundly different. For example, nothing seems more universal and basic than the parts of speech (noun, verb, adjective, adverb, preposition, etc.). Nevertheless there is irrefutable evidence that many languages do not share the parts of speech that seem so obvious for English. For example Mundari, an Austro-Asiatic language does not make a distinction between nouns and verbs (Bhat, 2000, p. 56). Any lexicalised predicate can be used as a verb, in the sense that it can be used as predication and takes tense and aspect markers, agreement, voice, etc., and as a noun, in which case it takes case markers and is used referentially. Words therefore denote both things and events. For example 'lutur' means both ear and listen and 'kumRu' a thief and to steal. Some languages do not have prepositions (but use double verb constructions instead), others, like Boro, a Tibeto-Burman language, do not have adverbs but use morphological markers instead (Bhat, o.c. p. 59). Thus masa means 'to dance' and 'masaglo' 'to dance quickly'.

There is a constant significant evolution in the grammars of human languages. Thus a grammatical category similar to prepositions may evolve from double verb constructions or the category auxiliary may develop by differentiating the class of verbs (Traugott and Heine,1991).

All this points to a first major puzzle: How can we explain that there are both universal tendencies in human languages and at the same time such a bewildering variety? How can we explain that there is an ongoing profound change in human language at all levels?

2.2 Genetic versus Culture

The tension between universals and particulars is related to another debate: How is the human language system transmitted. Is it primarily in a genetic fashion, i.e. through the human genome? Or is it primarily in a cultural fashion, i.e. through learning?

The genetic approach to language evolution necessarily favours universals but has difficulty to explain strong variation and rapid evolution. The cultural approach favours particulars and has no difficulty to explain rapid evolution but must still address why there are universal tendencies and how language is transmitted.

Intermediary positions are possible. In particular, some researchers have argued for a strong Baldwin effect (Briscoe, 2000), which would imply however that there is a cultural evolution first and that the resulting linguistic behaviors are sufficiently stable to be genetically encoded later.

The genetic position is associated with researchers like Pinker (Pinker,1994) who have postulated a genetically encoded language acquisition device which embodies the constraints of universal grammar. Language acquisition in this framework is not really a learning process but a maturation process. Data present in the linguistic environment acts as a way to set the parameters of universal grammar (Lightfoot, 1991). The genetic encoding of language requires that language evolves in a genetic fashion, i.e. that important changes in human language communication must have arisen from mutations and the propagation of these mutations in the human gene pool (Bickerton and Calvin, 2000).

The cultural position relies on learning as the way in which language gets transmitted (Tomasello, 1999). It has been criticised by nativists because they point to a poverty of stimuli and a lack of direct linguistic feedback given to children. But those emphasising learning argue that there is in fact quite a lot of pragmatic feedback and they have been trying to find alternative learning methods essentially pursuing two approaches: individualistic versus cultural learning.

In the case of individualistic learning, the child is assumed to receive as input a large number of example cases where speech is paired with specific situations. She is assumed to extract through an inductive learning process what is essential and recurrent of these situations, in other words learn the appropriate categories underlying language, and then associate these categories with words. This viewpoint assumes a rather passive role of the language learner and no feedback given by the speaker (see (Fischer, et.al., 1994), (Clark, 1987). It is widespread among researchers studying the acquisition of communication and various attempts have been made to model it with neural networks or symbolic learning algorithms (Broeder and Murre, 2000). Induction by itself is a weak learning method, which does not give identical results on the same data and may yield irrelevant clustering compared to human clustering. To counter this argument it is usually proposed that innate constraints help the learner zoom in on the important aspects of the environment (Smith, 2001).

In the case of social learning, interaction with other human beings is con-

sidered crucial. The mediator could be a parent and the learner a child, but children (or adults) can and do teach each other just as well. Given the crucial role of the mediator. The goal of the interaction is not really teaching but something practical in the world, for example, to identify an object or an action. The mediator helps to achieve the goal and is often the one who wants to see the goal achieved.

The mediator has various roles: She sets constraints on the situation to make it more manageable (scaffolding), gives encouragement on the way, provides feedback, and acts upon the consequences of the learner's actions. The feedback is not directly about language and certainly not about the concepts underlying language. The latter are never visible. The learner cannot inspect telepathically the internal states of the speaker and the mediator cannot know which concepts are already known by the learner. Instead feedback is pragmatic, that means that it operates in terms of whether the goal has been realised or not. Consider a situation where the mediator says: "Give me that pen", and the learner picks up a piece of paper instead of the pen. The mediator might say: "No, not the paper, the pen", and point to the pen. This is an example of pragmatic feedback. It is not only relevant to succeed subsequently in the task but supplies the learner with information relevant for acquiring new knowledge. The learner can grasp the referent from the context and situation, hypothesise a classification of the referent, and store an association between the classification and the word for future use. While doing all this, the learner actively tries to guess the intentions of the mediator. The intentions are of two sorts. The learner must guess what the goal is that the mediator wants to see realised (like 'pick up the pen on the table') and the learner must guess the way that the mediator has construed the world. Typically the learner uses herself as a model of how the mediator would make a decision and adapts this model when a discrepancy arises.

Social learning enables active learning. The learner can initiate a kind of experiment to test knowledge that is uncertain or to fill in missing wholes. The mediator is available to give direct concrete feedback for the specific experiment done by the learner. This obviously speeds up the learning, compared to a passive learning situation where the learner simply has to wait until examples arise that would push the learning forward.

2.3 Coherence

The members of a language community must all share (at least to a large extent) the same language conventions and the same conceptualisations, otherwise communication is not really possible. A key puzzle of evolutionary linguistics is where this coherence may come from and different explanations have been put forward along the lines of the universal/genetic versus cultural/learning debate.

The genetic framework explains language coherence through gene sharing. Genes of successful organisms, which presumably means in the case of language human beings which are better in producing and comprehending the language in the community, propagate in a population until there is complete homogeneity. It must be noted that this process is very slow (10,000 years in the case of a

human gene for lactose which could be studied in this respect) and that there is never complete homogeneity because of natural variation in the gene pool. But perhaps the language genes postulated by Pinker, et.al. are like the genes for determining the number of fingers of the hand. They are so widespread that differences are not noticed.

The genetic framework explains conceptual coherence by assuming that the concepts needed for language communication evolve also through mutation and subsequent propagation in the population. Once again, there is a difficulty with the speed of genetic evolution which is too slow to explain the rapid rise and spreading of new concepts in human populations. A good example are the many concepts associated with Internet (browsing, home page, server, internet provider, etc.) which are now common knowledge but did not exist even fifteen years ago. It seems difficult to maintain that these were all innate. Another issue concerns the question of storage. Worden (1998) has argued that there is simply not the available genetic storage space for the massive amount of concepts and linguistic specifications that are usually assumed to be innate, particularly when the human genome is compared to its closest species, which do not have language.

On the other hand, if different language users learn the language and concepts underlying language through social and cultural learning how can this become shared? Several researchers (see review in (Steels, 1997)) have shown through a number of computer simulations that two principles may explain this, both coming from the study of biological systems: self-organisation and structural coupling. Self-organisation is clearly seen in path formation in ant societies and many other pattern formation and collective phenomena (Camazine, et.al., 2001). In the case of ants, there is random variation (all ants crawling around randomly) and a positive feedback loop that influences variation. Concretely when an ant finds a path, it leaves behind a chemical trail (pheromone) which attracts other ants. When they also come back on the same path the trail becomes stronger and hence more ants get attracted. The end result is the complete self-organisation of all ants on a path without a central coordinator nor prior knowledge.

The principle of self-organisation can be applied to explain how certain conventions spread in a population. If speakers use the conventions that were most successful in the past as a guide, we get a positive feedback loop. The more conventions are used the more successful they are and so they get used even more. This leads to a winner-take-all effect in which one word dominates for the expression of a particular meaning (see figure 1) (Steels, 1996). Structural coupling, a concept first introduced by Maturana and Varela (1998), is needed to explain how concepts may become shared without direct feedback. Structural coupling means that two systems develop independently but due to the fact that they both provide inputs and feedback to each other, they become tightly coordinated. In this case, the learning system generating new meanings and the learning system lexicalising these meanings receive feedback from each other. New concepts generate new words and the successful use of a word gives a boost to the concept that was used with this word. It has been shown in com-

Figure 1: Results from an experiment in the emergence of a lexicon in a population of agents. The graph shows all the words for a given meaning and their percentage of use. A winner-take-all situation is clearly observed after a struggle in which several words compete for the same meaning.

puter simulations (Steels, 1997) and in experiments with robots (Steels, et.al. 2002) that this indeed gives rise to coherence not only for words but also for meanings.

2.4 Complexity

A final issue concerns the increase in complexity. Also here we get radically different explanations depending on the position with respect to the genetic vs. cultural debate. In a genetic framework, the complexity of language can only fundamentally increase due to genetic mutations or the reshuffling of parameter settings (Lightfoot, 1998). The introduction of a new part of speech would presumably require a change in the language genes.

In a cultural framework, language users are viewed as active agents that shape and reshape their language. Complexity increases due to the growing needs of the language community which pushes expressivity forward. These needs are met by the extension of the lexicon which may lead to increases in the complexity of the sound system, so that words can still be well distinguished, or to the expansion of grammatical structures to be able to say more with less. Many of the expansions are based on problem solving and make use of generic cognitive abilities. For example, analogy has been pointed out to play an important role in the recruitment of existing words or constructions for new functions (Heine). Expressivity generally comes in conflict with learnability. And so there are also forces to simplify wordforms or simplify grammatical constructions. The messy features of a specific language are explained by the need to optimise the conflict between expressivity and economy.

3 Evolutionary Game Theory

It is possible to use the methodology of computer simulations and robotic experiments to examine any of the hypotheses mentioned here. For example, if one argues in favor of the principles and parameters theory one can put forward a detailed universal grammar and show on a database of example sentences empirically collected from interactions between adults and children how precisely the parameters get set (see examples in Berwick (1998) and Briscoe (2001)). If one believes that 'the learning bottleneck' (i.e. the fact that each generation has to learn the language of its predecessors) explains why languages have become compositional, one can make a precise model that gives 'linguistic agents' a choice between compositional and holistic expression and see which choice is made if the language needs to be transmitted culturally (Kirby, 1999).

In our own work we favor the cultural approach to language. We view language as a complex adaptive system that has evolved in a cultural fashion under the natural constraints given by the human sensori-motor apparatus, the cognitive apparatus, and the environments and ecologies in which humans find themselves (Steels, 2000). Moreover we support the hypothesis that the learning process, both for conceptualisation and for language itself, is highly social. Our experiments attempt to substantiate these hypotheses, partly by showing that the cultural self-organisation of language is possible given social learning and partly by showing that other learning methods or innate schemas are too slow to adapt to changes or impossible due to the lack of quality data or the inherent combinatorial explosion hidden in language learning. The rest of the paper provides a bit more detail on two example experiments: one focusing on the emergence of sound repertoires and one looking at the emergence of words and word meaning.

3.1 Imitation games for sound repertoires

Kaneko and Suzuki (1994) showed how the framework of evolutionary game theory could be used to study the evolution of sound repertoires in birds. Our group has started to use the same framework for phonetics from around 1995, and applied it to different aspects of soundsystems: vowels (de Boer, 1997) and syllables (Steels and Oudeyer, 2000). All this builds further on earlier work by phoneticians to show that the sound systems of natural languages are not arbitrary but the consequence of various sensori-motor and cognitive constraints (Lindblom, et.al. 1984). The rest of this section describes the work of de Boer (1997) as an example how a repertoire of sounds may become agreed upon in a distributed group of robots. The question being addressed in these experiments is how robots can come to share a system of vowels without having been given a pre-programmed set nor with central supervision, and how the universal tendencies found in human vowel systems can be explained through a process of cultural evolution under natural selection.

In the robotic simulations, the sensori-motor apparatus of the robots consists of an acoustic analyser on the one hand, which extract the first formants from the signal, and an articulatory synthesiser on the other hand which models relevant aspects of the human vocal tract. The robots play an imitation game. One robot produces a random sound from its repertoire. The other robot (the imitator) recognises it in terms of its own repertoire and then reproduces the sound. Then the first robot attempts to recognise the sound of the imitator again and if it is similar to its own, the game is a success otherwise a failure. This setup therefore adopts the motor theory of perception whereby recognition of a sound amounts to the retrieval of a motor program that can reproduce it.

To achieve this task, the robots in the De Boer experiment use two cognitive structures: The vowels are mapped as points into a space formed by the first, second and third formants (see figure 2) and a nearest-neighbor algorithm is used to identify an incoming sound with the sounds already stored as prototypes. These prototypes have an associated motor program that can be used

Figure 2: Example of the evolution of a vowel system. Vowels are represented in formant space (first and second formant). Each dot represents the vowel prototype for one agent. Prototypes progressively cluster into groups and occasionally a new cluster appears.

to reproduce the sound. When an imitation game succeeds, the score of the prototype goes up, which means that the certainty that it is in the repertoire increases. There are two types of failure. Either the incoming sound is nowhere near any of the sounds already in the repertoire. In that case it is added to the prototype space and the robot tries to find its corresponding motor program by producing and listening to itself, progressively adjusting the motor program until it produces the desired sound.

Alternatively, the incoming sound is near an existing sound but the reproduction is rejected by the producing robot. This means that the imitator does not make sufficiently fine-grained distinctions. Consequently the failure can be repaired by adding this new incoming sound as a new prototype to the repertoire and associating it with a motor program learned again by hill-climbing. In order to get new sounds into the repertoire, robots occasionally "invent" a new sound by a random choice of the articulatory parameters and store its acoustic image in the prototype space. Sounds which have consistently low scores are thrown out and two sounds that are very close together in the prototype space are merged.

Quite remarkably, the following phenomena are perceived when a consecutive series of games is played by a population of robots: (1) A repertoire of shared sounds emerges through self-organisation (see figure 2). (2) The repertoire keeps expanding as long as there is pressure to do so. (3) Most interestingly, the kinds of vowel systems that emerge have the same characteristics as those of natural vowel systems (De Boer,2001). The experiment therefore not only shows that the problem can be solved in a distributed fashion but also that it captures some essential properties of natural systems.

Three principles have been used: Reinforcement learning based on feedback after each game (Sutton and Barto, 1998) explains how individual agents may learn the vowels that are present in their environment. Self-organisation (in the sense of Nicolis and Prigogine (1998) explains how a group individuals arrives at a shared repertoire. It arises when there is a positive feedback loop in an open non-linear system. Here there is a positive feedback between use and success. Sounds that are (culturally) successful propagate. The more a sound is used the more success it has and so it will be used even more. Self-organisation explains that the group reaches coherence, but not why these specific vowels occur and not others. For this we need a third principle, namely natural selection familiar from Darwinian explanations of biological complexity. The scores of vowels that can be successfully distinguished and reproduced given a specific sensorimotor apparatus have a tendency to increase and they hence survive in the population. Novel sounds or deviations of existing sounds (which automatically

get produced due to the unavoidable stochasticity) create variation, and sensori-motor constraints select those that can be re-produced and recognised. The closer we can model human natural sensori-motor behavior and environmental conditions the more realistic the vowel systems become.

3.2 Guessing games for concepts and words

The notion of a language game was promoted by Wittgenstein to emphasise that language and meaning are not based on context-independent abstractions but arise as part of concrete interactive situations. We have found language games to be an excellent vehicle for studying the emergence of the words and meanings in robotic experiments within the framework of social learning (Steels, 2001).

A language game is a situated interaction between two agents. The interaction not only involves language aspects, i.e. the parsing and producing of utterances, but also the grounding through sensory processing, execution of appropriate gestures or other actions in the world, and, most importantly, steps for learning new parts of language if necessary: new words, new meanings for existing words, new phrases, new pronunciations of known words.

Complex dialogs involve multiple language games interlaced with each other. The meaning of a word or phrase comes from its role in a language game, just like the meaning of 'queen' in chess. This explains why the meaning of a word cannot be defined easily in absolute terms but arises from the situation and context. It explains why humans have no trouble to disambiguate words or phrases: The interpretation processes take place in the context of a situated language game which strongly restricts what is being talked about.

We have used this framework of language games to study the origins of meanings and lexicons in groups of agents. A first example game that we have study intensely is the Guessing Game (Steels, 1998). In the Guessing Game the speaker tries to draw attention of the hearer to an object in the environment. For example, Mary sits at the table and asks her neighbour Pierre for the salt by saying "salt". She perhaps points at the same time to the salt. The table, all the objects on it, the people around the table and their actions, form the context of the game. The salt is called the topic. We notice immediately that the word spoken, namely "salt", is only a small part of what is going on. The hearer must also perceive and conceptualise the situation, interpret the gestures made by the speaker, guess what action the speaker may want, etc. All of these are an intrinsic part of the language game.

There are many ways the Guessing Game can fail. Pierre may have incorrectly understood the word, or simply not know the word, or believe that the word has another meaning. This failure becomes noticed by subsequent action (for example Pierre hands Mary the water instead of the salt). Every language game must contain provisions for detecting failure and repairing it. The speaker typically provides more information, possibly in a non-verbal way through additional gestures. If failure is due to lack of knowledge, the language game is an opportunity for learning. For example, when the hearer does not know the word "salt" (perhaps Pierre is French), he can use this example to acquire a

Figure 3: Set up for the Talking Heads experiment with two pan-tilt cameras looking at a white board on which colored geometric figures are pasted. The agents using these robotic bodies play a Guessing Game.

new word. If he failed to conceptualise the scene (perhaps Pierre comes from a culture where salt never is purified to white grains), he may acquire or enrich his repertoire of concepts.

Many variations can easily be imagined. But in any case, it is crucial that (1) speaker and hearer keep scores how well the associations between form and meaning were doing in the game to enable a well founded choice of the one that has potentially the most success, (2) there is a positive feedback loop between use and success in the sense that associations with a higher score increase their chance of being used and thus increase their potential for leading to a successful communication in the future, and (3) there is a strong structural coupling between concept formation and language, achieved when language gives feedback on the adequacy of concepts.

We have formalised and implemented all aspects of the guessing game on simple robots equiped with pan-tilt cameras (figure 3). The context consists of a small area on the white board covered with geometric figures and the topic is one figure, for example a red square. A decision-tree like algorithm was used for conceptualisation. Input to the decision-tree is output from a battery of statistical pattern recognition and computer vision algorithms. Thus, left meant that the x coordinate of the middle point of a figure was less than the average x coordinates of all figures in the context and right meant that it was greater than this average, large meant that the size of the figure was greater than the average size of all figures, etc. A selectionist learning method was used for concept acquisition: Decision-trees grow in a random fashion when there is a failure to find a distinctive concep that distinguishes the topic from the other objects in the concept, and branches get pruned when they are irrelevant or unsuccessful in subsequent language games Steels (1997b).

To play a game, the robots capture an image, segment it, derive features using statistical pattern recognition techniques, and give pragmatic feedback by gesturing towards objects with their cameras. Utterances are single words and the lexicon consists of associations between single words and visually grounded predicates. Each assocation is stored as a triple ¡r,s,k¿ where r is a visually grounded feature or combination of features, s a symbol, and k a score to reflect how successful this association has been in the past games, and hence how successful it might be in the future. Each individual robot has its own lexicon. There is no global knowledge nor central control.

A somewhat simplified version of the game involves the following steps (see figure 4):

1. Shared attention. By pointing, eye gazing, moving an object, or other means, the speaker draws the visual attention of the hearer to the topic or at least a narrow context which includes the topic. The speaker may

emit a word aiding to share attention, like "look", and observes whether the hearer gazes towards the topic. Based on this activity, both agents can be assumed to each have captured an image that reflects the shared context.

2. Speaker behavior. The speaker then conceptualises the topic yielding a representation r. Conceptualisation means that a combination of concepts is found that distinguishes the topic from the other objects in the context. Let us assume for the simple version of the game, that this is a single predicate which is true for the topic but not for the other objects. For example, if every object on the table is blue but the topic is white, then color is a good way to refer to the topic. The speaker then collects all associations <r,s,k> in his lexicon and picks out the one with highest score k. The s from this association is the best word to communicate from the speaker's point of view. It is transformed into a speech signal and transmitted to the hearer.

3. Hearer behavior. The hearer receives the speech signal, recognises the word s, and looks up all associations ¡r',s,m¿ in his memory.

   (a) If the hearer did not have an association in memory for s, this means that s is a new word for the hearer. The hearer signals incomprehension and the speaker points to the topic. So the hearer can perform his own conceptualisation of the scene, finding a categories indicating how the topic is different from the other objects, picks one possibility (if there is more than one) yielding a representation r", and stores the new association ¡r",s,i¿ where i is an initial default score.

   (b) If there are associations, the hearer applies each representation r' to the current scene (perhaps starting with those with the highest score m) to see whether any one picks out a unique object. If that is the case, this is the topic. There may be ambiguity, i.e. more than one possible topic, in which case the referent picked out by the association with the highest score is chosen. The hearer then points to the topic.

4. Feedback. Suppose that the hearer found a referent (step 3.2.), then there are two outcomes:

   (a) The speaker agrees that this is the right referent, i.e. it was the topic she originally had in mind and signals agreement. In that case both speaker and hearer increase the score of the association they used and decrease the score of competing associations. For the speaker a competing association is one which involves the same meaning but a different word. For the hearer a competing association is one which involves the same word but a different meaning.

   (b) If the speaker signals that the hearer failed to recognise the topic, then the score of the used associations is decreased by both speaker

Figure 4: Left: processes carried out by the speaker. Right: processes carried out by the hearer. There are also feedback processes moving in alternate directions until the agents settle on coherent choices for all the stages.

and hearer. The speaker gives additional feedback until speaker and hearer share the same topic. The hearer can then conceptualise the topic from his point of view and either store a new word (as in 3.2.) or increase the score of an existing association (as in 4.1.).

5. Speaker or hearer may fail to conceptualise the scene, in which case a concept acquisition algorithm is triggered, which should try to acquire a new conceptualisation using the current situation as a source for learning

Here is a typical dialog with the images and concept repertoires shown in figure 4. Both agents compute a number of features for each figure in the image: X the horizontal position of the middle point of the figure, Y the vertical position, H the height, W the width, A the angularity (the number of angles), R, G, Y, B the amount of red, green, yellow, and blue in the figure, and L the brightness:

Object 0 (0,1) X: 0.37, Y: 0.71, H: 0.48, W: 0.21, A: 0.45, R: 0.17, G: 0.0, Y: 0.0, B: 0.39 , L: 0.28

Object 1(1,0.96) X: 0.7, Y: 0.69, H: 0.38, W: 0.22, A: 0.45, R: 0.98 G: 0.0, Y: 0.52, B: 0.0, L: 0.36

Object 2 (0.42,0.0) X: 0.51, Y: 0.31, H: 0.21, W: 0.51, A: 0.70, R: 0.00, G: 0.99, Y: 0.73, B: 0.0, L: 0.46

The first object (with coordinates 0,1) is the topic. Based on the decision trees (shown in figure 3), the agent conceptualises this object in terms of distinctions on the blue channel. A shade of blue (between 0.25-0.5) is distinctive for the topic but not for any other object in the context. The speaker has three words in the lexicon for this: XAGADUDE (score 0.1), NIBIDESU (score 0.0) and TETIPI (score 0.0). The first word is chosen. so the speaker says:

Speaker: XAGADUDE

The hearer does not know this word and therefore signals incomprehension.

Listener: Huh?

Figure 5: Example of a guessing game played by two robots. Left the images captured by the speaker (top) and the hearer (bottom). Notice that they are not exactly the same. On the right hand side the decision trees are shown of the speaker (top) and the hearer (bottom). The repertoires are not the same, although the same distinction happens to be used by both agents in the game discussed in the text.

The speaker now points to the topic using the camera. The hearer then performs its own categorisation using its own decision trees, which happens to yield the same conceptualisation. The hearer then adds a new association in the lexicon. Note that it could very well have happened that the hearer used another conceptualisation for this scene, for example based on brightness or height, in which case there is a divergence in the lexicon. Such a divergence would show up when in a later game the same agents are confronted with a disambiguating situation.

We have done experiments with a growing population of up to 3000 robots with multiple installations and sharing of bodies by more than one robot. The robots played in total 500,000 language games over a period of 3 months. These experiments show that (1) a lexicon indeed arises in the population from scratch. There was a core of a few hundred words and a total lexicon of 8000 words. (2) The lexicon gets transmitted as new generations come in. (3) There is a high degree of communicative success which is due to sufficient sharing of words for specific meanings (as shown in figure 1) and sufficient sharing of meanings. The three principles discussed earlier are at work: reinforcement learning, self-organisation and structural coupling. A broader discussion of this experiment and an analysis of the causal factors explaining its success can be found in Steels, et.al. (2002).

3.3 Classification games on autonomous robots

More recently several experiments on the AIBO dog-like robot were performed within the same language game framework (Steels and Kaplan, 2001). In many respects these experiments are a giant step beyond the previous one. First of all the AIBO is a fully autonomous mobile robot with more than a thousand behaviors, coordinated through a complex behavior-based motivational system (Fujita, 1998). It follows that getting attention from the robot and sharing the same perspective on the world prior to a game is complex. The dialog can be enhanced by gestures and movements by the robot and on-board visual processing and sensing can be used. The experiments used speech input and output, using off-the-shelf speech components. Obviously the use of spoken language increases still further the uncertainty of the communication. Second, the dialog took place between a human and a robot. This introduced many additional complexities but gave us the opportunity to design and test very concrete models of social learning.

Nevertheless several language games were successfully implemented, starting

Figure 6: A Guessing Game between the AIBO and a human experimenter Frederic Kaplan, involving the use and acquisition of a word for ball.

with the guessing game. Rather than using the decision trees discussed earlier, conceptualisation was performed using a nearest neighbor algorithm with a memory of stored object views (similar to the one used in Bartlett (1997)). The object memory is acquired using instance-based learning. Every language game is an opportunity to acquire a new view of an object or to learn that there is a new class of objects of which the current example is a first instance. This experiment therefore showed that any kind of concept acquisition can be used. The topic can not only be a single object but also an action or a property of a situation. Other games focused on naming body parts and actions to be used in commands.

An example dialog is the following, starting when the experimenter shows a red ball (figure 6).

1. Human: Sit. 2. Human: Sit down.

AIBO has already acquired names of actions. Forcing the robot to sit down is a way to make it concentrate on the language game. The human now shows the ball to the robot.

3. Human: Look 4. Human: ball

The word "look" is helping to cause a focus on the guessing game based on visual input. The robot performs image capturing and segmentation. The game will be possible if a segment has been found. The robot tries to recognise the object using a nearest neighbor algorithm.

5. Aibo: Ball?

The robot asks for feedback of the word to make sure that the word has been understood. Ball is the word that will then be associated with the object.

6. Human: Yes

There is positive feedback on the word pronounced. This feedback causes the rest of the guessing game to proceed, i.e. storage of new word if not yet known, increase of the score, etc.

We have shown that the visual data generated in this experiment is so confusing that an unsupervised learning algorithm cannot find the concepts that are required to use the (English) words correctly, thus suggesting that there must be a strong causal influence of language on concept formation (Steels and

Kaplan, 2001). We also see evidence in the same experiment why the learning of language has to be social. The mediator must restrict the attention of the learner to help focus on what is important, on what needs to be learned. The mediator must help to ensure that the learner has good and relevant data. When the role of the mediator is reduced we get less adequate results both for concept acquisition and for word learning (Steels and Kaplan, 2001).

4 conclusions

Evolutionary linguistics is concerned with explaining the causal factors underlying the evolution of human-like communication. One way to test theories is to enage in computer simulations and robotic experiments. The paper discussed various issues in evolutionary linguistics, specifically how coherence and complexity may arise and how language is transmitted from one generation to the next. We have briefly introduced a few experiments that show the importance of social and cultural learning as well as concrete models to test these hypotheses. Self-organisation, structural coupling and reinforcement learning were shown to be among the primary forces. There are obviously many open problems, particularly concerning the origins of grammar, but there is no doubt that a new promising avenue has been opened up to investigate a fascinating enigma of human evolution: the origins and evolution of language.

5 Acknowledgements

I am indebted to the members of the Sony Computer Science Laboratory in Paris, particularly Frederic Kaplan, Angus McIntyre, Eduardo Miranda, Pierre-Yves Oudeyer and of the VUB Artificial Intelligence laboratory in Brussels, in particular Paul Vogt, Joris van Looveren, Bart De Boer, Tony Belpaeme, Edwin de Jong.

6 References

Bartlett M (1997) SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally-Inspired Approach to Visual Object Recognition Neural Computation, 1997, 9, 777-804

Berwick R (1998) Language evolution and the minimalist program: the origins of syntax. In: Hurford J , Studdert-Kennedy M and Knight C (ed.) (1998) Approaches to the Evolution of Language. Cambridge University Press, Cambridge. pp. 320–340.

Bhat D (2000) Word classes and sentential functions. p. 47-63. In: Vogel M and Comrie B (eds.) (2000) Approaches to the Typology of Word Classes. Mouton de Gruyter, Berlin.

Bickerton D and W Calvin (2000) Lingua ex Machina: Reconciling Darwin and Chomsky With the Human Brain, MIT Press, Ca.

Bowerman M and Levinson S (2001) Language acquisition and conceptual development. Cambridge Univ Press, Cambridge.

Broeder P and Murre J (2000) Models of Language Acquisition. Inductive and Deductive Approaches. Oxford University Press, Oxford.

Briscoe E (2000) Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device", Language 76.2, 2000.

Briscoe E (2001) Grammatical Acquisition and Linguistic Selection. In: Briscoe E (2001) (ed.) Linguistic evolution through language acquisition: formal and computational models. Cambridge University Press, Cambridge.

Camazine S Deneubourg J-L Franks N Sneyd J Theraulaz G and Bonabeau E (2001) Self-Organization in Biological Systems. Princeton University Press, Princeton.

Chomsky N (1975) Reflections on Language. Pantheon books, New York.

Chomsky N and H Lasnik (1993) The Theory of Principles and Parameters. In: Jacbos J von Stechow A, Sternefeld W and Vennemann T (eds) Syntax: An International Handbook of Contemporary Research. Walter de Gruyter, Berlin. p. 506-569.

Clark E (1987) The Principle of Contrast: A constraint on language acquisition. In: MacWhinney B. (ed.) Mechanisms of Language Acquisition. L. Erlbaum Hillsdale NJ.

Clark H. and Brennan S (1991) Grounding in communication. In: Resnick, L Levine J and Teasley S (eds.) Perspectives on Socially Shared Cognition. APA Books, Washington. p. 127-149.

Davidoff J Davies I, Roberson J (1999) Color categories in a stone-age tribe. Nature, vol 398. 230-231.

Dunbar R (1994) Grooming, Gossip and the Evolution of Language. Harvard University Press, Cambridge Ma.

Evans N (2000) Kinship verbs. p 103-172. in: In: Vogel P and Comrie B (eds.) (2000) Approaches to the Typology of Word Classes. Mouton de Gruyter, Berlin.

Fischer C Hall G Rakowitz S and Gleitman L (1994) When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. Lingua (92) 333-375.

Hurford J Knight C and Studdert-Kennedy M (eds.) (1998) Approaches to the Evolution of Language: Social and Cognitive bases. Cambridge University Press, Cambridge.

Kaneko K Suzuki J (1994), Imitation Games, Physica D 75, Elsevier Science.

Kirby S (1999) Function, Selection and Innateness: The Emergence of Language Universals. Oxford University Press, Oxford.

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

Ladefoged P and Maddieson I (1995) The sounds of the world's languages. University of Chicago Press, Chicago.

Langacker R (1987). Foundations of cognitive grammar, vol.1. Stanford: Stanford University Press.

Lightfoot D (1991). How to set parameters. MIT Press, Cambridge Ma.

Lightfoot D (1998). The Development of Language: Acquisition, Change, and Evolution. Blackwell, Oxford.

Maturana H and Varela F (1998) The Tree of Knowledge (revised edition). Shambhala Press, Boston.

Maynard Smith J (1975) Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Lindblom B MacNeilage P and Studdert-Kennedy M (1984) Self-organizing processes and the explanation of language universals. In Brian Butterworth B Comrie and Dahl O (eds.) Explanations for language universals. Walter de Gruyter and Co. pp. 181- 203.

Pinker S (1994) The Language Instinct. The New Science of Language and Mind. Penguin, Harmondsworth.

Ruhlen M (1994) On the Origin of Languages: Studies in Linguistic Taxonomy. Stanford University Press, Stanford Ca.

Schwartz J-L Boe L-J Vallee N and Abry C (1997) Major trends in vowel system inventories. Journal of Phonetics 25, pp. 233-253.

Smith L. (2001) How Domain-General Processes may create Domain-Specific Biases. In: Bowerman M and Levinson SC (2001) Language acquisition and conceptual development. Cambridge Univ Press, Cambridge. p. 101-131

Steels L (1996) Self-Organizing Vocabularies. In: Langton C. and Shimohara T (ed) (1997) Proceedings of the Artificial Life V. The MIT Press, Cambridge Ma. pp. 179-184.

Steels L (1997a) The Synthetic Modeling of Language Origins. Evolution of Communication Journal 1(1), 1-35. (1997)

Steels L (1997b) Constructing and Sharing Perceptual Distinctions. In: van Someren M and Widmer G (ed.) (1997) Proceedings of the European Conference on Machine Learning, Springer-Verlag, Berlin. pp. 4-13.

Steels L (1998) The origins of syntax in visually grounded robotic agents. Artificial Intelligence 103 (1,2) p. 133-156.

Steels L (2000) Language as a Complex Adaptive System. In: Schoenauer, M., editor, Proceedings of PPSN VI, Lecture Notes in Computer Science, Berlin, Germany, September 2000. Springer-Verlag. pp. 17-26.

Steels L (2001) Social learning and language acquisition. To appear in: McFarland D and Holland O (eds) Social Robots. Oxford University Press, Oxford.

Steels, L (2001) The Methodology of the Artificial Behavioral and Brain Sciences 2001 24(6). Commentary on Barbara Webb's article.

Steels L (2001) Language Games for Autonomous Robots. IEEE Intelligent systems, September/October 2001, p. 16-22.

Steels L and F Kaplan (2001) AIBO's first words. The social learning of language and meaning. Evolution of Communication 4(1).

Steels L and F Kaplan and A McIntyre and J Van Looveren (2002) Crucial factors in the origins of word-meaning. In: Wray A et.al. (2002) The Transition to Language. Oxford University Press. Oxford, UK, in press.

Steels L and Oudeyer P-Y (2000) The cultural evolution of syntactic constraints in phonology. In: Bedeau, et.al. (ed.) (2000) Proceedings of Artificial

Life VII. The MIT Press, Cambridge Ma. pp. 382-394.

Sutton R and Barto A (1998) Reinforcement Learning. The MIT Press, Cambridge Ma.

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

Tomasello M (1999). The Cultural Origins of Human Cognition. Harvard University Press.

Traugott E and Heine B (1991) Approaches to Grammaticalization. Volume I and II. John Benjamins Publishing Company, Amsterdam, 1991.

Vennemann T (1988) Preference Laws for Syllable Structure. Mouton de Gruyter, Berlin.

Wierzbicka A (1992) Semantics, Culture and Cognition. Oxford University Press, Oxford.