LANGUAGE AND THE EVOLUTION OF COGNITION

Peter Gärdenfors

Lund University Cognitive Science
Kungshuset, Lundagård
S–222 22 LUND

E-mail: peter.gardenfors@fil.lu.se

Abstract: The main purpose of this article is to discuss the kinds of mental representations that are required for language to evolve. Firstly, I distinguish between *cued* and *detached* representations. A cued representation stands for something that is present in the current external situation of the representing organism, while a detached representation may stand for objects or events that are neither present in the current situation nor triggered by some recent situation. The inner environment of an agent is defined as the collection of all detached representations of the agent. The fundamental difference between signal and a symbol is that the reference of a *symbol* is a detached representation, while a *signal* refers to a cued representation. *Icons* also refer to detached representations, but unlike symbols, the choice of representation is not arbitrary, since an icon in some aspects *resembles* the thing it represents.

A Gricean analysis shows that human linguistic communication presumes an advanced kind of inner environment in order to represent higher order intentions. Not only must we be able to represent the inner environment of other people (i.e., have a "theory of mind"), but we must also represent the other individual's representation of *our* inner environments.

In relation to the evolution of grammar, I distinguish between three levels of grammaticality in a communication system: Systems with no grammar, compositional systems, and systems with grammatical structure. Combining this tripartition with the distinction between cued and detached representations, one obtains six kinds of communication systems. Most animal signaling systems use cued representations and no grammar, but bee's dances have a compositional grammar, while still exploiting cued representations.

The evolutionarily first communication systems using detached representations were one-word languages. Gestural communication using icons preceded vocal language using symbols. Donald's proposal of a *mimetic stage* is an important step in the evolution of language. Gesturing and primitive speech then developed into a protolanguage, i.e., a communication system with a compositional structure, but without grammatical items. Requirements of fast and efficient communication finally resulted in the development of arbitrary symbols and syntactic rules, at the cost of iconicity, to arrive at a language with a full grammar.

1. CUED AND DETACHED REPRESENTATIONS

When we communicate by language, our utterances have meaning. The meaning of what we say is represented in our minds. But how do these representations function? It has turned out to be extremely difficult to teach other animals to communicate linguistically. Is there any difference between their mental representations and ours that can explain why they can't learn a complete language? How could such a difference have evolved? This article addresses these questions. The main purpose is to disclose the kinds of mental representations that are required for language to evolve.

The first thing to notice is that there are different kinds of mental representations. In Gärdenfors (to appear b), I distinguish *cued* and *detached* representations, which will turn out to be crucial for the arguments of the present article.

A *cued* representation stands for something that is present in the current external situation of the representing organism. When, for example, a particular object is categorized as food, the animal will act differently than if the same object had been

Lund University Cognitive Studies – LUCS 41 1995. ISSN 1101–8453.

categorized as a potential mate. In general, the represented object need not be actually present in the actual situation, but it must have been triggered by something in a recent situation. Delayed responses, in the behaviorist's sense, are thus based on cued representations. I am not assuming that the animal is, in any sense, aware of the representation, only that there is some generalizing factor that determines its behavior.

In contrast, *detached* representations may stand for objects or events that are neither present in the current situation nor triggered by some recent situation. A memory of something, that can be evoked independently of the context where the memory was created, would be an example of a detached representation. Other examples of detached representations are the "spatial maps" that were introduced by Tolman (1948) in order to explain the behavior of rats in different kinds of mazes.[1]

However, I don't claim that a sharp distinction between cued and detached representations can be maintained. It would perhaps be better to talk about *degrees* of detachment. For example, even the capacity for representing *object permanence* (in Piaget's sense) involves some level of detachment. A cat can, for example, predict that a mouse will appear at the other side of a curtain when it disappears on one side. It can "infer" information about the mouse even if there is no immediate sensory information, like when it is waiting outside a mouse-hole. In this sense it has a detached representation of a mouse that is, at least to some extent, independent of the information that is provided by the senses.

The role of detached representations in the mental life of an organism can be explained by relating it to an idea introduced by Craik (1943, p. 61):

> If the organism carries a "small-scale model" of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which are the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react on a much fuller, safer and more competent manner to the emergencies which face it.

Under the heading of the *inner environment* this kind of "small-scale model" has been made popular by Dennett:[2] "the inner environment is simply any internal region that can affect and be affected by features of potential behavioral control systems" (1978, p. 79). Such an environment is necessary for representing objects (like food and predators), places (where food or shelter can be found), actions (and their consequences), etc., even when these things are not perceptually present. The evolution of this kind of representation will clearly increase the fitness values of the animal.

As a tentative definition, the inner environment of an organism will be identified as the collection of all detached representations of the organism. Loosely speaking, the inner environment consists of all things the organism can actively "think" about.

The existence of an inner environment is necessary for many higher cognitive functions like planning, deception, and self-awareness (Whiten and Byrne 1988, Dennett 1991, Gulz 1991, Gärdenfors 1992, Gärdenfors to appear). A special case is when an organism in its inner environment represents the inner environment of another individual. This results in what has been called a "theory of mind" (Gopnik 1993, Gergely 1994). And as I will argue in Section 3, the inner environment is also a *sine qua non* for language.

It is difficult to assess when detached representations first appeared in the animal kingdom, but a wild speculation is that it is coordinated with the development of the neocortex, i.e., roughly with the appearance of mammals. However, it is only with the development of *crossmodal* representations that we obtain advanced forms of an inner environment (Davenport 1976, Murray 1990, Allott 1991).[3] It is interesting to note that the human language function does not reside in the same places in the brain as the call systems of the other apes (Deacon 1992). The call systems are *automatic* reactions which cannot be suppressed. The development of the areas in the frontal lobes allowed language to develop as a *voluntary*, i.e., detached, system (Barber and Peters 1992, p. 316).

In support of the general speculation concerning the correlation between detached representations and the neocortex, one can note that mammals *play*, but reptiles don't (Sjölander 1993). Playing is a way of building up a repertoire of behaviors that can be used at later occasions. However, this mechanism presupposes that the behaviors are represented in a detached way (see Gulz 1991). There is also evidence of *dreaming*, which apparently presumes an inner environment, only among the mammals.

---

[1] See Balkenius (1995) for an account of how the mechanisms of detached representations may arise as an extension of reactive behavior and various forms of learning.

[2] However, Dennett does not refer to Craik. For a related, more constructivist idea, see Sjölander (1993).

[3] Murray (1990) argues that the amygdala is crucial for crossmodal sensory–sensory associations.

2. SIGNALS, ICONS AND SYMBOLS

In my opinion, thinking does not presume a language. Humans, as well as animals, can simulate sequences of actions in their inner environments. Such simulations are, among other things, necessary for planning. For example, consider the high jumper who mentally penetrates his bodily movements before actually performing the jump.[4]

In contrast, I believe that language presumes the existence of an inner environment. In order to make this clear, I will introduce a way of distinguishing between *signals* and *symbols*. Both signals and symbols are tools of communication. The fundamental difference between them is that *the reference of a symbol is a detached representation, while a signal refers to a cued representation*. In other words, a signal refers to something in the outer environment, while a symbol refers to the inner environment. A similar characterization can be found in von Glasersfeld (1977, pp. 63–65), who traces the idea back to Langer (1948).[5] She clearly distinguishes symbols from signals:

> A term which is used symbolically and not signally does *not* evoke action appropriate to the presence of its object. [...] Symbols are not proxy for their objects, but are *vehicles for the conception of objects*. To conceive a thing or a situation is not the same as to "react toward it" overtly, or to be aware of its presence. In talking about things we have conceptions of them, not the things themselves; and *it is the conceptions, not the things, that symbols directly "mean."* Behavior toward conceptions is what words normally evoke: this is the typical process of thinking. (Langer 1948, p. 61)

With few exceptions, linguistic communication is achieved with the aid of symbols. Sjölander (1993, pp. 5–6) explains elegantly what is missing in animal communication:

> The predominant function of language is to communicate about that which is not here and not now. A dog can 'say': I am angry, I want water, I want to go out, I like you, etc. But it has no communicative means enabling it to 'say': I was angry yesterday, nor can it 'say': I will be angry if you lock me up tonight again, and I will chew up the carpet. Likewise, the dog can 'say': There is a rat here! but it cannot 'say': There is a rat in the next room.
>
> […] Clearly, if you live in the present, communicating mainly about how you feel and what you want to do in the moment, the biological signals inherent in each species are sufficient. A language is needed only to communicate your internal representation of what could be, what has been, and of those things and happenings that are not present in the vicinity.

Symbols refering to something in one person's inner environment can be used to communicate as soon as the listeners have, or are prepared to add, the corresponding references in their inner environments.[6] The actual conditions of the outer situation need not play any role for the communication to take place: two prisoners can talk fervently about life on a sunny Pacific island in the pitch dark of their cell.

The idea that symbols refer to detached representations is not quite the same as Hockett's (1960) notion of "displacement" which is one of the criteria he uses to characterize language. Hockett's notion is too weak as is clear from the following: "Any delay between the reception of a stimulus and the appearance of the response means that the former has been coded into a stable spatial array, which endures at least until it is read off in the response" (Hockett 1960, p. 417). This phrase has a clear behaviorist ring to it and any signal that occurs at a different place or time would count as displaced according to Hockett's criterion. Von Glasersfeld (1977, p. 64), makes the point as follows:

> But language allows us to talk not only about things that are spatially or temporally remote, but also about things that have no location in space and never happen at all. The very fact that we can make *understandable* linguistic statements about space and time,

[4] For a fascinating account of the neural representation of motor intention and motor imagery, see Jeannerod (1994).

[5] Already de Saussure (1984) proposes, in contrast to mainstream contemporary philosophical semantics, that "la signifié" is a mental entity. The following excerpt from the first paragraph of the first chapter illustrates this: "for some people a language, reduced to its essentials, is a nomenclature: a list of terms corresponding to a list of things. […] This conception is open to a number of objections. It assumes that ideas already exist independently of words […]. It does not clarify whether the name is a vocal or psychological entity […]. Furthermore, it leads one to assume that the link between a name and a thing is something quite unproblematic, which is far from being the case. None the less, this naive view contains one element of truth, which is that linguistic units are dual in nature, comprising two elements. […] the two elements involved in the linguistic sign are both psychological and are connected in the brain by an associative link. This is a point of major importance."

[6] For a model theoretic account of how such communication can be established, see Gärdenfors (1993).

> right and wrong, Humpty–dumpty, and the square root of minus one demonstrates rather incontrovertibly that language can deal with items that have nothing to do with "observable stimuli" or with the "referents" of the traditional theory of reference in linguistic philosophy.

Following Peirce's (1932) trichotomy of signals (which he calls indices), icons, and symbols, the role of *icons* can be characterized as follows. Like symbols, icons refer to detached representations, but unlike symbols, the choice of representation is not arbitrary. On the contrary, an icon in some aspects *resembles* the thing it represents.[7] We will return to the relation between icons and symbols below in Section 6.

Many animals have intricate systems of signals, for example, the dances of bees. This kind of dance has a kind of "lexicon," it exhibits combinatorial patterns of the elements in the lexicon, and it even satisfies Hockett's "displacement" since the dance refers to a nectar find that is remote from the hive where the dance is performed. However, even if the bees' dances seem to have a kind of grammar, they still consist only of *signals*. The bees categorize places where nectar can be found in a sophisticated way. The crucial point is that they only use their dances in a *cued* manner, and thus the dances are not symbols according to my criterion. The same point is made by von Glasersfeld (1976, p. 222): "In my terms, the bees do not qualify for symbolicity, because they have never been observed to communicate about distances, directions, food sources, etc., without actually coming from, or going to, a specific location."[8] And in Glasersfeld (1977, p. 65) he adds:

> To qualify as language, the bees' dance would have to be used also *without* this one–to–one relation to a behavioral response (e.g., in comments, proposals, or questions concerning foraging location), and this has never been observed. In short, a communication system that allows for *imperatives* only – no matter how sophisticated and accurate they might be – should not be called a language.

3. THE EVOLUTION OF COMMUNICATION

The fact that a language consists of symbols refering to detached representations is a necessary, but far from sufficient, condition to separate language from other forms of communication. I next turn to what needs to be added to this condition.

The first thing to notice is that human linguistic communication presumes an advanced kind of inner environment. To see this, let us turn to Grice's (1957, 1969) theory of meaning. His initial definition in the second paper reads as follows (1969, p. 151):[9]

> "*U* meant something by uttering x" is true iff, for some audience *A*, *U* uttered *x* intending
> (1) *A* to produce a particular response *r*.
> (2) *A* to think (recognize) that *U* intends (1).
> (3) *A* to fulfill (1) on the basis of his fulfillment of (2).

Although he defines "meaning," I am more interested in applying the definition to linguistic communication in general. The feature I want to focus on here is that condition (2) expresses a *third–order intention* (Dennett 1978, p. 277-278): U intends A to think that U intends something. Gomez (1994, p. 68) even claims that a truly requestive situation like "May I have some salt, please?" involves a *fifth-order* level of intentionality: U *wants* A to *understand* that she *wants* him to *understand* the she *wants* the salt.[10] Now, in what kinds of inner environments can such higher-order intentions be formed?

A crucial problem is determining how such an intricate system of representations could have evolved, and to what other cognitive functions it is related. In my opinion, the first step in the evolution of higher order intentions is when other agents are not only seen as things acting, but as having an *inner environment of their own*, with beliefs, desires, etc. Another way to express this capacity is to say that the representing organism has a *theory of mind*.[11] Once this level of representation is achieved, an organism can have goals concerning the intentions of other individuals, e.g., *want* somebody to *believe* that

[7] Barber and Peters (1992, p. 315) write: "An icon can be interpreted without previous agreement, through general knowledge of the world, and an index [i.e., a signal] through either knowledge of the world or pre-wired instinct. But an arbitrary symbol can only be interpreted through the direct process of agreeing on a convention and then learning it. That is, some preliminary mode of communication is needed to begin making the conventional agreements that underlie arbitrary systems. Icons and indices can serve this bootstrapping function because they can exist without conventional agreement. Thus spoken communication, like writing and sign, had to have begun iconically and/or indexically, and gradually shifted to arbitrariness."

[8] Benveniste (1966, p. 61) says similarly: "On n'a pas constaté qu'une abeille aille par example porter dans une autre ruche le message qu'elle a reçu dans la sienne, ce qui serait une manière de transmission ou de relais."

[9] This definition is revised several times in the second paper, but the more complicated versions have the same general structure as the definition given here.

[10] However, he also claims that the *mutuality* of intentional communication can be achieved by "attention contact" without metarepresentations of the inner environment of the other (Gomez 1994, p. 73).

[11] Gergely (1994, p. 54) provides a list of properties of a representational system that allows for a theory of mind.

an attack would fail. This is an example of a second–order intention.

It is only when this level is achieved that *deception* becomes possible. Deception, in its genuine sense, presumes a model of other minds. More precisely, deception presupposes that the deceiver has some representation of how the individual to be deceived will *interpret* the deceiving act. In other words, deception presupposes that the inner environment of the deceiver contains some form of *representation of the inner environment of the target individual*.

Whiten and Byrne (1988) present a series of examples of deception among primates. Most examples come from field observations of chimpanzees and baboons. However, there are also cases when it is clear that deception is *not* taking place: The partridge feigning a broken wing to lure away the fox from her chicks is not *fooling* the fox (Gärdenfors, to appear b). "Fooling" presumes an intention to make somebody else misinterpret the fooling act, and here there is no evidence that the partridge has any representation of what the fox thinks. She merely acts instinctively when the fox approaches, i.e., the representation of the danger is cued, and can hence not have any intention to fool.

Next step in the evolution of the inner environment of an individual U is for U to realize that *the inner environment of another individual A may in turn contain a representation of the inner environment of U*. Only then can one meaningfully express third–order intentions, e.g., that "U intends A to think that U intends something."

Self-awareness can then develop as a shortcut in this representation: I can in my inner environment have a representation of my own inner environment.[12] However, I submit that this kind of self-awareness could never develop without the previous establishment of a *you-awareness*, i.e., the representation of the inner environment of the other individual (see Mead 1934, Gärdenfors 1992, to appear a, Gomez 1994). Bråten (1988) even goes as far as to propose that in humans, a "virtual other," i.e., a representation of another being's mind, is an *innate* propensity of our inner environments. Some animals have a you-awareness, at least in the sense that they can act deceitfully, but it seems to be only humans who have a full self-awareness.[13]

The importance of this analysis with respect to language, however, is that communication in Grice's sense leads to an elaborate nesting of inner environments.[14] The upshot is, if I am correct, that a full-blown linguistic communication presumes a mind that is capable of you-awareness as well as self-awareness. A consequence of this is that language, in the normal sense, is most likely a *very* recent phenomenon in the evolution of thinking.[15] I will return to this topic in Section 5.

But can one think at all without language? We all have the experience of something like an omnipresent inner monologue (or dialogue) while we are engaged in thinking. I believe that this experience is deceptive. Firstly, we can "think" without language. Consider, for example, the previously mentioned mental simulation of a high jumper.[16] Secondly, and more importantly, the inner speech is best interpreted as just parts of the *simulations* in the inner environment. The inner soliloquy is part of what we *perceive* in the inner environment.[17] The *production* of the monologue is, however, hidden in the unconscious, just as we are not aware of how we find our words when we actually speak in the outer environment. As Chafe (to appear) notes, "language itself provides evidence that not everything in consciousness is verbal. Disfluencies show that people often experience difficulty in turning thoughts into words, suggesting that there is more to thought itself than inner speech" (p. 16).

[12]The representation of the inner environment is, of course, a simplification and idealization of the "real" inner environment.

[13]The entire volume by Parker, Mitchell and Boccia (1994) is devoted to this topic. The famous experiments by Gallup (1970) only show that the chimpanzees are aware of their own *bodies*, not their own minds. Thus this kind of experiment does not show that animals have self-awareness in the sense discussed here. Terrace (1985, p. 1026) correlates self-awareness with what can be *named* by an individual: "Human beings are able to name their inner states; animals are not."

[14]Terrace (1985, pp. 1015-1016) discusses an experiment by Epstein, Lanza and Skinner which purported to show that pigeons could communicate intentionally in the same way as chimpanzees do. Terrace explains carefully why the experiment shows nothing of the kind. Also see Davis (1989).

[15]For further discussion of this, see Donald (1991) and Jonker (1991).

[16]Already Sapir (1921, p. 10) saw this clearly: "A speech-sound localized in the brain, even when associated with the particular movements of the 'speech organs' that are required to produce it, is very far from being an element of language. It must be further associated with some element or group of elements of experience, say a visual image or a class of visual images or a feeling of relation, before it has even rudimentary linguistic significance."

[17]It is interesting to note that those apes who have been taught sign language never conduct a monologue with themselves, nor is their play accompanied by signing. This is in contrast with human children, where even deaf children who are confined to sign language engage in monologues and sign while playing.

4. SIX KINDS OF COMMUNICATION SYSTEMS

So far I have focused on the representational capacities of human minds that are necessary for linguistic communication and how these capacities may have evolved. The key idea has been that the symbols of language refer to detached representations. Language is, however, more than a collection of symbols. Another dimension is the *grammar* of symbols, i.e., the ways several simple linguistic symbols can be composed to form multipart expressions. Hence, we must consider the *evolutionary function* of the grammatical structure of language.

Let us assume that a community uses a communication system with a certain set of signs, icons or symbols. The communicative capacity of the community would be greatly heightened if the elements from the set could be combined to form *composite* signs, icons or symbols. Let me call such an expanded system a compositional communication system.[18] Full compositionality would result in a *generative* system, in the sense that the users could create new combinations of signs that have not been used before and where only the elements have been learned, but not the combination. Thus, a compositional expansion of a communicative system would clearly increase its evolutionary value. For example, the system of iconic gestures proposed by Donald as the crucial element of the "mimetic" stage of the human evolution is compositional and generative (Donald 1991, pp. 171–172).

Now, a drawback of compositional systems is that expressions composed of several elements very often are ambiguous. To give a trivial example, if somebody says or signs "Grog hit," it may mean that Grog hits something or that Grog is hit by something. One must rely on the *context* of the utterance to disambiguate the expression. Consequently, a communication system where the *form* of expressions helps in disambiguating the meaning could be used to make communication much more *context independent*. I believe that this is the main evolutionary function of *grammar*. There are various means to add structure to the composition of the single "original" elements: non–arbitrary word order, markers on the elements, grammatical elements with no independent meaning, intonation patterns, etc.

In brief, I want to distinguish three levels of communication system with respect to how the elements may be composed.

(i) *Systems with single elements:* The communication system uses only single signs, icons or symbols.
(ii) *Compositional systems:* Two or more signs or symbols from the vocabulary can be combined in a generative way.
(iii) *Systems with grammar:* The composed expressions contain different kinds of grammatical markers and constraints on word order.

Another feature that separates systems with grammar from compositional systems in general is that complex expressions can be *expanded* by adding or embedding words or phrases. For example, to a noun phrase one can add iterated adjectives, subordinate clauses, and prepositional phrases.

In Section 6, I shall argue that the level of grammaticality is, to a large extent, independent of the level of representationality. If we make, on the one hand, a distinction between the three levels of grammatical complexity, and, on the other hand, a distinction between cued and detached representations, we obtain six possible types of communication systems (see Table 1).

In this table, type 5 turns out to be the most interesting one from an evolutionary point of view, so let me comment on the other types first. The first type, where representations are cued and the signaling system has nor grammar, covers most of animal communication. This kind of communication uses only signals, where the coupling between the signal and its representation can be either innate or acquired. It should be noted that signals can be *gestural* as well as vocal.[19]

Type 2, where representations are cued but the signaling system is compositional, is rare. But presumably bees' dances should be classified as being of type 2. They have a limited "vocabulary" of signs that can be compositionally composed according to a simple "grammar," where the signal sequences represent a fairly large class of possible nectar locations.[20]

[18]This sense of "compositional" is different from the rather special meaning it has acquired in the philosophic/linguistic society, namely that the meanings of complex symbolic expressions can be determined as *functions* of the meanings of the single symbols. However, my use of "compositional" is not committed to any thesis about how the meaning of a complex symbolic expression is determined.

[19]Cf. Donald (1991) and Jonker (1991). The role of gestures in communication will be discussed below in Sections 6 and 7.

[20]Also the songs of certain birds, like blackbirds and nightingales, are generative in the sense that they can rather freely compose smaller song elements into longer sequences. However, it seems unlikely that the song elements have any separate meanings.

	Single elements	*Composition*	*Grammar*
Cued representations	Type 1 (Animal signs)	Type 2 (Bee's dances)	Type 3 Ø
Detached representations	Type 4 (One–word language)	Type 5 (Protolanguage)	Type 6 (Full language)

Table 1: Six types of communication systems. The examples of the different types will be explained in the text.

The generativity is, after all, limited, and, as was argued above, the representations are clearly cued. Benveniste (1966, p. 62) says about the mode of communication of the bees that "ce n'est pas un langage, c'est un code de signaux." One of the reasons which he gives for this conclusion relates to Gricean criteria of communication:

> Le message des abeilles n'appelle aucune réponse de l'entourage, sinon une certaine conduite, qui n'est pas une réponse. Cela signifie que les abeilles ne connaissent pas le dialogue, qui est la condition du langage humain. Nous parlons à d'autres qui parlent, telle est la réalité humaine. Cela révèle un noveau contraste. Parce qu'il n'y a pas dialogue pour les abeilles, la communication se réfère seulement à une certaine donnée objective. Il ne peut y avoir de communication relative à une donnée "linguistique"; déjà parce qu'il n'y a pas de réponse, la réponse étant une réaction linguistique à une manifestation linguistique; mais aussi en ce sens que le message d'une abeille ne peut être reproduit par une autre qui n'aurait pas vu elle-même les choses que la première annonce (Benveniste 1966, pp. 60–61).

Benveniste's distinction between "objective" and "linguistic" communication is, in my opinion, better expressed by the difference between signals and symbols, i.e., the distinction between communication using cued vs. detached representations.

I have not been able to diagnose any system of type 3. One reason for why such systems will not be found in a natural setting is that, on my analysis, the function of grammar is to make composed sequences less ambiguous and thus more context independent. However, signals refer to cued representations, according to the characterization in Section 2, and making them context independent would thus be pointless.

Let us then turn to the second row of Table 1, i.e., to communication systems based on detached representations. Here one could make a finer classification by distinguishing between iconic and symbolic representations (see e.g., Tomasello 1991). The simplest case of a system of type 4 would be a communication system based on single icons or symbols, i.e., a "one-word" language. Donald (1991, ch. 6) argues that the first major transition from primate cognition on the way to the modern human results in what he calls the *mimetic* culture. Mimetic communication involves, above all, iconic gestures and sounds. Donald furnishes different kinds of evidence to convincingly establish that the mimetic stage has existed in the human evolution, and he associates it with the era of *Homo erectus*. Mimetic communication started as a system of type 4, but may have evolved into a compositional system of type 5. However, the communication has presumably used mainly iconic signs and not arbitrary symbols.

Type 6, with both detached representations and grammatical structure, comprises full natural language. It should be noted that it is not necessary that the building blocks of such a language are arbitrary symbols, but it can be based on iconic representations, prime examples of this being the sign languages of the deaf.

A challenging question is whether type 6 contains any other kind of communication system. Perhaps there are natural systems that are grammatical and have detached representations, but which still do not satisfy all the Gricean conditions described above. Bennett (1976, pp. 171–175) introduces a set of weaker Gricean conditions and presents a story where the communication of an imaginary tribe satisfies these conditions but not the full Gricean conditions. Bennett's story is intriguing and seems fully possible, and a communication system of this kind may even have been an evolutionary precursor to our present kind of language. But there still remains a big leap between communication systems of the fourth type and those of the sixth.

5. PROTOLANGUAGE

This is where type 5 systems become relevant. In these, representations are detached and symbols are combined generatively, but with no grammatical

elements. Are there such communication systems? Have they existed in man's evolutionary history?

An interesting proposal is made by Bickerton (1990). He suggests that there are *protolanguages*, which, in my classification, would be of this fifth type. Bickerton claims that there are at least four different kinds of present day communication systems that can be classified as protolanguages. The first is the language of children under two. Here are some examples of some typical kinds of utterances from this stage (which normally occurs at an age between 18 and 24 months):[21]

> Big train; Red book *(attribution of qualities to objects)*
> Adam checker; Mommy lunch *(possessive relations)*
> Walk street; Go store *(location of actions)*
> Adam put; Eve read *(relation of agents to actions)*
> Put book; Hit ball *(relation of actions to patients)*

This kind of communicative system clearly depends on detached representations – the meanings of the individual words are well established in the mind of the child. However, the system of phrases is not grammatical – it only contains two-word expressions.

The second kind of evidence can be obtained from primate studies. It seems that the two-word stage of symbolic communication is the best that can be achieved by primates other than humans. The following are some examples of utterances of the chimpanzee Washoe, which functionally seem to match those of the child above (from Bickerton 1991, p. 114):

> Drink red; Comb black
> Clothes Mrs. G.; You hat
> Go in; Look out
> Roger tickle; You drink
> Tickle Washoe; Open blanket

The third piece of evidence is provided by "closet children," i.e., children who have been deprived of exposure to language during the crucial period of their lives. Bickerton (1990, pp. 114–118) recounts the story of Genie who had been imprisoned in her room until she was thirteen. She had normal intelligence, but she never fully acquired language, despite great efforts to teach her. She remained stuck at about the level of a two–year old.

The fourth kind of communication of the same style is first generation pidgin languages. In situations where speakers of unrelated languages come in contact, a rudimentary form of language develops which shows the same features as the previous examples. After the first generation, it may be acquired by locally born children and become a fully developed creole language. However, if the pidgin language is used only in, for example, sporadic trading contacts, it may remain on the same primitive level for many generations.[22]

On the basis of this evidence, Bickerton (1990, p. 122) concludes that

> there is a mode of linguistic expression that is quite separate from normal human language and is shared by four classes of speakers: trained apes, children under two, adults who have been deprived of language in their early years, and speakers of pidgin.

Bickerton never defines what constitutes protolanguage, but only characterizes it negatively by comparing it to ordinary language. He presents five types of differences: Protolanguage is less ordered than ordinary language, it contains no null elements,[23] it does not always respect the valence relations of verbs, it does not allow expansion of utterances, and it hardly contains any grammatical items (Bickerton 1990, pp. 122–126). The general lack of grammaticality means that protolanguage is a communication system of type 5 in my classification above.

If protolanguage is accepted as a special kind of communicative system, a very natural hypothesis concerning the evolution of language is that it has developed from a signaling system of type 1, via a one-word language of type 4 and a protolanguage of type 5, to a full language of type 6. As regards the timing of the transition from protolanguage to language, several authors (Bickerton 1990, Donald 1991, Fidelholz 1991, Deacon 1992, Lieberman 1992) speculate that this is essentially concurrent with the transition from *Homo erectus* to *Homo sapiens* (on the order of 200.000 years ago).[24] One of the anatomical changes that occurs in this transition between the two species is the lowering of the

[21]Taken from Bickerton (1990, p. 114), who attributes the examples, without reference, to Beatrice Gardner.

[22]Bickerton (1990, pp. 121–122) gives the example of Russonorsk which developed in contacts between Russian and Norwegian sailors.

[23]"Null element" is a syntactic notion refering to places in a sentence where one can infer (using government and binding theory) that some constituent should be present, but where there is no explicit constituent.

[24]Bickerton (1990, p. 174) even argues that the transition from protolanguage to language is *catastrophic*, in the sense that it originates from a mutation in a single individual. However, the evidence he presents for this thesis is, in my opinion, quite weak. Barber and Peters (1992, pp. 343–344) presents an interesting story about a *gradual* development of grammar among early members of *Homo sapiens.*

larynx, which clearly is connected with the development of a spoken language (Lieberman 1992).

6. THE INDEPENDENCE OF SPEECH AND THOUGHT

Returning to the two dimensions of communication systems proposed above, Vygotsky had as early as 1934 the idea the these dimensions are independent (although he makes no distinction between cued and detached representations):

> The preintellectual roots of speech in child development have long been known. The child's babbling, crying, even his first words, are quite clearly stages of speech development that have nothing to do with the development of thinking. These manifestations have generally been regarded as a predominantly emotional form of behavior. Not all of them, however, serve merely the function of release. Recent investigations of the earliest forms of behavior in the child and of the child's first reactions to the human voice ... have shown that the social function of speech is already clearly apparent during the first year, i.e., in the preintellectual stage of development. (Vygotsky 1986, p. 81)
>
> In brief, we must conclude that:
> 1. In their ontogenetic development, thought and speech have different roots.
> 2. In the speech development of the child, we can with certainty establish a preintellectual stage, and in his thought development a prelinguistic stage.
> 3. Up to a certain point in time, the two follow different lines, independently of each other.
> 4. At a certain point these lines meet, whereupon thought becomes verbal, and speech rational. (Vygotsky 1986, p. 83)

The independence of speech from thinking is consistent with the fact that Genie never reached full linguistic capacities, although her thinking was quite normal. Her speech development was hampered at the crucial age and could never be fully regained. Barber and Peters (1992, p. 328) conclude: "It therefore looks as though the acquisition of the *meaningful* parts of language (vocabulary and semantics) are dependent on Cognition, whereas development of the *grammatical system* of Language is relatively independent of it." Vygotsky's thesis that the primary role of speech is emotional and social also receives support from a recent article by Dunbar (to appear) who argues that "language evolved as a 'cheap' form of social grooming, so enabling the ancestral humans to maintain the cohesion of the unusually large groups." The disadvantages of ordinary grooming among apes as a social glue are that it is impossible to do anything else while grooming and only one individual can be groomed at a time. Language overcomes both these limitations. If this social view on the emergence of language is correct, the representational uses of language are evolutionarily later and may have begun as supervenient on the social communication.

So whence grammar? As the human societies grew more complex, speed and efficiency in communication was rewarded. Barber and Peters (1992, p. 311) argue that[25]

> the need for fast and efficient processing is thus a major force that drives language away from iconicity and toward systematicity – and this in turn drives language toward arbitrariness […], for the following reason. In the long run it is less effort to deal with a tightly patterned system with a small number of reusable parts (both items and rules) than to deal with a sprawling system with many, many unique parts. But the reduction to reusable parts and patterns destroys most of the iconicity […], and at the same time compresses a great deal more information into a small number of rules: it radically increases the "depth" of the system while decreasing the algorithmic complexity […]."

A similar point is made by Savage-Rumbaugh and Rumbaugh (1993, pp. 86–87), who note the need for communication that is *independent of context* as one of the evolutionary forces behind the development of grammar:

> It will also be argued that syntax, rather than being biologically predetermined, is a skill which arises naturally from the need to process sequences of words rapidly. As overall intelligence increased, spurred by the ever-increasing use of language for planning future activities, communications became increasingly complex and increasingly independent of context. When complex ideas began to require groups of words for their expression, it became essential to devise a means to specify which of the words in a group modified (or were related to) which other words. Syntactical rules were developed to

[25] Ellegård (1977, p. 142) speaks about the "double articulation" of language and remarks concerning the evolution of grammar: "My hypothesis is thus that the double articulation of human speech emerged as a necessary consequence of the increasing number of signs, and the increasing demands for fast and more or less automatized production. The reaction of the brain toward these demands was the double articulation in phonemes and morphemes." (My translation).

> solve this dilemma. Such rules were the inevitable outgrowth of complex symbolic communication involving multiple symbols.

Adding grammar to a communication system thus increases its efficiency. For a neuroscientist, the question is in what way the brain must change in order to achieve this capacity. It seems that the grammaticality of language has probably not evolved as an independent cognitive ability (in contrast to Chomsky's claims concerning a "language acquisition device"). Rather, it could build on already existing structures since it seems to be tied to a more general capacity of combining actions into sequences. Neurologically, sequencing is typically lateralized to the left hemisphere in humans. Corballis (1989) argues that in the course of evolution, sequencing emerged in the left hemisphere and was essential for *tool-making* and other practical skills.[26] In order to reproduce or create a tool, a sequence of actions had to be performed, and old elements of action sequences must be recombined to produce something new. The practical, mainly manual, ability forms the basis for all kinds of sequencing and was extended to sequencing of symbols, which then resulted in a grammatical language. Also *playing* involves sequences of motor actions performed on symbolically used objects (see Vauclair and Vidal 1994). And remember that Piaget has always emphasized that play and imitation are cardinal for the development of symbolic capacities. Apes seem to lack the sequencing capacity, which could explain why they are bad at imitating action sequences, why they never invent new plays, as well as why they cannot learn more than a protolanguage.[27]

Similarly, Kimura (1976) argues that, clinically, aphasia (language loss) is often correlated with apraxia (loss of complex movements) and results from lesions to the left association cortex. The left hemisphere seems to be specialized for serial motor control, for manual sequences as well as vocal. As a matter of fact, in signing disorders in the deaf, the breakdown can be traced to damages to the speech areas on the left half of the cortex. Gesturing may, indeed, be a stage that developed as a form of iconic communication long before vocal communication became generative.[28] As mentioned above, Donald (1991) argues that the mimetic ability was what gave *Homo erectus* a decisive advantage over earlier forms in the history of human evolution. Kimura (1976) concludes that the left hemisphere is "well adapted, not for symbolic function per se, but for the execution of some categories of motor activity which happened to lend themselves readily to communication."

7. CONCLUSION

Summing up, I envisage the following steps in the evolution of language and the underlying cognitive faculties. Starting from a primitive animal signaling system of type 1, a crucial step was the evolution of an inner environment which allows for iconic and symbolic reference. The first detached communication systems were presumably one-word languages of type 4. Gestural communication using icons preceded vocal language using symbols.[29] Thus, I believe that Donald's (1991) *mimetic stage* is an important step in the evolution of language. Vocal sounds gradually changed from primarily instinctive emotional/social signals to voluntary communicative symbols. Gesturing and primitive speech developed into a protolanguage, i.e., a communication system of type 5, which allows for some combinations of symbols, but not full grammaticality.

The survival benefits of fast and efficient communications then resulted in the development of arbitrary symbols and syntactic rules, at the cost of

[26] See also Allott (1991) and Tomasello (1991).

[27] See also Barber and Peters (1992, p. 344) and Donald (1991, pp. 70–75). On the other hand, Bickerton (1990, p. 139) argues that "it is tool-making and *protolanguage* that share the same processes."

[28] Lyons (1988, p. 159) adds the following argument: "Iconicity, of which onomatopoiea in spoken languages is the most obvious example, is generally regarded as one of the 'design-features' which separates non-linguistic, or pre-linguistic, systems of communication from fully fledged languages. But iconicity, more generally defined as non-arbitrariness of the association of form and meaning, is […] not a matter of yes or no, but of more or less; and there is much more iconicity in 'ordinary' natural languages, at all levels of their structure, than the conventional wisdom in linguistics would have us believe." Ellegård (1979) distinguishes four stages in the development of language: the pointing stage, the refering stage, the telling stage, and the depicting stage (my translation). In my classification, the pointing stage corresponds to a signaling system of type 1 except that it involves pointing which is an advanced form of signal; and the refering stage corresponds to a system of type 4, with detached representations using either icons or symbols. The telling stage and the depicting stage both seem to be of type 6, the difference being that the telling stage is only used to express *actual* events in order to communicate them to others, while in the depicting stage the story may be totally detached from reality.

[29] Compare Lyons (1988, p.160): "On the basis of this and other evidence, including the fact that gesture continues to play an important 'paralinguistic' role in the modulation and punctuation of normal spoken utterances, it is argued that languages, as we now know them in their fully developed form, may have developed, whether by relatively slow evolution or catastrophically, between 100000 and 40000 years ago, not as a direct outgrowth of the expressive, or emotive, use of vocal signals characteristic of non-human primates, but of a pre-existing system of manual gestures […]." See also Stokoe (1991) who argues that even in cultures that exists today, for example in many aboriginal tribes of Australia, a sign language functions in parallel with a spoken language. Thus the role of sign languages among the non-deaf seems to be underestimated.

iconicity, to arrive at a language with a full grammar, i.e., a communication system of type 6. In order to successfully establish linguistic conventions and to have real communication, the inner environment has to contain models of other agents' inner environments, i.e., a you-awareness, as well as a model of one's own inner environment, i.e., an I-awareness. Perhaps a phase in between protolanguage and full language was Bennett's (1976) sub–Gricean tribal communication where the nesting of belief systems still had not reached its present complexity.

The final point to be made is that the steps in the evolution of language that I have outlined here are not to be seen as replacing one another, but rather as adding new features to an already existing communication system. Thus I follow Lyons (1988, p. 156) in subscribing to "the hypothesis that human language is a multi-layered or multi-stranded phenomenon, each of whose layers or strands may be of different antiquity and of different origin."

ACKNOWLEDGEMENTS

Research for this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences. I want to thank Lukas Böök, Paul Bourgine, Ingar Brinck, Henrik Gedenryd, Pierre Livet, Kevin Mulligan, Joëlle Proust and Simon Winter for helpful discussions. The material in Sections 1 and 2 is also discussed in Gärdenfors (to appear a, to appear b).

REFERENCES

Allott, R. (1991): "The motor theory of language," in *Studies in Language Origins, Vol 2*, Raffler-Engel, W. von, Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 123–157.

Balkenius, C. (1995): *Natural Intelligence in Artificial Creatures*, Lund University Cognitive Studies 37, Lund.

Barber, E. J. W. and Peters, A. M. W. (1992): "Ontogeny and phylogeny: What child language and archeology have to say to each other," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 305–351.

Bennett, J. (1976): *Linguistic Behaviour*, Cambridge University Press, Cambridge.

Benveniste, É. (1966): *Problèmes de Linguistique Générale*, vol 1, Gallimard, Paris.

Bickerton, D. (1990): *Language and Species*, The University of Chicago Press, Chicago.

Bråten, S. (1988): "The virtual other," in *Between Rationality and Cognition*, M. Campanella, ed., Albert Meynier, pp. 205–235.

Chafe, W. L. (to appear): "Accessing the mind through language" to appear in *Of Thoughts and Words*, ed. by S. Allén, Proceedings of the Nobel Symposium on the Relation between Language and Mind, Stockholm, August, 1994.

Corballis, M. C. (1989: "Laterality and human evolution," *Psychological Review 96*, pp. 492–505.

Craik, K. (1943): *The Nature of Explanation*, Cambridge University Press, Cambridge.

Davenport, R. K. (1976): "Cross–modal perception in apes," in S. R. Harnad, H. D. Steklis, and J. Lancaster, eds., *Origins and Evolution of Language and Speech*, Annals of the New York Academy of Science 280, pp. 143–149.

Davis, L. H. (1989): "Selfconsciousness in chimps and pigeons," *Philosophical Psychology 2*, pp. 249–259.

Deacon, T. W. (1992): "Brain-language coevolution," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 49–83.

Dennett, D. (1978): *Brainstorms: Philosophical Essays on Mind and Psychology*, The MIT Press, Cambridge, MA.

Dennett, D. (1991): *Consciousness Explained*, Little, Brown and Company. Boston, MA.

Donald, M. (1991): *Origins of the Modern Mind*, Harvard University Press, Cambridge, MA.

Dunbar, R. I. M. (to appear): "Co-evolution of neocortex size, group size and language in humans," to appear in *Behavioral and Brain Sciences.*

Ellegård, A. (1979): "Om det mänskliga språkets ursprung," in *Kungliga Vitterhets-, Historie- och Antikvitetsakademins Årsbok*, Stockholm, pp. 131–147.

Fidelholz, J. L. (1991): "On dating the origin of the modern form of language," in *Studies in Language Origins, Vol 2*, von Raffler-Engel, W., Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 99–113.

Gallup, G. G. Jr. (1970): "Chimpanzees: Self-recognition," *Science 167*, 86–87.

Gärdenfors, P. (1992): "Medvetandets evolution," Chapter 5 in *Blotta tanken*, Doxa, Nora.

Gärdenfors, P. (1993): "The emergence of meaning," *Linguistics and Philosophy 16*, pp. 285–309.

Gärdenfors, P. (to appear, a): "Speaking about the inner environment," to appear in *Of Thoughts and Words*, ed. by Sture Allén, Proceedings of the Nobel symposium on *Mind and language*, Stockholm, August 1994.

Gärdenfors, P. (to appear, b): "Cued and detached representations in animal cognition," paper presented at the conference on *Cognition and Evolution*, Berder, March 1994, to appear in *Behavioural Processes.*

Gergely, G. (1994): "From self-recognition to theory of mind," in *Self-awareness in Animals and Humans*, S. T Parker, R. W. Mitchell, and M. L. Boccia, eds., Cambridge University Press, Cambridge, pp. 51–60.

Glasersfeld, E. von (1976): "The development of language as purposive behavior," in *Origins and Evolution of Language and Speech*, S. R. Harnad, H. D. Steklis, and J. Lancaster, eds., Annals of the New York Academy of Science 280, pp. 212–226.

Glasersfeld, E. von (1977): "Linguistic communication: Theory and definition," in D. M. Rumbaugh, ed., *Language Learning by a Chimpanzee*, Academic Press, New York, pp. 55–71.

Gomez, J. C. (1994): "Mutual awareness in primate communication: A Gricean approach," in *Self-awareness in Animals and Humans*, S. T. Parker, R. W. Mitchell, and M. L. Boccia, eds., Cambridge University Press, Cambridge, pp. 61–80.

Gopnik, A. (1993): "How we know our minds: The illusion of first-person knowledge of intentionality," *Behavioral and Brain Sciences 16*, pp. 1–14.

Grice, H. P. (1957): "Meaning," *The Philosophical Review 66*, pp. 377–388.

Grice, H. P. (1969): "Utterer's meaning and intentions," *The Philosophical Review 78*, pp. 147–177.

Gulz, A. (1991): *The Planning of Action as a Cognitive and Biological Phenomenon*, Lund University Cognitive Studies 2, Lund.

Hockett (1960): "Logical consideration in the study of animal communication," in *Animal sounds and Communication*, W. E. Lanyon and W. N. Tavolga, eds., American Institute of Biological Science, Washington, DC.

Jeannerod, M. (1994): "The representing brain: Neural correlates of motor intention and imagery," *Behavioral and Brain Sciences 17*, pp. 187–202.

Jonker, A. (1991): "On the origins of language and self–consciousness," in *Studies in Language Origins, Vol 2*, Raffler-Engel, W. von, Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 41–50.

Kimura, D. (1976): "The neurological basis of language qua gestures," in *Current Trends in Neurolinguistics*, H. Whitaker and H. A. Whitaker, eds., Academic Press, New York.

Langer, S. (1948): *Philosophy in a New Key*, Penguin Books, New York.

Lieberman, P. (1992): "On the evolution of human language," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 21–47.

Lyons, J. (1988): "Origins of language," in *Origins: The Darwin College Lectures*, A. C. Fabian, ed., Cambridge University Press, Cambridge, pp. 141–166.

Mead, G. H. (1934): *Mind, Self, and Society*, University of Chicago Press, Chicago.

Murray, E. A (1990): "Representational memory in nonhuman primates," in *Neurobiology of Comparative Cognition*, R. P. Kesner and D. S. Olton, eds., Lawrence Erlbaum Associates, Hilldale, NJ, pp. 127–155.

Parker, S. T., Mitchell, R. W., and Boccia, M. L., eds. (1994): *Self-awareness in Animals and Humans*, Cambridge University Press, Cambridge.

Peirce, C. S. (1932): *Collected Papers of Charles S. Peirce*, ed. Hartshorne, Weiss and Burks, Harvard University Press, Cambridge, MA.

Sapir, E. (1921): *Language*, Harcourt, New York, NY.

Saussure, F. de (1984): *Cours de Linguistique Générale*, T. de Mauro, ed., Paris.

Savage-Rumbaugh, E. S. and Rumbaugh, D. M. (1993): "The emergence of language," in *Tools, Language and Cognition in Human Evolution*, K. R. Gibson and T. Ingold, eds., Cambridge University Press, Cambridge, pp. 86–108.

Sjölander, S. (1993): "Some cognitive breakthroughs in the evolution of cognition and consciousness, and their impact on the biology of language," *Evolution and Cognition 3*, pp. 1–10.

Stein, B. E. and Meredith, M. A. (1993): *The Merging of the Senses*, MIT Press, Cambridge, MA.

Terrace, H. S. (1985): "In the beginning was the name," *American Psychologist 40*, pp. 1011–1028.

Tolman, E. C. (1948): "Cognitive maps in rats and men," *Psychological Review 55*, pp. 189–208.

Tomasello, M. (1991): "Processes of communication in the origins of language," in *Studies in Language Origins, Vol 2*, Raffler-Engel, W. von, Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 85–97

Vauclair, J. and Vidal, J.-M. (to appear): "Discontinuities in the mind between animals and humans," paper presented at the conference on *Cognition and Evolution*, Berder, March 1994.

Vygotsky, L. (1986): *Thought and Language*, The MIT Press, Cambridge, MA.

Whiten, A. and Byrne, R. W. (1988): "Tactical deception in primates," *Behavioral and Brain Sciences 11*, pp. 233–73.