

DEVELOPING AGENTS THAT CAN SPEAK WITH HUMANS: POINTERS FROM THE EVOLUTION OF LANGUAGE

CAROLINE LYON

*Science and Technology Research Institute, University of Hertfordshire,
Hatfield, Hertfordshire, AL10 9AB, UK
c.m.lyon@herts.ac.uk*

CHRISTOPHER L. NEHANIV

*Adaptive Systems Research Group, University of Hertfordshire,
Hatfield, Hertfordshire, AL10 9AB, UK
c.l.nehaniv@herts.ac.uk*

To interact in a natural manner with humans robots must “understand” some ordinary speech, however limited. This paper addresses a preliminary issue: the need to process short sequences of words as grammatical fragments, evidenced by empirical linguistic observations. The evolution of language can inform the development of agents’ communication, and insights from ethology, anthropology and neuroscience suggest that evolved sequential processors are probably exaptations. Possible steps in the emergence of compositionality also indicate a need to process sequences of linguistic items as units. Recent results from neuroscientific research suggest ways in which serial processors might be modelled

1. Introduction

The development of an artificial language that can enable synthetic agents or robots to interact with humans is an active area of research, and a number of different approaches are possible. For instance synthetic agents might communicate with each other in an artificial language that is comprehensible to humans, and a tutored human might use it to communicate with them.

However, learning such a communication system would not enable the agents to process ordinary natural language, to “understand” humans. If robots are to interact in a more natural manner with untutored humans they will have to handle some normal human speech, however limited its scope. A typical restricted scenario could be a variation on the familiar block world, where human and robot converse about moving various objects around.

In this paper we examine a small part of the development of such a communication system congruent with natural language. This can be informed by human language evolution and by child language acquisition (though these two processes

do not necessarily follow the same path). We look at the evolution of language to get pointers to possible directions for the first steps. Our work focuses on the fundamental issue of processing short sequences of words which comprise linguistic elements, phrases in most cases, and the modelling of serial order.

First, we present linguistic evidence to demonstrate why grammatical fragments need to be processed as basic elements, as well as single words and whole sentences, in Section 2. In Section 3 we examine possible reasons why language evolved in the way it has. We present insights from ethology and anthropology, and from relevant research in neuroscience.

Section 4 examines some of the short word sequences that are the particular focus of this work, and in Section 5 we conclude with some pointers to the practical consequences for the development of the agents' language. Recent neuroscientific research suggests ways in which serial processors could be modelled.

2. Short word sequences as linguistic elements and word ambiguity

In English, Japanese, Chinese and other languages there are many common homophones such as

two / to / too for / four their / there so / sew

words that sound the same but have different meanings. We can observe present day language slipping into homophonous forms (Warren, Rae, & Hay, 2002).

In spite of the evidence around us that a model of language must accommodate the phenomenon of ambiguity, and that it does not seem to impair our communicative capacity, some simulations of the evolution of human language overlook this. For instance Nowak, Komaraova, and Niyogi (2002, p 613) have asserted that "ambiguity is the loss of communicative capacity that arises if individual sounds are linked to more than one meaning" and that the absence of word ambiguity is a mark of evolutionary fitness. There is even a Homeric nod from Oudeyer who says "For efficient communication, it is better that different words are associated with different meanings. This obvious remark" (Oudeyer, 2007). It is possible that the avoidance of word ambiguity could be an objective of some artificial communication system, but not as part of an explanation as to how human language evolved. We take ambiguous words in context, and usually only one interpretation is possible.

The most frequently occurring homophones are those in which at least one of the ambiguous words is a function word, as in the examples above, (Ke, Wang, & Coupe, 2002), which we call type 1 ambiguity.

In the LOB corpus of about 1 million words 20 of the 50 most frequently occurring words are type 1 homophones (Johansson & Hofland, 1989, page 19)^a.

Pairs of homophonous content words, which have different semantic meanings, also occur, e.g. *byte / bite*. We call this type 2 ambiguity. The homophonous

^ato, in, for, be, I, by, not, but, are, which, you, there, been, one, we, their, would, so, no, will.

words may have a common ancestry, for instance *chip*, *mobile*, but have developed distinct meanings over time.

We also see content words with shades of meanings, such as Wittgenstein's well known example of the word "game", whose various meanings can be seen as family resemblances, "a complicated network of similarities" (Wittgenstein, 1953, sections 66-76). For another example consider how we use the word "space". This has a core the concept of "void", but it can mean a gap between words, a place to park a car, or the outer reaches of the universe.

Syntactic ambiguity is also common, since content words in English can often act either as nouns or verbs; for example in the blocks world we could meet words like

push *move* *place* *lift*

in utterances like "give it a push" (noun) or "push it" (verb).

In this paper we focus on the commonest type 1 ambiguity. An examination of samples of speech or text shows that humans usually have no difficulty in disambiguating these words because they are taken in context, typically very short, a phrase of just a few words. (A much longer context may be needed for semantic disambiguation. Consider a sentence such as: "Here are some chips".) We need to address the issue of processing short sequences of linguistic elements as coherent units. In the case of a language like English with pronounced word order constraints these linguistic elements will be words; for an inflected language they may be word parts, morphemes. In either case we have to process short sequences rather than single items.

3. The evolution of serial processors

At first sight it is surprising that we find so much ambiguity in natural language. There are plenty of "spare" phonemes that could be employed to produce distinct, non-ambiguous sounds. The number of phonemes varies widely between languages, from less than 20 to over 100 (Maddieson, 1984). Some of the most salient phonetic discriminators are rarely used.

The fact that we do not avoid ambiguous words but resolve their meaning by processing short sequences suggests that such serial processing methods were easily accessible. It seems likely that sequential processing is based on exaptations of faculties that were originally developed for different purposes. Lieberman (2002) shows that neural sequential processors in the basal ganglia, which are needed for motor control, are also involved in language production and perception.

Pulvermuller (2002) reports neuroscientific research showing how serial order is detected in the human brain. The basic linear synfire chain underpins serial order, and though it may not be applicable to sequences of meaningful linguistic units (ibid, page 155), he proposes an adaptation that is known in other neuroscientific domains.

Steel's Recruitment Theory of language is consistent with this neuroscience

research: “the human language faculty is a dynamic configuration of brain mechanisms, which grows and adapts recruiting available cognitive/neural resources for optimally achieving the task of communication” (Steels, 2007).

The Mirror System Hypothesis which relates the comprehension of an utterance to its production also suggests that linguistic communication makes use of more primitive motor processors (Arbib, 2002).

3.1. *Primate behaviour*

There is evidence that sequential processing is a necessary part of primate communication. Goodall (1986, page 144) reports that “chimpanzees’ repertoire of species-specific gestures and calls will gradually be patterned and organised into functional sequences” in a normal social environment. Chimpanzees raised in isolation for the first 2 years still showed a number of characteristic communication signals but these appeared in incomplete sequences and inappropriate contexts when the isolates were introduced to normal chimpanzee youngsters. This suggests that there is a genetic predisposition to process sequential signals in coherent units, but this has to be triggered by social or environmental stimuli.

3.2. *Negation and the first stages in compositional structure*

In a pre-language state of paratax, communicative gestures or utterances are independent elements, not linked together. In contrast a fully evolved language is characterised by its syntax, in which linguistic elements are related to each other in a hierarchical grammar, with dependencies between the different components. These components can have a compositional structure, so we can take a noun phrase such as “the red block” and replace *red* by another adjective to produce a new grammatical phrase. However, before this state we can hypothesize an intermediate stage of proto-syntax.

The simplest form of compositionality is a combination of two concepts, via concatenation or superposition. This precedes more advanced compositional forms, where the ability to categorise words is necessary. Using the term “proto-syntax” in this way is consistent with its use in logic, where Quine (1940) takes it to mean syntax without the concept of class membership.

An obvious candidate for an early form of proto-syntax is negation, in which a negating term is added to a primitive utterance or a holophrase: negation can be verbal, non-verbal or both. There can be a parallel expression of communicative signals, via prosodic or gestural marking (Neidle, Kegl, MacLaughlin, Bahan, & Lee, 1999), with the advantage that scope can be indicated by the overlap in the superposition (cf. (Nehaniv, Lyon, & Cangelosi, 2007)). Alternatively, two concepts can be concatenated, the case we consider here. It is more informative to negate an existing utterance, thus linking two concepts, than to produce a new, unrelated holophrase. It can be used for objects: “not food”; for actions: “not touch”; for

descriptors: “not good”. It is a simple form of compositionality, without other grammatical features, such as part-of-speech categorisation.

Other candidates for proto-syntactic structure are interrogation and conjunction. An assertion can be converted to a question by adding an interrogatory marker, such as “est-ce que” in French. And the concatenation of two holophrases is an example of conjunction.

There are a number of reasons why compositional forms should have evolved. First, they enable learning from sparse data, as demonstrated by Kirby (2002) and others. They promote semantic cohesion, since related concepts are linked. They give expressive power - a limited number of elements can be combined to give an indefinite number of expressions. In addition to these reasons we can add the hypothesis that humans are predisposed to sequential processing, to concatenate utterances into new linguistic elements.

3.3. Anthropological indicators

There is much debate about early hominid life styles, but group hunting and fishing undoubtedly figured at times, and is also seen in other mammalian species. Now, among the many varied purposes of linguistic communication there is a need for planning and executing group activities. In making such plans some language forms will become increasingly useful, if not essential, such as prepositional phrases:

wait *behind* the rock hide *in* the tree *at* the end *of* the path

They relate verbs to nouns, and nouns to nouns. Such phrases make up linguistic entities of three or more elements, and these short sequences have to be processed. Before we reach the stage of producing a full hierarchical grammar we need to handle sub-sentential phrases.

4. Short sequences of words in natural speech

Much work has been done in computational linguistics on analysing grammar at the sentence level. However, effective communication includes sub-sentential, grammatical elements, as is apparent from a glance at newspaper headlines, the titles of papers for a conference, or everyday speech. For example:

“Where is the butter?” ”In the fridge”.

The phrase “in the fridge” is meaningful even though it is not a complete sentence. However, it is still a *grammatical* fragment, as opposed to a sequence like “fridge the in”. Tomasello (2003, page 111) reports that mothers talking to their 2 year old children used fragments of sentences or phrases 20% of the time.

4.1. Regular phrases and “constructional islands”

In simulations and models of the evolution of language the analysis of paradigmatic categories, typically content words such as nouns, verbs, adjectives, has been an active research area.

Now, just as important for effective communication are the frequently occurring function words. Short sequences, such as prepositional phrases, are grammatical fragments that need to be processed as linguistic elements. Prepositions constitute a distinct part-of-speech class, they behave in the same way much of the time (though there are ambiguities with particles: consider “look up the chimney” and “look up the number”)

However, there are many other grammatical fragments, common expressions, that do not fit neatly into classes of structures. Tomasello (2003, page 149) calls these “constructional islands” . Consider the following quantifiers that might occur in a blocks world dialogue:

all both each some

In some respects they behave the same way, so we can say:

all of the bricks both of the bricks each of the bricks some of the bricks

but then we find that they differ:

all the bricks both the bricks each the bricks some the bricks**

Any artificial language congruent with human language will have to accommodate both the regular structures based on part-of-speech classes and the irregular ones.

5. Pointers for the development of agents' language

If we take an incremental approach to investigating stages in language development there are various starting points focused on the analysis of different aspects of language. One possibility is to reduce our language to small sets of words and/or a few paradigmatic syntactic categories. With this type of development the human might eavesdrop on robots and understand their communications, but the robots would not be able to “understand” normal human language.

Another approach is to take as a basic requirement the need for the robot to process natural human speech, however limited, using automated speech recognition, now quite a mature technology, and speech synthesis.

Taking this approach we can envisage a simple scenario, such as the blocks world, in which each of the players can ask the other to take a certain action, acknowledge that a request has been heard, and carry out simple tasks. Language structure would be modelled in incremental stages. Initially, the dialogue would be unstructured holophrases, strings of words that are taken as a whole, a state of paratax. We can imagine that the robot or agent learns the meaning of these holophrases through imitation, or from signs of approval or disapproval. The next stage would be to a state of proto-syntax, where terms are concatenated:

do A and do B or not A

Already some hierarchically structured skills can be transmitted to robots in this way via social learning (Saunders, Nehaniv, & Dautenhahn, 2006), and associating aspects of such learned structures with communicative signalling seems feasible. A later step in the development of the communication system would be

to move to further compositional forms.

A reason in favour of this approach is that it might follow the path of evolution of language in humans, as discussed above. Furthermore, it has been shown, using Information Theoretic analysis, that processing speech in grammatical segments could confer an evolutionary advantage, before the benefits of having a full hierarchical grammar are realised (Lyon, Dickerson, & Nehaniv, 2003). Segmented speech is better understood than an unsegmented stream, and a stage in the evolution of language could have had its own benefits before the advantages of a full grammar were realised.

5.1. Modelling sequential data

If this approach should be taken, a focus of investigation would be the modelling of sequential data.

A conceptual modelling based on neurophysiological research (Pulvermuller, 2002) could be investigated. Empirical discoveries about synfire mechanisms, serial processors, have provided a basis for phonological processing. Though Pulvermuller doubts their application to processing word sequences, for reasons that include timing constraints, he proposes further structures that are compatible with what is already known about neuronal organisation. In a grossly oversimplified description, the hypothesis is that there could be mediated serial processors for word sequences. Two linguistic components produced in sequence would trigger the firing of a third neuronal element, a *sequence detector* that indicated a serial unit.

The relationship between simulations and real world observations is not that one matches the other precisely. From a neuroscience point of view “[it is] not so much that actual simulations taught us lessons ... [but they] made scientists think about details of neural circuits, and this led to important insights.” (ibid, page 109).

For the computer scientist real systems inspire synthetic models. We cannot hope to emulate natural neural circuits constructed out of wet ware, but on the other hand some constraints such as the timing of neuronal activities may be circumvented. By examining real processes and actual language as it is used we may hit upon profitable approaches to the development of artificial systems, and gain insights into the acquisition of language.

References

- Arbib, M. (2002). The mirror system, imitation, and the evolution of language. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts*. MIT Press.
- Goodall, J. (1986). *The chimpanzees of Gombe: patterns of behaviour*. Harvard University Press.

- Johansson, S., & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar*. Clarendon.
- Ke, J., Wang, F., & Coupe, C. (2002). The rise and fall of homophones: a window to language evolution. In *Proceedings of 4th international conference on the evolution of language*.
- Kirby, S. (2002). The Emergence of Linguistic Structure: an overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*. Springer.
- Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology*.
- Lyon, C., Dickerson, B., & Nehaniv, C. L. (2003). The segmentation of speech and its implications for the emergence of language structure. *Evolution of Communication*, 4, no.2, 161-182.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University press.
- Nehaniv, C. L., Lyon, C., & Cangelosi, A. (2007). Current work and open problems: A roadmap for research into the emergence of communication and language. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language*. Springer.
- Neidle, C. J., Kegl, J., MacLaughlin, D., Bahan, B., & Lee, R. G. (1999). *The syntax of American sign language: Functional categories and hierarchical structure*. MIT Press.
- Nowak, M. A., Komaraova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, 417, 611 - 617.
- Oudeyer, P.-Y. (2007). Language evolution as a Darwinian process: computational studies. *Cognitive Processing*, 8, 21-35.
- Pulvermuller, F. (2002). *The Neuroscience of Language*. Cambridge University Press.
- Quine, W. (1940). *Mathematical Logic*. W. W. Norton and Co.
- Saunders, J., Nehaniv, C. L., & Dautenhahn, K. (2006). Teaching robots by moulding behavior and scaffolding the environment. In *1st annual conference on human-robot interaction (hri2006)* (p. 142-150). ACM Press.
- Steels, L. (2007). The recruitment theory of language origins. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language*. Springer.
- Tomasello, M. (2003). *Constructing a language*. Harvard University Press.
- Warren, P., Rae, M., & Hay, J. (2002). Goldilocks and the three beers. In *9th Australian international conference on speech science and technology*.
- Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell. (Translated by G. Anscombe)