

A brain for language

Luc Steels

Sony Computer Science Laboratory - Paris

steels@csl.sony.fr

1 Introduction

Natural languages constitute one of the most complex phenomena arising in living systems. Because languages are so rich and complex, they can be studied from many different viewpoints:

- Linguists, who study language as a system as such, have hence been extremely interested in the question: Why is language the way it is? What kind of universal tendencies do we see and why are they there? (Greenberg, 1966).
- Psycholinguists have been focusing on the question what architecture and processes need to be available to speakers to participate in language understanding and what kinds of mechanisms can explain that language can be learned. (Levelt, 1989)
- The complex processes required for language are somehow carried out by the human brain. This raises the question, taken up by neurobiology, where in the brain language processing takes place, and what kind of neural circuitry is involved. Mirroring the psycholinguistic question of language learning, neurobiology also asks the question how the neural networks involved in language develop or adapt. (Friederici, 1999).
- The rise of computers and the need to interact with them in a natural language like form has generated considerable interest recently in language technology. Most of this technology focuses on some kind of processing without taking meaning into account. But there is a longstanding interest,

particularly in AI, in the question whether and how machines can be built that are capable of natural language understanding. This will only be possible if we go beyond statistical processing of language. We need compositional analysis at a fine level of detail, a procedural semantics establishing a relationship between sentences and the world, and above all a productive system to formulate new sentences or interpret sentences never heard before.

Although practitioners of these various fields have often believed in the past that their subject areas are totally unrelated, there is fortunately a strong trend in contemporary cognitive science to see possible exchanges across disciplines. My own interest lies strongly in the question how language understanding machines can be built, although I believe at the same time that this can only be done based on deep insights into the linguistic, psychological and neurobiological processes in human language understanding. Conversely I believe that we can only develop deep models of language, its processing, and its brain implementation only through the experimental construction of artefacts which approach closer and closer to human language capability. This effort is far from finished.

2 Two approaches to grammar

Let me start from a linguist's point of view: There are clearly universal tendencies in languages. For example, if a language shows a distinction between subject, object(s) and verb, then we often see that the subject comes first followed by the verb and then the object(s), as in English: "John likes novels", or French "Jean aime les romans". However not all languages follow this SVO pattern, for example, Japanese uses SOV. The verb comes at the end, as in "Sumisusan wa Nihon e ikimashita" (Mr. Smith went (ikimashita) to Japan (Nihon)). Of course when a language does not employ a clear distinction between subject, verb, and object(s), then to talk about such patterns is largely irrelevant. For example, "Marlunnik ammassattorpunga" (Greenlandic Eskimo) translates literally as "Two-instrument-plural sardine-eat-1st-singular-Indicative" to mean "I ate two sardines". The objects have been incorporated into the verb and case dependencies and other information is expressed using lexical morphemes functioning as affixes. Although a verb has different cases it is unnatural to insist on subject verb object distinctions (Van Valin and LaPolla, 1997).

This example of syntactic regularity, and others can be found at all levels and for all aspects of language, suggests that although there are universal tendencies in natural languages they should not be construed as immutable general laws. The situation is similar to biological species which also show many regularities, particularly when examining the members of the same broad families, but there are at the same time many differences among species and even a lot of variation among individuals within a species. Two approaches have been adopted to explain these universal tendencies: the universal grammar approach and the complex adaptive systems approach.

2.1 Universal Grammar

The first possibility is that language shows universal tendencies because they form part of the human genetic make-up, similar to the way we all have two hands and walk upright. Thus the different grammatical categories (like adjective or subject), the general patterns on how members of a category can be combined, the types of meanings that can be expressed, parsing and production strategies, language acquisition strategies, the mapping from syntactic form to meaning and back, the ability to engage in dialog, and so on, are all hypothesised to be innately known. Language is largely transmitted genetically and so we can expect there to be a special section of the brain especially wired for natural language. The obvious differences between languages and the evolution of language is accounted for by assuming that there are linguistic parameters set by the brain during development under cultural influence, but all languages are assumed to be based on the same *Bauplan* (Lightfoot, 1991).

This explanation, which was first put forward and defended by Chomsky and linguists in the formal generative tradition (Chomsky, 1981), has some strong implications for other disciplines. Neurobiologists are expected to find very special regions dedicated to language and to identify the genes responsible for them (Rice, 1996). The brain regions involved in language are expected to exhibit unique architectural properties and lesions to these regions should correlate with observed language pathologies. Psychologists can contribute by helping to identify what universal processing mechanisms exist across languages and how parameters get set during development. Engineers can focus on implementing the Universal Language Acquisition Device which is instantiated by exposure to a specific language (an example is shown in Briscoe, 1999).

2.2 A Complex Systems Approach

This Universal Grammar approach has been so popular during the second half of the 20th century that we would almost forget that there is an alternative, which we can call (to use modern terminology) the complex adaptive systems approach. This approach resonates with 19th century evolutionary linguistics as well as contemporary functional (Dik, 1997) and cognitive (Langacker, 1987) approaches to language. Language communication is viewed as a continually changing complex adaptive system, not unlike an economy or an evolving ecosystem. Individual language users are the units in the system and they engage in local interactions with each other. Global sharing arises through self-organisation and the structural coupling between the evolving language competence and the processes developing meaning. The language users have a specific state of knowledge about the structure of the language which they use for their own communicative behavior.

The language system is adaptive in two senses: Individuals produce and interpret sentences partly in a routinised way (otherwise we cannot explain how they can speak so fast). But occasionally the language needs to be expanded by a speaker to deal with novel concepts and situations and these expansions need to be learned so that they can propagate to the rest of the population. Language users optimise their behavior to be more successful in future communications and to minimise the energy and memory resources they need to apply. Because individuals adapt, the language as a whole changes and evolves and this in turn determines how individuals must change if they want to be understood by the rest of the population. Systematicity is always temporary and perhaps less pronounced than a Universal Grammar approach tends to suggest (Hopper, 1987).

The universal tendencies we observe in language are emergent properties which are the consequence of functional requirements, just like the presence of wings can be explained by functional requirements from what it takes to fly. The emergent properties are also influenced and hence explainable in terms of the various constraints under which a language system has to evolve. The constraints relate to the sensori-motor apparatus and cognitive architecture available to human language users, the demands of real world communication without telepathy, within limited time, and with a noisy transmission medium, constraints on learnability, etc. Instead of a genetic origin of language, the complex adaptive systems view leans towards a cultural origin, transmission and evolution of language.

This alternative view resonates strongly with recent work in complex systems in biology and economics, but has also been defended within linguistics itself and

worked out through concrete examples. One example is in the domain of phonetics where some of the top researchers have been trying to explain why natural languages show specific distributions in their sound repertoires based on functional, cognitive and sensori-motor constraints (Lindblom, MacNeilage, Studdert-Kennedy, 1984).

The complex systems approach has again important implications for the various disciplines interested in language. The linguist will now take a more empirical view, looking how new constructs like adjectives, articles or auxiliaries may appear or disappear in a language or how languages may shift from a morphological strategy for expressing case to a word order strategy (Heine, Claudi, Huennemeyer, 1991). The brain scientist is no longer expected to be looking for a set of genes that specifies in full detail the neural microcircuitry implementing universal grammar but now focuses on how the various cognitive functions available to humans have become recruited for language and how the system evolves and develops under cultural constraints (Deacon, 1997). In other words, a very different kind of model of the brain structures involved in language would be hypothesised.

The engineer following a complex adaptive systems approach will also build a completely different system compared to one pursuing a Universal Grammar approach. She would try to build a more general cognitive architecture and a sensori-motor apparatus that approaches that of humans and then perform experiments to see whether such artificial systems can develop languages with the complexity and characteristics of human languages and whether they are able to acquire natural language within similar time and data constraints as experienced by humans. Examples of this work are discussed in Steels (1997), Kirby (1999), De Boer (1997)

Both approaches to language are, I believe, equally plausible and coherent. So it is a matter of working them out and testing them in full detail, as well as performing empirical research to see which one is the most realistic. In my own work and that of my collaborators, we focus on a complex adaptive systems approach and try to develop experimental results rather than theoretical arguments to show why we believe this to be a more realistic approach. The remainder of this paper reports briefly on some of the achievements so far.

3 Experimental Infrastructure

The first thing we have done is develop a general infrastructure for allowing effective experimentation. Since language is about the real world as experienced through sensors and actuators, we have no other choice but to do experiments with artificial robotic agents (Steels and Vogt, 1997). The agents should have some sort of body, a sensori-motor apparatus (cameras, tactile sensors, motors for moving the head or moving around), low-level sensori-motor capabilities, and most importantly a cognitive architecture enabling them to engage in language interactions. The infrastructure we have built is open to any kind of architectural proposal one might want to investigate.

Obviously we need to do experiments with large numbers of agents. This can be achieved by loading agents in different bodies and transporting their mental states from one body to another so that agents can engage in interactions from many perspectives and in many different environments without requiring that they physically move around. We have built a 'cognitive teleporting' infrastructure such that the robots are networked through the internet. This has the additional advantage that experimenters can create agents through the web and send them around in different places and that experiments can be monitored and inspected from wherever the experimenter happens to find herself. This general infrastructure has been fully operational now for over a year and experiments have been done with populations of up to 3000 agents over a period of several months. Sites have been operational in Paris, Brussels, Amsterdam, London, Tokyo, Lausanne, and Antwerp. In total close to a million language interactions have taken place and agents have made tens of thousands of travels over the Internet between different sites. We hope that this infrastructure will become a general test ground for exploring various theories on the emergence and evolution of language understanding in autonomous situated agents. Indeed, the infrastructure is neutral with respect to which theory of language one adopts and is ideally suited to compare the adequacy and performance of different theories.

Our own first experiments (known as the Talking Heads experiments) (Steels and Kaplan, 1998) examined rather simple cognitive mechanisms. The agents had a limited set of sensory channels (aspects of color, shape, and position), mechanisms for segmentation, for identifying the most salient features of the objects in the 2-dimensional scenes before them, and for categorising objects based on evolving discrimination networks. They had a lexical component based on a 2-way associative memory associating words with meanings and meanings with

words. The agents played a game called the guessing game in which one agent tries to identify an object in the scene captured by a camera to another agent using verbal means. When the agents do not have sufficient categories or sufficient words, the discrimination trees or lexical memory expands. We have seen in these experiments that given relatively stable environmental circumstances, shared lexical systems can evolve as long as the population flux is not too high. We have also seen that the lexicon keeps evolving due to fluxes of agents, new environments, etc. (Steels and Kaplan, 1999).

The remainder of this paper focuses on experiments that go significantly beyond these early results because they address the problem of the origins and acquisition of grammar. This implies that two problems are attacked: the origins of more complex compositional meaning and the origins of grammar to express or parse complex meaning. This research is far from finished but sufficiently advanced to justify the first reports.

4 The origins of compositional meaning

Given the size and complexity of natural language meaning, it is obviously a very deep challenge to find mechanisms that can explain how meaning emerges in interactions with the environment and with other language users. We start from two assumptions. The first one is that (at least in its primary function) language is intended for communication. Communication is a form of coordinated action. The speaker hints at an action or a series of actions that she wants the hearer to perform. These actions are either physical actions in the world ("Give me that book") or mental actions to focus on certain items in the context, to store facts for later use ("the book is on top of the refrigerator in the kitchen"), etc. To perform a communication the speaker therefore has to plan what actions she wants the hearer to perform. A possible plan is then translated into an utterance satisfying the conventions adopted by the language community and this plan needs to be decoded and then interpreted by the hearer. Success in communication arises when the effect of the hearer's actions is the one expected by the speaker.

The second assumption is that two traditions in cognitive science will have to be integrated: The *neural network* tradition, which has developed a variety of dynamical systems for categorising real world sensory experiences, for example by weighted decision networks or distance matching to prototypes, and for learning the appropriate weights and thresholds in these networks, and the

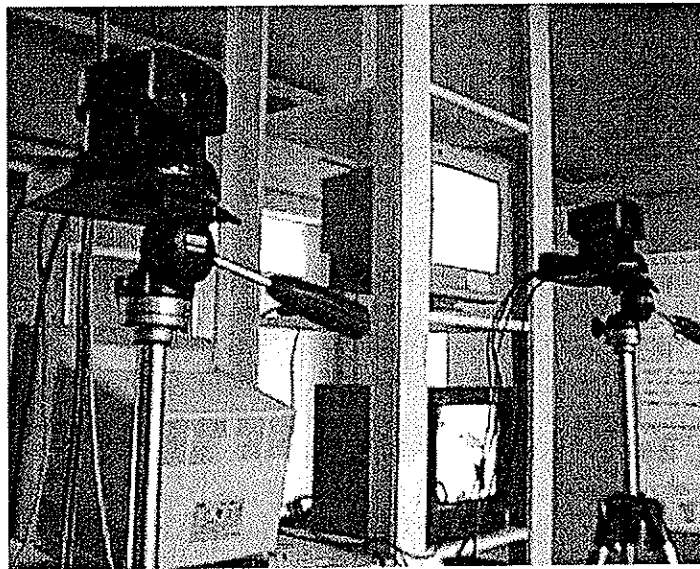


Figure 1: Image of 'Talking Heads' experimental setup with a steerable camera capturing images of geometrical figures in front of them.

logic/symbolic tradition which has developed powerful mechanisms for combining primitive meaningful units into larger compositional structures and for performing inferences or procedural interpretation of such structures. The first tradition is strong in relating categories to real world sensory data. The second tradition is strong in compositionality.

The marrying of two traditions suggests a two-layered system, the first performing subsymbolic computation to generate various ways to conceptualise the scene about which communication takes place, the second supporting in the speaker a planning process to determine what kind of cognitive actions the hearer should perform. The hearer must also perform processes at these two levels: She must compositionally interpret the partially communicated plan and perform a procedural semantics with respect to real world images by using the neural network layer. Part of the difficulty of natural language communication is that the plans formulated by the speaker are only vaguely hinted at. A lot of intelligence is required from the hearer to interpret them correctly.

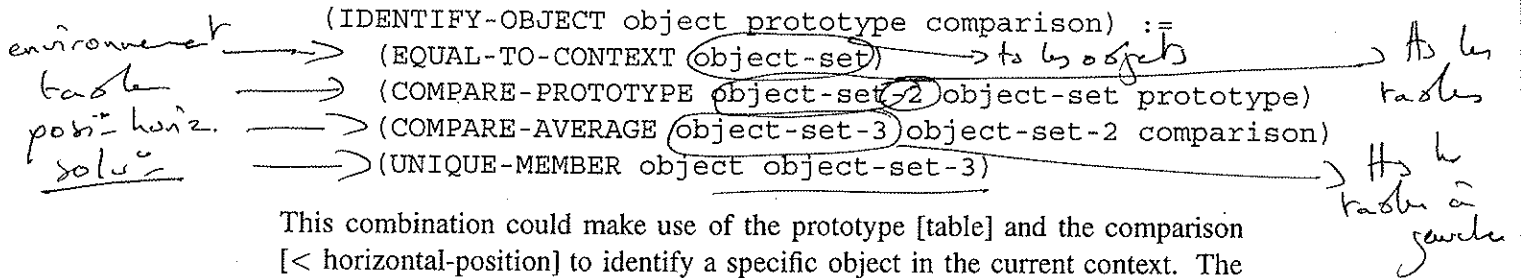
Over the past two years, I have built a computational system that has this kind of two-layered architecture. There is not enough space here to go into great detail, or to summarise the main technical precursors (partly coming from constraint-based systems and symbolic AI programming partly from neural networks and dynamical systems). The primitive components at the subsymbolic layer consist of small networks that learn and perform various kinds of categorisation and use the categorisation for primitive cognitive tasks, like filtering a set of objects into two subsets depending on whether or not they satisfy a category, ordering a sequence of objects based on a category, etc.

One such network, which I will call COMPARE-PROTOTYPE, performs a filtering operation by comparing the elements of a set (further called the source-set) based on their distance to a prototype. The prototypes start out on the basis of concrete input examples of which most of the contingent properties are gradually stripped away. Those elements of the set that are close to the prototype are retained and collected in another set (further called the object-set). Such a network appears useful for the procedural semantics of many nouns. Another network, which I will call COMPARE-AVERAGE, performs a similar operation but now by comparing the values of the elements along a particular dimension (e.g. Horizontal Position) to their average and retaining those that are less than or alternatively larger than the average. Such a primitive network is relevant for the procedural semantics of concepts like 'left' or 'right'. I expect that for natural language semantics we need thousands of such small networks each capable of a particular cognitive operation

1 definition
of the prototype?

and exploiting sensori-motor spaces or their derivation. 2

The second layer is concerned with the combination of these primitive networks in flexible compositional structures. For a phrase like "the left table" as in the sentence "put the box on the left table", a complex combination is required, like the one below:



This combination could make use of the prototype [table] and the comparison [$<$ horizontal-position] to identify a specific object in the current context. The EQUAL-TO-CONTEXT operation sets the object-set to the elements in the present context and UNIQUE-MEMBER picks out one element from an object-set which is assumed to be a singleton. The various arguments (object, prototype, object-set, etc.) are slots that need to be filled in the computation by the primitive networks. M

A study of natural language quickly reveals that the computational processes required in the interpretation of these networks will be very non-trivial. First of all it will have to use data flow (rather than explicit control flow) in the sense that information should propagate in any direction whenever possible. For example, in the phrase "the ball rolls to the edge of the table", we can only uniquely identify the ball after we have identified the edge and also the table. But there may be more than one table, there is any case more than one edge for each table, and there might be several balls - but perhaps only one rolling to the edge of the table. So the computation takes the form of an attempt to find a set of fillers for all slots that is internally consistent and compatible with the present context.]

Second, the computation will have to examine many different possibilities at the same time, as already illustrated by the previous example. In the current implementation this is done by exploring in parallel many different possible worlds, expanding into competing worlds when there are many different hypotheses and collapsing worlds in which some relations are invalid, i.e. in which the underlying network signals a failure to establish the relation.

Third, the computation needs to be usable in any direction. Specifically in the planning process fillers need to be found (for example for the prototype and comparison slots used in the IDENTIFY-OBJECT example given earlier) to test

whether the combination indeeds yields the referent that the speaker wants to communicate.

The planning process can therefore be understood as a search in the space of possible combinations of primitive components. Occasionally there is a successful combination which can then be abstracted out and compiled as a building block for the future. This way a repertoire of complex components gradually arises and so planning becomes mostly the retrieval of high level ready-made plans rather than requiring the microplanning from scratch for a new utterance. The repertoire of stereotyped plans is organised in a hierarchy based on the developmental history of the plans and the structure of this hierarchy can be exploited for language in the sense that speakers can (and do) assume that listeners have similar internal organisations of their meaning repertoires.

At the moment we have a complete simulation operational of this two-layered architecture and done various experiments in planning and executing complex combinations of primitive cognitive operations in order to satisfy communicative goals situated and grounded in visual perception.

5 The emergence of grammar

The planning process produces a set of trees where each tree evokes a combination of primitive components. There can be more than one tree because combinations of primitives can yield a result which is itself used as a slot in another component. The dependencies between the trees is represented by associating an index with every tree which can be used as reference somewhere in another tree. The task of the production process is to turn this tree into an utterance and of the parsing process to reconstruct the semantic trees from the utterance.

I hypothesise that language speakers use a variety of different strategies to do so, each strategy yielding a different sort of language. An example strategy is to use lexical tags attached to basic lexical items to indicate the slots, components and indices. Another strategy is to use word order for some of the same information. The first strategy would be used for example in a language with a strong case system and lexicalisation of case distinctions through affixes. English uses for the same problem a more syntactic approach where word order expresses case relations (i.e. which slot is filled by what). Languages use multiple strategies and strategies shift during the history of the language. Often there is a period of productive use of a strategy, followed by sedimentation and fossilisation, and the

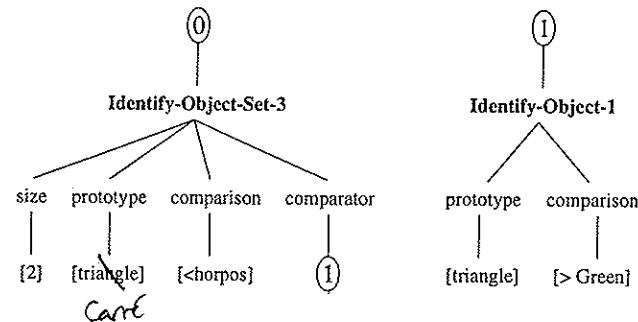


Figure 2: Example of semantic trees produced by the planning process. The trees reflect a similar meaning then invoked by the phrase "The two squares left of the green triangle."

resulting debris is then used as material for other strategies.

A strategy always needs four components:

- A method for routine *production*, which requires specific storage structures like an associative memory for morphologically-oriented strategies or a pattern memory for syntactically-oriented strategies.
- A method for expanding the grammar and lexicon when routine solutions are not available but novel structures need to be verbalised. I call this the *invention* strategy.
- A method for routine *parsing*, which requires similar storage structures but a different usage of the same knowledge.
- A method for *learning* unfamiliar constructions.

Here is an example strategy, the simplest I can think of, which performs full lexicalisation, in other words all the aspects of a semantic structure have corresponding morphemes. Each path in each tree corresponds to a word and the order of morphemes in a word goes from top to bottom. This strategy is of course artificial and not used as such by any natural language, but the idea is that by researching such strategies we can gain progressively a better insight into those that are more realistic and clearly used in natural languages.

The production and invention component of a full lexicalisation strategy is straightforward. The agent just looks up in his lexical memory what the best

lexicalisation is for each node in the tree (i.e. the one with the highest score), traverses each tree from top to bottom and thus collects the different words. If on the way there is a node which has no expression in the language yet, then a new morpheme is created. Note that the index of a tree is also lexicalised and then re-used similar to the use of pronouns. For example for the left tree in figure 2, this strategy yields two words:

(da do di du) (da do bi bu)

assuming that the agent has decided to use the following associations:

da = 1 (an index)
do = identify-object-1 (a component)
di = prototype (a slot)
du = [triangle] (a prototype)
bi = comparison (a slot)
bu = [> green] (a comparison)

Each association has a score within the memory of the agent, reflecting how strongly this association is believed to be valid in the language according to the feedback on language interaction this agent has received.

The parsing strategy is less trivial because of the unavoidable ambiguity and uncertainty causing one word to have more than one possible meaning (and one meaning more than one possible world). But it can be done in the following way: The hearer collects all the possible meanings of each morpheme from the lexicon. Then the hearer collects all possible uses of each meaning in terms of the meaning repertoire. Indeed, the hearer can derive for each bottom-node in the tree what the possible components and slots are in which it fits. It can derive for each possible slots in which components it can be. This generates a lot of possible hypotheses which are then shaken out to retain only those that are internally consistent. The hearer then tries to interpret the remaining semantic structures in terms of the situation of the communication and ideally arrives at one possible plausible result. Note that this process implies a large amount of parallelism, of which hearers are mostly unconscious except in the case of garden path sentences or other problems.

It is entirely possible that the communication fails in the sense that the hearer does not know some of the morphemes used by the speaker, does not know the primitive or complex meaning components hinted at, or simply has the wrong associations between words and meanings. In that case, the learning strategy must

come in action. The one I have used in the present implementation of the full lexicalisation strategy works as follows. The hearer collects the best possible interpretation of the utterance transmitted by the speaker. In other words, the hearer performs his own conceptualisation of the communicative goal and thus guesses what kind of semantic structure the speaker might have intended. Usually there are many possible partial matches between the utterance and the target semantic structure. The best one is chosen and then the language memory is updated. New associations might be stored, the score of associations that are part of the successful match are increased and its competitors decreased. Also in a successful communication, speaker and hearer must adapt their scores to be more successful in future games.

Simulations with robotic agents have shown that this full lexicalisation strategy is effective in the sense that agents build up a shared repertoire of morphemes to express their expanding repertoire of stereotyped plans. Word order is entirely irrelevant in this language. The order of morphemes was adopted under pressure from the difficulty of language acquisition, otherwise there would be an explosion of possibilities. This shows that we always have to keep the four aspects of language processing in mind: parsing, production, invention *and* learning.

Other strategies can readily be invented - as long as they are capable to handle all four aspects of the problem. Another strategy for example is to use word order for expressing which slots are present, as opposed to lexical tags. This can be achieved by associating with each component a pattern that prescribes the order in which the components need to be presented. The production apparatus lexicalises nodes in the trees as before, except the slots. For each component, a pattern is retrieved and then used to order the different words related to the same component. When a new pattern needs to be invented, the speaker can pick any kind of order, perhaps the one that resulted from the semantic planning process.

The parser is similar to the one outlined earlier. The hearer looks up the different meanings for each word, and through the use of the meanings he can discover already what kind of components might have been intended and what roles filler-items may play. From the component, the required order can be retrieved and this way the hearer can determine based on word order what slots the items are filling. Syntax acquisition comes in action when the word order is different from the one expected or when it is not known yet. Similar to the full lexicalisation strategy, the hearer must try to find the best matching semantic structure and then update the lexicon and grammar to reflect the utterance produced by the speaker.

Yet another strategy, which I have also implemented, is to be much more eco-

nomical in expression. The speaker starts by expressing the bottom nodes of the semantic structures and only includes higher level nodes (through word order or lexical tags) when absolutely required. The speaker can simulate the difficulty of parsing by first parsing the utterance internally himself, and only add more information when the utterance is ambiguous with respect to the semantic information and the state of the grammar. This strategy results in a "telegraphic style" of the language reminiscent of pidgin or two year old "protolanguages", particularly in the early phases when not many ambiguities exist yet due to the limited size of the agents' meaning repertoire.

I believe that in the near future, we should formulate many more of these strategies, which each require slightly different processing mechanisms and storage structures. The strategies can be implemented and their properties investigated. Properties relate to stability with respect to transmission, resistance to population flux, ability to cope with increase in complexity of the meaning repertoire, resistance to noise in utterance transmission, etc.

6 Conclusion

Language exhibits various kinds of universal tendencies which are a clue to the kinds of processing that is taking place and the brain architecture that is capable of this processing. I have sketched a research path that takes a complex adaptive systems approach to language. Language speakers are assumed to continuously adapt their state of knowledge of the language on the basis of one-on-one interactions. They expand the language based on strategies licensed in the language in case they need to express entirely novel meanings and they use the context and their evolving meaning repertoire to guess the meaning of utterances which are only partly understood.

Although there is still quite a distance to go, particularly if we insist on experimentation on grounded robots, it appears that this research path is progressively bringing us closer to understand one of the most complex biological phenomena to arise and to build language understanding machines that can play an active role in shaping and adapting the language of a community.

7 Acknowledgement

This research was conducted at the Sony Computer Science Laboratory Paris with major contributions by Angus McIntyre, Frederic Kaplan, and Joris van Looveren (VUB AI laboratory Brussels). Additional inputs and discussions have come from Paul Vogt, Edwin De Jong and Bart de Boer all at the VUB AI Laboratory.

8 References

- Briscoe, T. (1999) *Linguistic evolution through language acquisition: formal and computational models*. Cambridge Univ. Press, Cambridge.
- Chomsky, N. (1981) *Lectures on government and binding*. Foris Publications, Dordrecht.
- Deacon, T. (1997) *The Symbolic Species. The co-evolution of language and the brain*. Norton and co. New York.
- De Boer, B. (1997) Emergent Vowel Systems in a Population of Agents. In Harvey, I. et.al. (eds.) *Proceedings of ECAL 97*, Brighton UK, July 1997. The MIT Press, Cambridge Ma.
- Dik, S. (1997) *The Theory of Functional Grammar*. Mouton De Gruyter. Berlin.
- Friederici, A. (ed.) (1999) *Language comprehension. A Biological Perspective*. Springer-Verlag. Berlin.
- Greenberg, J. (1966) *Language universals*. Mouton, Den Haag.
- Heine, B., U. Claudi and F. Huennemeyer (1991) *Grammaticalisation. A conceptual framework*. The University of Chicago Press, Chicago.
- Hopper, P. (1987) Emergent Grammar. *Berkeley Linguistic Society* 13, 139-157.
- Hurford, J. C. Knight and M. Studdert-Kennedy (eds.) (1999) *Approaches to the Evolution of Human Language*. Cambridge Univ. Press. Cambridge.
- Kirby, S. (1999) *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford University Press, Oxford.
- Langacker, R. (1987) *Foundations of Cognitive Grammar*. Stanford University Press, Stanford Ca.

- Levelt, W. (1989) *Speaking: From Intention to Articulation*. MIT Press, Cambridge.
- Lightfoot, D. (1991) *How to set parameters: arguments from language change*. The MIT Press, Cambridge Ma.
- Lindblom, B., P. MacNeilage, and M. Studdert-Kennedy (1984) Self-organizing processes and the explanation of phonological universals. In: Butterworth, G., B. Comrie and O. Dahl (eds.) (1984) *Explanations for Language Universals*. Walter de Gruyter, Berlin. pp. 181-203.
- Rice, M. (1996) *Toward a Genetics of Language*. Lawrence Erlbaum Ass. New Haven.
- Steels, L. (1997) The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1-35.
- Steels, L. and Kaplan, F. (1998) Situated Grounded Word Semantics. In *Proceedings of IJCAI-99*, Stockholm. Morgan Kauffman Publishing, Los Angeles. p. 862-867.
- Steels, L. and Kaplan, F. (1999) In *Proceedings of COLING-ACL*, Montreal, August 1998, p. 1243-1249.
- Steels, L. and P. Vogt (1997) Grounding Adaptive Language Games in Robotic Agents. In Harvey, I. et.al. (eds.) *Proceedings of ECAL 97*, Brighton UK, July 1997. The MIT Press, Cambridge Ma.
- Van Valin, R. and R. LaPolla. (1997) *Syntax. Structure, meaning and function*. Cambridge University Press, Cambridge.