

# Evolution, communication, and the proper function of language

Gloria Origgi, Dan Sperber

► **To cite this version:**

Gloria Origgi, Dan Sperber. Evolution, communication, and the proper function of language. Peter Carruthers, Andrew Chamberlain. Evolution and the Human Mind: Language, Modularity and Social Cognition, pp.140-169, 2000. ijn\_00000015

**HAL Id: ijn\_00000015**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_00000015](https://jeannicod.ccsd.cnrs.fr/ijn_00000015)**

Submitted on 10 Jun 2002

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Gloria Origgi**                      &                      **Dan Sperber**  
Department of communication                      CNRS, Paris  
University of Bologna

**EVOLUTION, COMMUNICATION AND THE PROPER FUNCTION OF  
LANGUAGE\***

**(A discussion of Millikan in the light of pragmatics and of the psychology of  
mindreading)**

(In Peter Carruthers and Andrew Chamberlain (eds.) *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge University Press. 140-169)

\* We thank Peter Carruthers, Andrew Chamberlain, Ruth Millikan, and Deirdre Wilson for their most useful comments on earlier drafts of this chapter.

---

Language is both a biological and a cultural phenomenon. Our aim here is to discuss, in an evolutionary perspective, the articulation of these two aspects of language. For this, we draw on the general conceptual framework developed by Ruth Millikan (1984) while at the same time dissociating ourselves from her view of language.

**Biological and cultural evolutionary processes**

The phrase "evolution of language" refers to two related but quite distinct processes: the biological evolution of a language faculty, and the historical-cultural evolution of languages. The historical-cultural evolution of languages itself requires the repetition across populations and over generations of the individual process of language acquisition. Individuals who have acquired the language of their community can engage in verbal communication. Through a myriad of acts of communication, they achieve a variety of effects, intended or unintended. The aggregation of these effects explains both the biological evolution of the language faculty, and the historical-cultural evolution of languages.

The biological evolution of a language faculty and the historical-cultural evolution of languages are related in interesting ways. If we assume, with Chomsky, that human languages require, to be acquired, a language faculty, it follows that the biological emergence of this

faculty is a precondition for the cultural emergence of any human language. On the other hand, if we think, without Chomsky this time, of the language faculty as a biological adaptation, then, presumably its function - at least its proximate function on the successful performance of which other functions depend - is to make language acquisition possible. A language faculty is adaptive only in an environment where languages are spoken and where, therefore, inputs indispensable for language acquisition are found. Adaptations qua adaptations emerge only in an environment where they are adaptive. So it seems that the existence of a spoken language is a precondition for that of a language faculty. But then, the language faculty and a spoken language are each a precondition for the other. There are various ways to finesse this bootstrapping problem. We will conclude this paper by proposing a possible way to resolve it.

Even if the proximate function of the language faculty is to permit the acquisition of language, what makes *this* adaptive is the adaptive value of language use itself. In fact, most adaptationist explanations of the biological evolution of the language faculty just take for granted or ignore its obvious proximate function, that of permitting language acquisition. They explain the emergence and stabilisation of the language faculty by the adaptive value of language use, that is, in terms of quite remote functions of the language faculty itself.

How, then, does language use contribute to biological fitness? Language use consists in the expression and communication of thoughts. Expression without communication, as when we think in words, may be adaptive because of its contribution to cognitive performance (Bickerton, 1995, chapter 3; Carruthers, 1996, Chomsky 1980: 229-230). We will not consider this (possibly important) aspect of the adaptiveness of language use in this discussion. The adaptive value of public languages is, we assume, mostly due to their use in communication. But what makes communication itself adaptive? Communication has a great variety of effects. It allows individuals to benefit from the perceptions and inferences of others and increases their knowledge well beyond that which they could acquire on their own. It allows elaborate forms of co-ordinated planning and action. It can be used for manipulation, deceit, display of wit, seduction, maintenance of social relationships, all of which have fitness consequences.

Many of the debates on the biological emergence and evolution of the language faculty revolve solely around the relative importance of these diverse functions of linguistic communication (e.g. Dunbar, 1996; Hurford et al. 1998, part I). Even accepting the implicit move from language faculty to language use, there is something missing here. It is as if the evolution of organs of locomotion such as wings or legs were discussed only in terms of the effects of locomotion such as fleeing predators, finding food, or finding mates, without considering the proximate function of these organs, namely locomotion itself. Different organs of locomotion determine qualitatively and quantitatively different performances. The evolution of these organs cannot be properly understood without taking into consideration these specific performances, i.e. the ways in which these organs perform their proximate function of locomotion. In much of the literature, verbal communication is treated as a well-understood process, the character of which can be taken for granted rather than examined when discussing evolutionary issues. This, however, is an illusion. The mechanism of verbal communication is contentious. Different views of this mechanism have different evolutionary implications, and this is one of the two main issues we want to investigate here.

A language faculty is an adaptation because it permits the acquisition of linguistic competence, which permits verbal communication, which can be used in a great variety of

ways, some with beneficial effects. Identifying those remote effects of the language faculty that have contributed to the biological fitness of language users should provide some essential pieces of the overall puzzle. However, this is unlikely to help much with the specifics and the articulation of the two evolutionary processes involved: the biological evolution of a language faculty and the cultural evolution of languages. The proper way of describing this articulation is the second main issue we want to discuss here.

There have been, in the past twenty years or so, interesting discussions of the relation between biological and cultural evolution. On the one hand processes of gene-culture co-evolution have been hypothesised. It is reasonable to surmise that solving the bootstrapping problem we mentioned at the outset will involve modelling such a co-evolutionary process between languages and the language faculty. On the other hand, various conceptual frameworks for dealing in a unified manner with both biological and cultural evolution have been proposed (Boyd & Richerson 1985; Cavalli-Sforza & Feldman 1981; Dawkins 1976, 1982; Dennett 1995; Durham 1991; Lumsden & Wilson 1981; Millikan 1984; Sperber 1996). The best known is probably Dawkins's. The conditions for undergoing Darwinian selection, Dawkins argues, can be fulfilled not only by biological replicators such as genes, but also by artefacts such as computer viruses, or by bits of culture that get copied again and again and which he calls "memes". If one accepts this framework, then languages, or at least linguistic devices such as words or grammatical forms, can be seen as paradigmatic examples of memes. There are problems, however, with the meme framework: either the Darwinian model of selection is applied as is to cultural evolution, and this is too rigid, as many including Dawkins himself have noted. Or else the meme framework must be loosened, but it is unclear how this should be done, and to what extent the explanatory power of the approach might survive such loosening (see Sperber 1996, chapter 5).

In this presentation, we will consider another conceptual framework, more familiar to philosophers than to evolutionary theorists, that of Ruth Millikan. It is intended from the start to approach biological and cultural phenomena in the same basic way, and it is, in this respect at least, both more precise and less rigid than Dawkins's. Moreover, in her book *Language, Thought and Other Biological Categories* (1984), Millikan uses this framework to discuss in detail the case of language. Her not-at-all-hidden agenda, in so doing, is to debunk a view of verbal communication defended in particular by Paul Grice (1957, 1989) that has gained not universal but wide acceptance among philosophers of language and linguists. According to the Gricean view Millikan attacks, comprehension systematically involves identifying an intention of the speaker. According to the view Millikan defends, comprehension typically consists in coming directly to believe what is being asserted or in coming directly to want to comply with what is being requested. Here, we will articulate our discussion of the evolution of language around Millikan's proposals. Specifically, we will attempt to pry apart her conceptual framework, which we find well worth exploring further, in particular in reflecting about the case of language, from her own view of language, with which we quite disagree.

### **Millikan's teleofunctional framework**

In her 1984 book, *Language, Thought and Other Biological Categories*, Ruth Millikan presented a general account of biological and cultural items in terms of "proper functions" historically responsible for the reproduction and proliferation of these items. She proposed a particularly interesting distinction between direct and derived proper functions. (We will ignore several other conceptual distinctions that she introduced in this book and made little use of thereafter. Generally, our goal is not to present a critical exegesis of Millikan work, but

to reconcile aspects of her basic approach with a view of language she opposes.) Millikan's theory of proper functions is a way of explaining different cases of reproduction, in particular biological and cultural, within a single framework. Linguistic devices, purposive behaviours, artefacts, and body organs provide examples of such cases.

What is a proper function? Quite standardly, Millikan distinguishes the proper function of an item from its actual effects, that is, what in fact it succeeds in doing on various occasions, and from the functions that various users intend it to perform on various occasions. She then defines, quite originally, not one but two types of proper function: direct and derived. For an item *A* to have a *direct* proper function *F* it has to fulfil the following condition:

***Direct Proper Function:*** "A originated as a "reproduction" [...] of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed *F* in the past, and *A* exists because (causally, historically because) of this or these performances" (Millikan, 1993 : 12).

An item may typically have a great many recurring effects: its *direct* proper function is the one that is *historically* responsible for its reproduction. A heart makes noise, contributes to the body's weight, and pumps blood. Only the latter effect is its proper function. Even a malfunctioning heart still has the direct proper function to pump blood, because it has been reproduced through organisms that, thanks in part to their own heart pumping blood, have had descendants similarly endowed with blood-pumping hearts. Millikan's notion of direct proper function is a rendering of the biological notion of function as used in evolutionary theory, but without any reference to the particular conditions of biological reproduction and selection. As we will see, it applies equally well to an item such as a word.

A device having a direct proper function may perform it by producing items that are adapted to specific environmental circumstances. For instance the pigment-arranging device of the chameleon's skin performs its functions of hiding the chameleon by producing colour patterns matching the background on which the animal is sitting. When the chameleon is sitting on a matching surface, the function of hiding it from predators is performed by a particular colour pattern. It is reasonable to say that this pattern, though it may never have been produced before, has a proper function. However, this should be described not a direct, but a derived proper function. For an item *A* to have a *derived* proper function *F* it has to fulfil the following condition:

***Derived Proper Function:*** "A originated as the product of some prior device that, given its circumstances, had performance of *F* as proper function, and that, under those circumstances, normally causes *F* to be performed by *means* of producing an item like *A*" (*ibid.*).

Whereas, by definition, a direct proper function is performed by a great many items with the same causally relevant properties, a derived proper function may be performed by individual items that have each different causally relevant properties. An item with a derived proper function is one that has been produced by a device (for instance, the chameleon's pigment-arranging device) that produces different items in different contexts (for instance different colour patterns depending on the surface on which the chameleon happens to be sitting). To take another example, the gosling's imprinting mechanism has the *direct* proper function of allowing each and every gosling to fix an image of its mother so as to follow her. The fixation by gosling George of an image of his mother Samantha is a product of this imprinting

mechanism in the special circumstances of George's birth. This particular imprinting, unique to George has the *derived* proper function of helping George follow Samantha.

Note that the derived proper function of a given item can be given two descriptions, one general, the other specific. A general description is without reference to the particulars of the case. For instance any particular colour pattern on the skin on a chameleon has the derived function of making it less visible on the surface on which it is sitting; any imprinting in the brain of a gosling has the derived function of helping it follow its mother. A specific description refers to the particulars of the case and may be different for each item. For instance *this* pattern has the derived function of making *this* chameleon less visible on *this* surface on which it is sitting; the imprinting in George's brain has the function of helping him follow Samantha. Under its general description a derived proper function is one typically shared by many items. Under its specific description, a derived proper function may be a one time affair: the particular colour pattern of chameleon sitting on an improbable background may occur only once in the history of the species, and therefore the derived proper function of hiding this chameleon on this background may be a function that, under this description, is performed only once.

Roughly, the distinction between direct and derived proper functions explains respectively how an item stabilises due to the function its ancestors have performed, and how a new particular item, not reproduced from any ancestral model, is nevertheless generated to perform a proper function - though an indirect one.

### **Millikan, language and communication**

Culture is comprised of all items that are reproduced and proliferated through communication in the widest sense, including unintentional transmission of information (for a more elaborate characterisation of culture, see Sperber 1996). The direct cultural function of a cultural item is, unproblematically, the effect that prior items of the same type have performed in the past and that have caused the item to be reproduced again and again. For instance, a hammer, even if it is actually used as a paperweight, has the direct cultural function of helping to drive nails, because it is the repeated and successful performance of this effect (helping to drive nails) by hammers that has caused them to be produced again and again.

Linguistic items are cultural items, and it is sensible to ask what direct proper functions of a cultural kind they have. In Millikan's terminology, language is a complex of different devices. A "linguistic device" can be a word, a surface syntactic form, a tonal inflection, a stress pattern, a system of punctuation and "any other significant surface elements that a natural spoken or written language may contain" (Millikan 1984:3). A linguistic device has proliferated because it has served a describable, stable proper function.

Language use is a purposeful activity that needs some regularity for its successful performance. More specifically, there must be a regular pattern of correspondence between a speaker's purpose in uttering a given language device and the hearer's response to this utterance. It is this reliability that accounts for the device being used again and again. Among the effects that may be correlated with a linguistic device, its direct proper function is what keeps speakers and hearers using and responding to the linguistic device in a standard way and therefore stabilises the device. What is often called the "conventional use" of a linguistic device corresponds to this *stabilising direct proper function*. Thus the stabilising direct proper

function of a given word is to contribute its conventional meaning to the meaning of the utterances in which it is used.

The use of a given linguistic device on a given occasion, by a speaker with his or her own purposes, endows this token of the device with a derived proper function. This derived proper function may be a mere tokening without modification of its direct proper function (as when a word is used to convey just its conventional meaning) or it may be different from its direct proper function (as in the case of an indexical, or of a non-conventional metaphor). For example, at first blush (and we will propose a different account later), the indexical "now" has the stabilising direct proper function of referring to the time of utterance, and this direct function is performed through each token of "now" performing the derived proper function of referring to a specific time.

Though Millikan does not develop this, linguistic devices also have derived proper functions of a biological kind. A word, say the English word "now", has both public tokens (one every time it is uttered) and mental tokens. The mental tokens themselves are at two levels. There is a mental token each time the word "now" is uttered or comprehended, i.e. a mental representation of the uttered word. There is also, at a more fundamental level, in all individuals capable of using the word "now", an entry for "now" in the mental lexicon which is part of their knowledge of the English language. This mental lexical entry is a mental version of the public language word. It is a cultural item, with a cultural direct proper function. At the same time, it is a device produced by the individual's language faculty or Language Acquisition Device performing its direct function in the particular environment of an English speaking community. The direct biological function of acquiring a language is performed by producing mental devices adapted to the local language community. Therefore mental "now" in a person's mental lexicon (and all the mental linguistic devices of English or of any other language) have biological derived proper functions, just as does George the gosling's imprinting of his mother Samantha's image. The difference is that the gosling's imprinting mechanism fulfils its direct biological function by producing a single item with a derived biological function and no cultural function at all, whereas the Language Acquisition Device fulfils its function by producing tens of thousands of items with derived biological functions and direct cultural functions.

Here then, thanks to the notion of derived function, is a way of describing linguistic devices as belonging simultaneously to biological and cultural histories. A linguistic device in the mind of an individual belongs to biological history in being the product of a biologically evolved language faculty that performs its function by producing such devices, adapted to the local linguistic community. The same linguistic device belongs to a cultural history: it has been reproduced in the mind of the individual, as in that of all the members of the linguistic community, because of its past and repeated performance of a specific linguistic function. The proliferation and stability of linguistic devices can be explained through a combination of cultural and biological (more specifically cognitive) factors. This seems to us much more insightful than a strictly cultural story.

Note that this account differs from a meme model of linguistic evolution in two respects. In the meme model, linguistic (and more generally cultural) evolution is homologous to biological evolution in that it too is essentially driven by a process of Darwinian selection. Biological and cultural evolution however are not otherwise articulated (apart from the obvious point that cultural evolution requires a species with biologically evolved capacities that makes it capable of culture). In contrast, by using Millikan's distinction between direct

and derived proper functions, we can describe the articulation between biological and linguistic evolution. Also, whereas the meme model assumes that memes are replicated, typically by "imitation", there is no postulation, in the present account, of a copying process. Generally, the word "reproduction" is ambiguous between a sense of *repeated production* and a sense of *copying*. Items of the same type can be produced again and again without being copied from one another, for instance by being produced from the same mould.

As Chomsky pointed out long ago, members of the same linguistic community do not learn to speak by copying the sentences they have heard. Most sentences of a language are uttered, if at all, only once, and, therefore, the overlap between the sets of sentences heard by two learners of the same language is quite small. If they learned their language by copying, language learners would end up speaking not just languages quite different from one another, but also languages quite different from those humans speak.

In fact, language learners sift, sort, and analyse linguistic inputs, and use them as evidence for grammar construction. From quite different inputs sets, they converge on similar grammars - they "reproduce" more or less the grammar of their community - thanks to a biologically evolved disposition to treat linguistic inputs to precisely this effect. A similar point can be made at the level of the lexicon. The contextual evidence on the basis of which a meaning can be attributed to a new word tends to be different in every case, and, moreover, quite often, a word is used with a contextual meaning different from its "literal meaning". Still, language learners converge on the same meanings for the same words, not by copying - and what exactly is there to copy on the semantic side? - but by deriving converging conclusions from quite different and sometimes divergent pieces of evidence. It may be assumed that the conclusions language learners derive about word meanings are guided by the language faculty, which constrains the kind of words that can occur in the lexicon (count nouns, mass nouns, transitive verbs, intransitive verbs, prepositions, etc.), and also, possibly, by cognitive constraints on the structure of concepts. To sum up this point, the stabilisation of linguistic devices is explained not by some kind of imitation of linguistic behavioural inputs, but by the constructive processing of these inputs by a biologically evolved language faculty. Such an account, though not exactly Millikan's own, fits much better within Millikan's conceptual framework than within the standard meme framework.

Millikan's motivation, in developing her theory of proper functions and in applying it to language was, primarily, to give an original account of meaning and intentionality, and to defend a certain view of linguistic communication that we do not share. Here is a stark statement of this view: "Speech is a form of direct perception of whatever speech is *about*. Interpreting speech does not require making any inference or having any beliefs [...] about speaker's intentions" (Millikan 1984: 62). According to Millikan, it is a sufficient condition for linguistic communication that the linguistic devices used succeed in performing their stabilising proper functions. For example, in the case of indicative sentences "speakers proliferate tokens of the indicative mood mainly insofar as these tokens produce, at any rate, *beliefs* in hearers [...] For this to be true it is not necessary that speakers should explicitly "intend" that their hearers believe what they say in a sense of "intend" that would require thinking of these beliefs or even having concepts of beliefs. [...] A proper function of speakers' acts in speaking could be to produce true beliefs in hearers even if the speakers had no concept of mental states and no understanding of the hidden mechanism whereby rewards result from speaking the truth" (1984 : 58). Similarly, Millikan argues, the direct proper function of imperatives is to produce compliance. Thus, an imperative utterance such as

"Eat!" performs its proper function when it causes the hearer to intend to eat and to act accordingly.

Until very recently, all explanations of the very possibility of communication were based on one version or another of the idea that a communicator encodes a content into a signal, and that the audience decodes the signal back into more or less the original content. After Grice, a second, wholly different mechanism was identified that also made communication possible: a communicator could communicate a content by giving evidence of her intention to do so, and the audience could infer this intention on the basis of this evidence. Of course, the two mechanisms, the coding-decoding, and the evidence-inference one, can combine in various ways. Today, most students of linguistic communication take for granted that normal comprehension involves a combination of decoding and of Gricean inference processes. By rejecting the Gricean approach (or confining to an occasional and marginal role), Millikan must, willy-nilly, fall back on some version of the coding-decoding explanation of verbal communication. There just is not to this day, in Millikan's work or anywhere else, a third type of explanation of the very possibility of communication.

In many respects, Millikan's view of verbal communication is a highly original one. Still, it is, we claim, a version, however atypical, of the code model of human communication (this is not, of course, Millikan's terminology). A code can be viewed as a systematic pairing of stimuli and cognitive responses shared by communicators, such that the production by a communicator of a stimulus belonging to the code has, both for communicator and audience, the function of producing the associated response in the audience. We do not dispute that human languages are codes in this sense. We do not dispute that the use of a shared code provides a sufficient explanation for many forms of communication. Indeed, it does explain how non-human animal communication works. But is what makes human communication possible the sharing of a common linguistic code? According to the code model, it is. According to the alternative, inferential model we will elaborate below, the sharing of a common linguistic code is what makes human communication so complex and powerful. What makes human communication possible at all, however, is human virtuosity in attributing intentions to one another.

In its standard form, the code model assumes that a human language is a pairing of sound and meanings, and that the meanings encoded by the sounds are, at a sentence level, propositional contents and attitudes, and at a sub-sentential level, constituents of these propositional contents and attitudes. Millikan's takes a different and original view of the cognitive responses paired with linguistic stimuli. In a nutshell, the responses she envisages are closer to perception on one side, to action on the other side, than the more abstract responses envisaged by standard accounts. Still, her model is a true code model of communication in that it explains communication by the systematic pairing of linguistic stimuli and responses. The representational resources of bees and their code are extremely different from the representational resources and language of humans, but some of the basic aspects of communication are, in a Millikanian perspective, the same. In both case, communication typically is a form of belief and desire transfer: cognition by proxy - or to use Millikan's phrase "natural teleperception" - made possible by a reliable pairing of stimuli and responses.

Most current discussions of the evolution of language give little or no place to pragmatics, and explicitly or tacitly accept the code model of linguistic communication. Human languages are seen as, precisely, a rich kind of code that allows for the encoding and decoding of any communicable thought.

A perfect code is one without ambiguity: each stimulus-type is paired to only one response-type. Simple perfect codes are common in animal communication. However, the code model does not require such perfection. Ambiguities do not necessarily compromise the model, provided that there is some method for automatically resolving them. Thus tokens of the same bee dance give, at different times of day, different indications regarding the location of food, but bees readily integrate relevant information about the position the sun is in their decoding of the dance, and understand the dance unambiguously. Human languages are obviously not perfect codes. Typical sentences contain multiple ambiguities. Thus, the one-word sentence "Eat!" might be interpreted as an order, a request, an encouragement, or an advice. It could be metaphorical, or ironic, etc.

As Millikan acknowledges, "understanding a language is never just decoding" (Millikan, 1998a:176). There must be further processes that use the output of decoding and information about the situation to fix the contextual meaning of the utterance. For Millikan, except in marginal and untypical cases, these further processes consist in strict disambiguation, that is, in the selection of one of the possible decodings of the utterance. All the possible contextual meanings of a linguistic device (in normal language use) must be conventionally associated (in the sense of Millikan 1998a) with this device. This actually implies truly massive ambiguity of nearly all linguistic expressions. As Millikan puts it:

"A language consists in a tangled jungle of overlapping, criss-crossing traditional patterns, reproducing themselves whole or in part for a variety of reasons, and not uncommonly getting in each other's way. Places where these patterns cross can produce ambiguities. These are sorted out not by conventions, but by the hearer managing to identify, by one means or another, the source of the pattern, that is, from which family it was reproduced" (1998a:176).

Although she does not dwell on the issue, Millikan's view implies, we insist, massive ambiguity. The idea closest to that of massive ambiguity is probably that of massive polysemy currently explored, for instance, in the work of Pustejovsky (1996). However, the idea of polysemy is that of many senses being generated in context and according to grammar-like rules, rather than that of many conventional senses each belonging to a distinct reproductively established family. (Polysemy would deserve an elaborate discussion from an evolutionary point of view, but we cannot pursue it here.) The task of the hearer of, say, the utterance "Eat!" is, on the polysemy account, to *generate* a contextually appropriate meaning for the lexical item "eat", whereas, according to Millikan, the hearer's task is to *recognise* to which one of the many family that proliferate phonetically indistinguishable but semantically different tokens of "eat" this particular token belongs (and the same problem has to be resolved with the imperative mood: to which of the many syntactically indistinguishable but semantically different tokens of the imperative does a particular token belong).

### **Massive ambiguity vs. Grice's "Modified Occam's Razor"**

Massive ambiguity and associated disambiguation processes (or, for that matter, massive polysemy and associated sense-generation processes) are not the only way to try and accommodate the fact that the same linguistic expression can convey many different meanings. In fact, Millikan's approach was developed as an alternative to Paul Grice's. Grice's influential approach is guided by a methodological principle he called Modified Occam's Razor: "Don't multiply senses beyond necessity." From a Gricean point of view, linguistic

meanings provide indications, and not necessarily full encodings, of speakers' meanings, and the same words *used with the same linguistic meaning* can quite ordinarily serve to convey different speaker's meanings. Comprehension is not a process of just decoding and disambiguating, but also of inference that goes beyond disambiguation.

In all modern pragmatic approaches inspired by Grice - and in particular in Relevance Theory (Sperber & Wilson [1986] 1995), the approach we favour and will adopt in the rest of this chapter -, three ideas go together: the goal of semantic parsimony expressed in Modified Occam's Razor, the distinction between sentence meaning and speaker's meaning, and the claim that to understand an utterance is to discover the speaker's meaning (using sentence meaning merely as a means towards that end). Millikan, rejecting the view that understanding an utterance is understanding what the speaker meant in uttering it, has in effect, to give up of the goal of semantic parsimony.

It might seem that, in accounting for the richness of communicated meanings, there is a balanced choice between two possible approaches. According to a first approach, which was, for Grice, at the time when wrote on the issue, exemplified by Ordinary Language philosophers, meanings communicated are, with marginal exceptions, meanings linguistically encoded. For instance, if the English word "and" can be understood sometimes as the corresponding logical connective, sometimes as *and then*, and sometimes as *and therefore*, there must be at least these three meanings in the mental lexical entry that English speakers have for "and". Reacting against Ordinary Language Philosophy, Grice pioneered another approach aimed at explaining richness of communicated meaning not at the linguistic-semantic level in terms of disambiguation, but at the pragmatic level in terms of inference. Thus, Grice argued, "and" semantically has just the logical-connective meaning, and all other interpretations are pragmatic speaker's meanings derived inferentially in context.

In fact, it is questionable whether the disambiguation and inferential derivation approaches really provide two alternative accounts of the richness of communicated meanings, more or less on par with each other. Grice's ideas have given rise to a whole field of research, pragmatics, pursued more and more within the framework of cognitive psychology. On the other hand, the disambiguation approach to the richness of communicated meanings consists in little more than theoretical hand-waving. To quote again Millikan, hearers resolve ambiguities "by one means or another." True, but then, the more massive the ambiguity implied by the theory, the less plausible that human minds can deal with it. Any theory that implies massive ambiguity faces a problem of psychological plausibility, and is betting on the outcome of future scientific development. Present studies of disambiguation in psycholinguistics (which tend to show that all the senses of a lexical item are unconsciously activated), and in pragmatics (which point the Gricean way) do not support the view that the richness of communicated meanings is based on massive ambiguity. This argument is, incidentally, similar to the sensible argument levelled by Millikan against Grice: that his account the recovery of speaker's meaning involves psychologically implausible complex reasoning.

Of course, it is often a wholly open empirical question whether a given interpretation of a given lexical item or of some other linguistic device is linguistically encoded or contextually inferred. On the other hand, there is a clear and ready answer to the empirical question whether the meanings that, in general, a word or a linguistic device may serve to convey form a small finite set. The answer, we would argue (and will shortly illustrate) is a resounding no. If indefinitely many new meanings can be communicated by means of the same linguistic

device used in a normal way, then the very notion of disambiguation (or, in Millikan's terms, of identifying from which family a linguistic token was reproduced) is of limited use in explaining the contextual aspects of comprehension. Meanings are not just disambiguated, they are in part disambiguated, in part constructed in context.

Let us illustrate. Julia puts a piece of cheesecake in front of Henry and says: "Eat!" In so doing, she intends him to find it desirable to eat the cheesecake there and then. Linked to the use of the imperative mood, Julia's utterance may have the character of a permission if it is manifest to the interlocutors that Henry would want to eat the cheesecake but might fear that it would be impolite to do so without having been invited; it may be an encouragement if it is manifest to the interlocutors that Henry's desire to eat the cheesecake is weak; or it may be an order, an enticement, or some less easily definable form of request, wish, advice etc. Millikan would assume that every distinct force that the imperative serves standardly to convey must be one of the conventional meanings of the imperative, and that the hearer somehow (and without attending to the speaker's beliefs and intentions) infers which of these meanings is being reproduced in the situation. Relevance theory assumes on the other hand that the imperative encodes merely desirability (whether to the speaker or to the hearer), and that its use in a given utterance and context allows the hearer to infer what specific form of desirability is meant by speaker.

Say Julia intends that Henry should recognise that she is encouraging him to eat the piece of cheesecake. She intends that his recognition of her intention to encourage him should indeed encourage him. If, as result of Julia's utterance, Henry understands that she is encouraging him to eat the cheesecake, then comprehension has been successful. This is so whether or not Henry complies: Julia's communicative intention is fulfilled by Henry's comprehension, that is, by his recognition of her meaning. Of course the goal that she was pursuing *through* communication, her "perlocutionary" intention, to use Austin's term, viz. to cause Henry to eat the piece of cheesecake, may be frustrated, but this is another story: understanding an encouragement or a request and complying with it are two different things. Similarly, from the speaker's point of view, securing comprehension and causing compliance (or, with other types of utterances, other perlocutionary responses) are two distinct purposes, achieving the first being a means towards achieving the second. The speaker, however, is much more in control of the hearer's comprehension than of the hearer's further cognitive or behavioural responses. So, of the two effects, securing comprehension and causing further responses, the first, being more regular, is more likely to play a stabilising role in the evolution of linguistic devices. More about this later.

"Eat!" can also serve to convey an ironical or a metaphorical speaker's meaning. Imagine for instance that, to highlight the thickness of a stout beer Henry has ordered, Julia tells him "Eat!" instead of "Drink!" An ambiguity-based analysis might consist in having "eat" be ambiguous between (among many other senses) ingesting solid food and ingesting thick drinks, and having the hearer somehow disambiguate. But this would be a case of multiplying senses beyond necessity. A Gricean approach would consist in assuming that only the standard linguistic sense of "eat" is involved here. According to Grice's own analysis of metaphor, Henry, encountering a linguistic meaning incompatible with what he can presume of Julia's communicative intention, searches for a meaning related to, but different from the literal one, a meaning she could have intended and expected to convey by means of her utterance (namely, drink a drink so thick that it resembles regular food). Henry then infers that this must indeed have been Julia's meaning. According to relevance theory, the same example would be explained by assuming that Henry accesses, in his mental lexicon, the standard entry

for "eat", and uses the information thus activated as a starting point for constructing a contextually relevant meaning that he then attributes to Julia. Millikan can treat such metaphors - which are neither dead, nor out of the ordinary, neither clearly conventional nor particularly creative - either as ambiguities, or else as Gricean exceptions to the normal flow of verbal communication. If such metaphors are cases of ambiguity, then every word has a great many stably attached metaphorical senses. If these are Gricean cases, then communication is much more Gricean than Millikan would have it.

In any case, a Millikanian speaker-hearer has in his or her memory many more senses for each lexical item (or for other linguistic devices such as the imperative mood) than does a Gricean or a relevance-guided speaker-hearer. Is this extra weight in memorised lexical information compensated by a lighter inferential task in comprehending utterances? Gricean inferential patterns involve using higher-order metarepresentations of the speaker's beliefs and intentions as premises and are notoriously cumbersome. Relevance theory departs from Grice precisely in assuming and describing a much lighter inference pattern where only the conclusion, but not the premises, need be about the speaker's intention. Since Millikan gives no indication of the inferential pattern involved in the kind of massive disambiguation she is hypothesising, there is no reason to assume that it would be lighter than relevance-based, or even than standard Gricean inference. In fact, the only plausible accounts of context sensitive disambiguation are to be found in Grice-inspired pragmatics and involve standard forms of pragmatic inference.

Moreover, even massive disambiguation may not be sufficient for the task at hand. If Henry had simply decoded Julia's utterance and disambiguated it (by whatever means) as, say, a literal request to eat, he would still not know what and how much was to be eaten, nor when. He might just eat a crumb, and thereby fulfil Julia's request literally interpreted. Even if Henry, somehow, disambiguated "Eat!" in this case as containing a reference to a direct object, and if, somehow, he inferred that the referent was the piece of cheesecake, this would not suffice. Should Henry take home the cheesecake, put it in his freezer, and eat it a month later, he would have acted in such a way as to render true the decoded, disambiguated, and referentially specified meaning of Julia's utterance, but, of course, he would have neither understood her nor complied with her intention. In all these respects, it is hard to see how Henry could understand Julia's utterance without paying attention to what Julia means by the utterance. Millikan asserts that comprehension is just a belief or desire transfer, but she does not begin to address decisive empirical issues in the study of comprehension that have been highlighted in modern pragmatics.

Let us qualify the last statement. In fact, Millikan does provide a highly Gricean pragmatic account of the word "this" (used as a whole noun phrase and without gestural demonstration as in: "this is how to live!"). She writes:

"... "this" often holds a place for improvisation. [...] the speaker has the hearer's capacities, viewpoint, and dispositions in mind as he utters "this" and utters it purposing that the hearer supply a certain referent for it, that is, that he translate it into an inner term having a certain referent. This referent is to be something proximate, or a sign or reminder of which is proximate, but beyond that the hearer is often pretty much on his own. He picks up his cues from the rest of the sentence and from his knowledge of what he and the speaker both know of that it would be reasonable for the speaker to expect him to think of first. When all goes well, speaker and hearer thus achieve a co-ordination, but

not a co-ordination that results from the speaker's and the hearer's speech-producing and understanding abilities having been *standardized* to fit one another" (1984: 167).

Clearly, Millikan equates standardisation and full-fledged determinate meaning, and all the rest is mere "improvisation". We would argue, on the one hand, that there is some modicum of standardisation involved in the use of "this", that makes it a word of English rather than of Italian, and a word different from "that". "This" does not encode, but is indicative of, the speaker's meaning in a standardised way. The indication is weak, it leaves a lot to be inferred, but it does indicate to an English hearer that what is to be inferred is an easily accessible referent. We would argue, on the other hand, that even when the words used do have a full-fledged meaning, their use still leaves room for what Millikan calls "improvisation" and which is just the inferential part of communication. So, for instance, the word "square" has a definite meaning, but when a speaker says "this field is square," she does not commit herself to the field actually having exactly four right angles and four equal sides. What she does is give an effective indication from which the hearer can infer her meaning, which, depending on the context, may involve a greater or lesser degree of approximation to squareness.

### **Comprehension as recognition of speaker's meaning**

Comprehension, as understood in modern pragmatics, crucially involves the recognition by the hearer of a specific intention of the speaker, the "speaker's meaning." The fact that the hearer is seeking to reconstruct the speaker's meaning is what focuses, constrains and indeed makes possible inferential comprehension (and, to begin with, inferential disambiguation, the necessity of which Millikan well recognises). We won't here give more positive arguments for the view that comprehension is recognition of speaker's meaning. The whole of modern pragmatics is predicated on this assumption, and its findings are arguments in favour of it. Of course, this does not make the assumption right, but those who deny it, are, in effect, implying that pragmatics as currently pursued is a discipline without an object, somewhat like the study of humours in ancient medicine. Surely, the burden is on them to show how pragmatics fails, and what is a better alternative to explain comprehension.

We will however address the view expressed by Millikan, that there is some serious implausibility in the very idea that comprehension is about speaker's meaning. Millikan does not deny the existence of speaker's meanings, but she sees their communication through linguistic means not as the normal form of linguistic communication, but as a departure from this normal form. "The truth in Grice's model," she says, "is that we have the ability to interrupt and prevent the automatic running on of our talking and our doing-and-believing-what-we-are-told equipment." We do this when we have discovered "evidence that the conditions for normally effective talking and for correct believing-on-the-basis-of-what-we-hear are not met" (Millikan 1984:69). In ordinary communication, she claims, going the Gricean way would be incredibly inefficient. However, for all we know, disambiguation that would not involve attending to the intentions of the speaker - if possible at all, which we doubt - might be even more dramatically inefficient.

Still, we do share Millikan's worry that comprehension as described by most Griceans is indeed implausibly cumbersome. There are two aspects to this. On the one hand the process of comprehension as described by Grice involves, in many cases, fairly sophisticated reasoning about the speaker's mental states. As we already mentioned and will discuss again below, this

is not the case in relevance theory, where the speaker's meaning is normally inferred without using as premises assumptions about the speaker's mental states.

On the other hand the very notion of speaker's meaning can be seen as implausibly complex. In Grice's original account (1957) a speaker's meaning involved a moderately complex two-level intention: roughly the intention to achieve a certain effect on the audience by means of the audience's recognition of this intention. In order to accommodate some objections, from, in particular, Strawson (1964) and Schiffer (1972), Grice with some reservations, and others more resolutely embraced the idea that communicative intentions involve many, or even infinitely many levels, or are infinitely nested. Millikan objected at length, and rightly, against the psychological implausibility or irrelevance of communicative intentions so understood. However, a Gricean-inspired approach to communication need not be committed to these complexities. Relevance theory's account of a communicative intention takes the objections into account but, just like Grice's original account, involves only two levels. According to this particular approach, a speaker has two intentions. She has the informative intention to make it manifest to the hearer that a certain state of affairs is actual or is desirable, and she has the communicative intention to achieve this informative intention by making it mutually manifest to the hearer and herself that she has this informative intention (For a detailed defence of this account and arguments that it is sufficient and genuinely involves only two levels, see Sperber & Wilson 1995, 1996).

It might still be felt that there is some implausibility in attributing to speaker-hearers, and in particular children, the ability to represent, as a matter of course, second-order metarepresentational intentions. However, to represent a second-order metarepresentational intention does not mean representing each and every time its internal structure. We standardly attribute to speakers of English the knowledge that *John killed Bill* entails *John caused Bill to die* without assuming that they mentally represent the latter each and every time they understand the former. Still, this attribution of knowledge is psychologically relevant: we assume that an English speaker who believed that Bill was alive, or that John had not caused anyone to die, would not be inclined to believe that John had killed Bill. Similarly, Henry can merely represent that Julia means that he should eat the piece of cheesecake now, without expanding the meaning of "means", except when needed.

Imagine the following scenario: Julia puts a piece of cheesecake in front of Henry and another one in front of Paul. Henry exclaims "this looks delicious!" and Paul sneers "cheesecake again!". Henry looks at Paul and hears Julia say "Eat!" Henry knows that Julia intends both of them to eat, but he - rightly as it happens - takes her meaning to be that Paul should. Without any difficulty, Henry thus dissociates Julia's informative intention to cause both of them to find eating the cheesecake desirable (already manifested by her putting the pieces of cake in front of them) from her communicative intention to make it manifest to Paul that she intends him to find eating the cheesecake desirable (manifested by her saying "Eat!"). Let us stress the relevant particulars of this case. Henry is not looking at Julia and therefore has no behavioural cues to the fact that Julia is addressing Paul. The utterance would be perfectly interpretable if understood as addressed to Henry, or to both Henry and Paul. Yet it is quite natural for Henry to infer that the utterance is addressed to Paul only, and that the unexpressed subject of "Eat!" is Paul. His inference is guided, we would argue, by considerations of relevance. Given the circumstances, Julia's utterance best achieves the expected level of relevance if understood as addressed to Paul. We take this example to illustrate the fact that hearers are capable, as a normal part of the process of comprehension, of inferentially discriminating different levels of intentions in the speaker. We have no evidence regarding the

age at which a child would be likely to perform such inferences and to interpret Julia the way Henry does, but there is nothing implausible in assuming that this would occur quite early in the development of verbal abilities (more about this below).

### **Fitting (post-)Gricean pragmatics into Millikan's conceptual framework**

Assume that verbal comprehension is recognition of speaker's meaning. Assume that what a linguistic utterance does is not to encode speaker's meaning, but to provide rich evidence from which the audience can infer speaker's meaning. Could languages playing such a role be described within the general conceptual framework put forward by Millikan? What would then be the direct and derived proper functions of linguistic devices? Before giving a general answer, let us take three examples, that of "now", of "eat", and of the imperative.

It is a misleading oversimplification to say that the indexical "now" refers to *the* time of utterance. Even ignoring various complications, and in particular the use of "now" in free indirect speech, the time indicated by "now" can be any time span, long or short. For instance, "now" in "I feel great now" could refer to the very minute of utterance, to a period of few days, or to a period of many years. "Now" does not encode any one of these time spans, nor is it ambiguous among them. Rather, it is indeterminate. The speaker's meaning, however, though it may be vague, is generally determinate. Therefore, in order to understand the speaker's meaning, the hearer must discover which time span is intended. So, we suggest, the direct proper function of "now" is to give evidence of the fact that the speaker's meaning includes a reference to a certain time span within which the utterance occurs. This direct function is performed through each token of "now" performing the derived proper function of indicating a specific time span.

Unlike the adverb "now", the import of which *must* be contextually specified for it to contribute to the meaning of any utterance in which it occurs, the verb "eat" has a full-fledged meaning. On occasions, it is used to convey just this meaning. For instance, in "Henry ate a piece of cheesecake," the meaning of "ate" seems to be just that of "eat" (plus some specification of the past tense). However, quite often, "eat" is used to indicate a meaning that may be more specific, less specific, or more specific in certain respects and less specific in other respects than the lexical meaning of "eat". For instance, a person declining an invitation to join a dinner party by saying "I have eaten" is indicating not just that she has eaten, but also that she has eaten a quantity such that she has no desire to eat any more (having eaten just a peanut would make her utterance literally true, but would nevertheless make her a liar). In this case, the meaning conveyed by means of "eat" is more specific than the lexically encoded meaning of "eat" (this example is discussed in greater detail in Wilson & Sperber forthcoming). In the example of Julia saying metaphorically "Eat!" to Henry who has ordered a thick stout, the lexicalised meaning of "eat" has to be made less specific (by ignoring the restriction to "food" in the sense where "food" is opposed to "drink") in order to understand Julia's meaning. Imagine now that Henry were asked if he would like to join a dinner party and answered: "I have had three stouts. As far as I am concerned, I have eaten." In this case, Henry's meaning conveyed by means of the word "eat" would be less specific than the lexical meaning of "eat" in being extended to the ingestion of thick drinks. At the same time, it would be more specific than the lexical meaning of "eat" in that it would indicate that he has ingested a quantity such that he had no desire to eat anymore. Thus the direct proper function of "eat" is to give evidence of the fact that the speaker's meaning includes a concept best evoked by "eat", a concept which may, but need not be, the very concept lexically encoded by "eat". This direct function is performed through each token of "eat" performing the derived proper

function of evoking, in the context, a specific concept which is part of the speaker's meaning on that occasion.

The imperative mood, we argued, does not encode any particular illocutionary force such as request or advice, nor is it ambiguous among all the particular forces it may serve to convey. (That is, speakers and hearers don't have a mental list of possible forces among which they must choose each time the imperative mood is tokened.) The imperative mood merely indicates desirability. Indicating that the action or the state of affairs described in the imperative mood is desirable typically falls short of yielding, by itself, a relevant enough interpretation. On the other hand, given expectations of relevance and contextual information, desirability may be understood as desirability for the speaker (as in the case of a request), or for the hearer (as in the case of an advice), or for both (as in the case of a wish). When desirability is understood as being for the speaker, the use of the imperative may further be understood as indicating expectations of compliance (as in the case of an order), or preference for compliance (as in the case of an entreaty), and so on. So, we suggest, the direct proper function of the imperative mood is to give evidence of the fact that the speaker is presenting the action or the state of affairs described as desirable in some way. This direct function is performed through each token of the imperative mood giving evidence that, together with contextual information, indicates which specific form of desirability is intended by the speaker.

The description of these three examples, "now", "eat", and the imperative mood, can be generalised to all meaning-carrying linguistic devices (see Carston 1998, Sperber & Wilson 1998, Wilson & Sperber forthcoming, for a thorough discussion from a pragmatic point of view). A linguistic device does not have as its direct proper function to make its encoded meaning part of the meaning of the utterances in which it occurs. It has, rather, as its direct proper function to indicate a component of the speaker's meaning that is best evoked by activating the encoded meaning of the linguistic device. It performs this direct function through each token of the device performing the derived proper function of indicating a contextually relevant meaning.

We follow Millikan in considering that the direct proper function of a linguistic device is what keeps speakers and hearers using and responding to the linguistic device in a reliable way, thus stabilising the device in a community. Our disagreement with Millikan has to do with the level of processing at which linguistic devices elicit the reliable response to be identified as their direct proper function. For Millikan, this reliable response is to be found at the level of belief or desire formation, or even at the behavioural level in the case of compliance. In particular, the function of a word is to contribute its "conventional meaning" to the overall meaning of an utterance which will then be accepted as a belief or a desire (depending on the mood) by the hearer. The function of the imperative is to cause desire and compliance, and so on. The problem, we argued, is that the same linguistic stimulus may elicit a great many different responses at the belief or desire level. In other words, at that level, responses are not reliably paired to stimuli. To invoke massive ambiguity and say that indistinguishable phonological or syntactic forms are, in fact, tokens of many different linguistic devices, is a way to shift the problem, not to resolve it. It amounts to saying that the reliability of linguistic stimuli is contingent on the ability of the hearer to identify the type to which the token belongs. As long as there is no account of how this can be reliably achieved, the very existence of reliable responses to linguistic devices at the level of belief or desire formation is in doubt, and so is the claim that the direct proper function of these devices is to be found at this level.

What is the alternative? Linguistic devices produce highly reliable responses, not at the level of the cognitive outputs of comprehension such as belief or desire formation, and even less at the level of behavioural outputs such as compliance, but at an intermediate level in the process of comprehension. Linguistic comprehension involves, at an intermediate and largely unconscious level, the decoding of linguistic stimuli that are then used as evidence by the hearer, together with the context, to arrive inferentially at the speaker's meaning. The same unambiguous linguistic item, decoded in the same way each and every time, can serve as evidence for quite different meanings in different contexts. (We do not, of course, deny the existence of true linguistic ambiguity, but there is much less of it than the code model of linguistic communication ends up implying, and moreover, the same inferential processes that explain other aspects of inferential comprehension explain disambiguation.) Linguistic devices have proliferated and stabilised because they cause these highly reliable cognitive responses *at this intermediate level*. Linguistic devices provide speakers and hearers with informationally rich, highly structured, and reliably decoded evidence of speaker's meaning. Note that this proper function of linguistic devices is not one speakers and hearers are aware of, let alone something they choose.

There could, in principle, be an intelligent species that communicated the way Millikan believes humans do: with speakers using utterances directly to cause belief or desire transfer, and hearers merely decoding and disambiguating these utterances and automatically turning the resulting interpretation into a desire or belief of their own. The language of such a species should present many fewer ambiguities than actual human languages, and only ambiguities that can be easily resolved either on the basis of the linguistic context (the "co-text"), or by applying simple rules to pick out the pertinent piece of information from the environment (as, for instance, in replacing the first person pronoun with a reference to the actual speaker).

The reaction of a hearer to a speaker in such a species, using a language à la Millikan, would look very much like that of a person hypnotised to the hypnotist, where belief and desire transfers do actually occur. This raises, of course, the problem of explaining how hearers could escape being systematically deceived and manipulated by speakers. Communication is a form of co-operation. Co-operation is vulnerable to free-riding, which, in the case of communication, takes the form of manipulation and deception. In the study of any communicating species, explaining how come the benefits of communication are not offset by the cost of deception is a major problem (Dawkins & Krebs 1978, Krebs & Dawkins 1984, Hauser 1996).

In the case of human communication, explaining how the costs of possible deception are contained crucially involves the fact that comprehension and acceptance are two distinct steps in the overall process. It may be (at least in some socio-cultural contexts) that people believe most of the things they are told, but this is not because they are hypnotised or gullible. It is rather that they mostly interact with relatives and friends with whom they cooperate and from whom sincerity can be expected in ordinary conditions. People are typically distrustful of information provided by strangers, or by competitors, or even by relatives and friends in situations of conflict. Communicated information is sifted, rather than automatically accepted as Millikan argues. Another part of the explanation of the viability human communication is the fact that comprehension is, *pace* Millikan, a form of mindreading and links easily with attending to the speaker's benevolence and competence (for a more thorough discussion of the metarepresentational mechanisms involved in sifting communicated information, see Sperber forthcoming).

So, yes, assuming that the problems raised by ambiguity and deception were somehow avoided or solved, there could be a species that communicated in the way Millikan believes humans do. On the other hand, there is nothing in Millikan's teleofunctional framework that implies that communication can only evolve in the way she claims it did. There could be a species that communicated in the way Grice or relevance theory says humans do, and, in fact, we believe that humans are such a species. At this point we have reached one of our goals: to pry apart Millikan's overall framework from her view of language, and fit this framework together with a view she opposes and according to which linguistic comprehension is a form of mindreading. In the next two sections we explore some of the evolutionary implications of this view of language.

### **Linguistic communication and mindreading**

In the past twenty years, the study of the capacity to attribute mental states such as beliefs or intentions to others has become a major focus of cognitive science under names such as Theory of Mind or Mindreading (e.g. Carruthers & Smith 1996). There is a growing body of evidence and arguments tending to establish that a mindreading ability is an essential ingredient of human cognition, and moreover, is a domain-specific evolved adaptation (rather than an application of some general intelligence, or cultural competence). What are the relationships between mindreading and the language faculty? Millikan argues that linguistic communication is independent of mindreading, whereas Grice and post-Griceans assume that linguistic communication involves a form of mindreading where, by speaking, the speaker helps the hearer read her mind. These two views of comprehension as a cognitive process fit differently with developmental and evolutionary considerations.

At the developmental level, Millikan assumes that linguistic abilities develop before mindreading, and sees this as further evidence against a Gricean view of linguistic communication. At first blush, the evidence might seem to be in her favour. Whereas language comprehension starts developing in the second year of life, it is only around the age of four that children pass the much-studied "false-belief task" (in which they are asked to predict where a character will look for an object that she falsely believes to be in one location when, in fact, it has been moved to another). Success at the false-belief task is often treated as the criterion establishing mindreading abilities. Indeed, success at the task is a clear demonstration of mindreading abilities. Failure, however, is by no means a demonstration of total lack of such abilities. Mindreading is not an all-or-none affair. It develops in stages from infancy (Baron-Cohen 1995, Gergely & al. 1995). People with autism, a condition now understood as involving a deficit in mindreading abilities, lack the ability to a greater or lesser degree (Frith 1989, Happé 1994).

The attribution of a meaning to a speaker, and the prediction that a person with a false belief will act on this belief, though both involving mindreading, are two very different performances. The formal resources involved in the two cases are not the same. In the case of speaker's meaning, what is needed is the ability to represent an intention of someone else about a representation of one's own - a second-order metarepresentation of a quite specific form. (From a modularist point of view, it is quite conceivable that children might develop the ability to represent speaker's meaning before being able to deploy other types of second-order metarepresentations.) In the case of false beliefs, a first-order metarepresentation of a belief of someone else is sufficient, but what is needed is the ability to evaluate the truth-value of the metarepresented belief and to predict behaviour on the basis of false belief. We are not aware of any argument to the effect that the ability needed to pass the false-belief task is a

precondition for the ability needed to attribute speaker's meaning. There is nothing inconsistent or paradoxical therefore in the idea of an individual capable of attributing speaker's meaning and incapable of attributing false beliefs (and conversely).

There are, on the other hand, functional reasons to expect the ability to attribute false beliefs to develop after the ability to communicate verbally. The attribution of false beliefs to others plays an obvious role in the ability to filter false information communicated either by mistaken or by deceitful speakers. It plays an obvious role also in the ability to deceive others by communicating false information. These abilities are asymmetrically dependent on the ability to communicate. Suppose, moreover that, as we have argued, comprehension consists in the attribution of a meaning to the speaker. Then there are reasons to expect attribution of false beliefs to develop after attribution of speaker's meaning.

The fact that success at the false-belief task occurs three years or so after the beginnings of verbal comprehension is no evidence against the view that comprehension is a form of mindreading. Are there, though, positive arguments or evidence to the effect that, say, two-year-olds (who fail the false-belief task) do attribute meaning to speakers? We would be tempted to say that we all *know* that they do. As speakers, we take for granted that when we say something we mean something, and that people - including very young children - who understand what we say understand what we mean (understand *us*, in an ordinary sense of the expression). But of course, this may be a piece of mistaken naïve psychology. A scientifically more compelling argument is this: young children do disambiguate, identify referents, and understand implicatures. As we argued before, the only actual explanations of such achievements (as opposed to hand-waving in the direction of unspecified explanations) draw on (post-)Gricean pragmatics and presuppose the capacity on the part of the comprehender to attend to speaker's meaning. Further positive evidence of an experimental kind is provided by Paul Bloom's work which shows that the acquisition of lexical meanings - which is involved in very early language acquisition - requires attention to speaker's intentions (Bloom 1997).

At an evolutionary level, the biological evolution of language is, for Millikan, quite independent from that of mindreading. From a Gricean viewpoint, the evolution of language should be linked to that of mindreading, since utterances are encodings of speaker's thoughts, and are typically recognised as such by the audience (Pinker 1994). Linguistic communication enhances mindreading abilities (and even, some might argue - e.g. Dennett 1991 -, makes true mindreading possible in the first place), and also exploits these abilities in complex cases where Gricean inferences must supplement linguistic decoding. It is reasonable therefore, from a Gricean point of view, to assume a co-evolution of language and mindreading, without committing oneself any further.

From a relevance theory point of view, it is also reasonable to assume a co-evolution of language and mindreading, but there are reasons to commit oneself to a more precise articulation of the two. In standard Gricean approaches, inference is seen as needed in discovering the implicit part of the speaker's meaning, while the explicit part is seen as decoded (and disambiguation is not much discussed). Accordingly, there could have been an initial stage in the evolution of language where utterances were wholly explicit and decoded, with Gricean inferences about implicit content evolving only at a later stage. In other terms, Gricean communication could result from a partial change of function of what might have been, at an earlier stage, a strict code. According to relevance theory, on the other hand, human verbal communication is never a matter of mere decoding. In fact, in its basic structure, inferential communication does not even depend on linguistic stimuli: other

behavioural stimuli, e.g. improvised mimes, may provide adequate evidence of a communicator's intention. Linguistic utterances, however, provide immensely superior evidence for inferential communication. They can be as richly and subtly structured as the communicator wishes, and they are reliably decoded by the audience at an intermediate level in the process of comprehension. The function of linguistic utterances, then, is - and has always been - to provide this highly precise and informative evidence of the communicator's intention. This implies that language as we know it developed as an adaptation in a species already involved in inferential communication, and therefore already capable of some serious degree of mindreading. In other terms, from a relevance theory point of view, the existence of mindreading in our ancestors was a precondition for the emergence and evolution of language.

### **The bootstrapping problem and its solution**

Most evolved domain-specific cognitive abilities have a specific domain of information (a "proper domain" - see Sperber 1996, Ch. 6) available in the environment well before the ability develops, and they can be seen as adaptations to that aspect of the environment. For instance, different individuals have distinctive faces; an evolved face recognition ability is an adaptation to the prior presence of these faces in the environment and an exploitation of their informational value. A mutant endowed with a face recognition ability could benefit from it, even if he or she were the only individual so endowed. Some cognitive abilities, however, have a specific domain of information that is initially empty and that gets filled only by the behaviour of individuals who already have and use the ability in question. For instance, an ability to enter into reciprocal exchanges is an adaptation to the opportunities offered by other individuals who are also endowed with this ability. A unique mutant endowed with a reciprocal exchange ability could not benefit from it until other individuals became also so endowed. Thus the emergence in evolution of abilities that need to be shared by several individuals in order to be adaptive raises a specific bootstrapping problem.

Innate codes found in non-human animals are cases in point. What would be the use of an innate code in a single individual, as long as other members of its species, lacking such a code, could neither decode its signals, nor send it signals of their own? To point out that any actual code is likely to result from several mutations and to have evolved in small steps spreads the problem but does not resolve it. There are, however, at least three ways to tackle this puzzle. The first is to assume that an innate code spread in a population as a neutral trait, initially without benefit but also without significant cost, so as not to be selected out. The trait then became advantageous and was selected *for* (Sober 1984), when enough individuals sharing it could use it in their interactions and benefit from it. Such a development can occur rapidly, say among the offspring of the initial mutant individual endowed with the trait. Another plausible speculation is that the trait was initially selected for thanks to some other beneficial effect, and that its function as a code emerged as a new function added or substituted to some previous one. A third, more controversial speculation is that the signals of the code emerged first as "cultural" items, transmitted through learning and not through genes; it then became advantageous to possess them innately, sparing the cost of learning (this strictly Darwinian but Lamarckian-looking possibility is known as a Baldwin effect).

Human languages, however, are not innate codes. The human language faculty is not an ability to produce and interpret signals, it is an ability to acquire culturally transmitted languages. Thus the bootstrapping problem raised by the emergence of the human language faculty is not as easily speculated away as that raised by that of the innate code of most

animal communication. Even if a Language Acquisition Device, starting as a neutral trait, became shared by a number of individuals, this would not be advantageous to them, since there would still be no language to acquire. The argument applies not just to the initial emergence of a rudimentary language faculty, but also to any later biological development of this faculty. The emergence of an ability to acquire a different, presumably richer language, is not advantageous in the absence of such a language to be acquired.

This bootstrapping problem is at its worst if one accepts the code model of verbal communication. Coded communication works at its best when the interlocutors share exactly the same code. Differences in code typically lead to communication failures. Now, a modification in the language faculty of one individual, if it had any effect at all on the structure of its internalised language, would introduce a mismatch between her linguistic code and that of other people, and would have a detrimental effect on her ability to communicate. An individual endowed with a language faculty different from that of others, even if it were "more advanced" in some sense, would stand to suffer rather than to benefit from it.

If, on the other hand, we adopt the inferential model of communication, the puzzle becomes much more tractable. Inferential communication is a matter of reconstructing the communicator's informative intention on the basis of the evidence she provides by her utterance. Successful communication does not depend, then, on the communicator and addressee having exactly the same representation of the utterance, but on having the utterance, however represented, seen as evidence for the same intended conclusion. Different decodings may provide evidence for one and the same inferential interpretation. Here, a metaphor may help. Think of a meanings as points in semantic space. Then according to the code model, any device encodes such a point (or several such points when it is ambiguous). According to the inferential model, on the other hand, a linguistic device encodes a *pointer* in semantic space (or several such pointers when ambiguous) that makes accessible, with ordered saliencies, a series of points. According to the code model, a mismatch between the codes of interlocutors must result in the selection of different points, i.e. different meanings, by the communicator and audience. Not so according to the inferential model: differently situated pointers may point to the same meaning. The inferential model is thus compatible with a much greater degree of slack between the codes of interlocutors.

Acquiring and using a non-standard version of the common code need not involve any cost, it may even be advantageous. In particular, a language faculty that leads to the internalisation of a grammar that attributes more structure to utterances than they superficially realise (that project onto them "unexpressed constituents" for instance) may facilitate inferential comprehension (Sperber 1990).

Imagine a stage in linguistic evolution where the languages available consisted in simple sound-concept pairs, without any higher structure at all. "Drink" in such a primitive language encoded the concept *drink* and nothing else, "water" encoded the concept *water* and nothing else, and so on. With such a limited code, the decoding by a hearer of a concept encoded by a speaker falls quite short of achieving communication between them. An addressee associating for instance the concept *water* with the utterance "water" is not thereby being informed of anything. Even a concatenation of expressions in such a language such as "drink water" does not have as its decoded interpretation what we all understand from the homonymous English expression. It does not denote the action of drinking water. Rather two concepts, *drink* and *water*, are activated without being linked either syntactically or semantically. The mental

activation of one or several concepts without syntactic linkage does not describe a state of affair, whether actual or imagined. It does not express a belief or a desire.

If, however, the people using such a rudimentary code were capable of inferential communication, then the activation in their mind, through decoding, of a single concept might easily have provided all the necessary evidence needed to reconstruct a full-fledged, propositional speaker's meaning (see Stainton 1994 for a related point). Imagine two individuals of this ancestral species walking in the desert. One points to the horizon and utters "water". The other correctly infers that the speaker means *here is some water*. They reach the edge of the water, but one of them collapses, exhausted, and mutters "water". The other correctly infers that the speaker means *give me some water*. To the best of our knowledge, there is no evidence that the signals of animal communication ever permit such an open range of quite diverse interpretive elaborations.

Imagine now a mutant whose language faculty is such that she expects elementary expressions of the code she is to acquire to be either arguments or one- or two-place predicates. She classifies "drink" as two-place predicate, "water" as an argument, and so on. When she hears her collapsing companion mutter "water," what gets activated in her mind as a result of decoding is not just the mere concept *water*, but also a place-holder for a predicate of which *water* would be an argument. Her decoding, then, goes beyond what had been encoded by the speaker, who, not being a mutant, had spoken the more rudimentary language common in the community. This mismatch, however, far from being detrimental, is beneficial to the mutant: her inferential processes are immediately geared towards the search for a contextually relevant predicate of which *water* would be an argument.

When she talks, our mutant encodes by means of signals homonymous with those of the community not just individual concepts, but predicate-argument structures. When she utters "water," her utterance also encodes an unexpressed place-holder for a predicate; when she utters "drink," her utterance also encodes two unexpressed place-holders for two arguments; when she utters "drink water," her utterance encodes the complex concept of *drinking water* and an unexpressed place-holder for another argument of *drink*, and so on. These underlying linguistic structures are harmlessly missed by her non-mutant interlocutors, but are useful to other mutants, pointing more directly to the intended interpretation. In the language of these mutants, new symbols, for instance pronouns for unspecified arguments, may then stabilise. This illustrates how in an inferential communication system, a more powerful language faculty, which causes individuals to internalise a linguistic code richer than that of their community, may give them an advantage and may therefore evolve (whereas in a strict encoding-decoding system, a departure from the common code may be harmful or harmless, but not advantageous).

This line of reasoning applies to the very emergence of a language faculty: being disposed to treat an uncoded piece of communicative behaviour as a "linguistic" sign may have facilitated the inferential discovery of the communicator's intention, and led to the stabilisation of this stimulus type as a signal.

## **Conclusion**

Millikan's conceptual framework allows one effectively to articulate various issues raised by the biological and cultural evolution of language. At the same time, her own view of language makes it more difficult to deal with these issues. In particular, it leaves one with an extra

problem of massive ambiguity, and it makes the bootstrapping problem, if anything, less tractable. Fortunately, Millikan's conceptual framework can be dissociated from her view of language. It can be applied to Gricean or relevance-theoretic approaches to language, with, we hope to have shown, some interesting results.

### References

- Baron-Cohen, S. (1995) *Mindblindness*, Cambridge, Mass.: MIT Press.
- Bickerton, D. (1990) *Language and Species*, Chicago : The University of Chicago Press.
- Bloom, Paul (1997). Intentionality and word learning. *Trends in Cognitive Sciences*, 1: 9-12.
- Boyd, R.; Richerson, P.J. (1985) *Culture and the Evolutionary Process*, Chicago : The University of Chicago Press.
- Byrne, R.W.; Whiten, A. (eds.) (1988) *Machiavellian Intelligence : Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, Oxford : Clarendon Press.
- Byrne, R. W.; Whiten, A. (eds.) (1997) *Machiavellian Intelligence II : Extensions and Evaluations*, Cambridge : Cambridge University Press.
- Carston, Robyn (1998). *Pragmatics and the Explicit-Implicit Distinction*. University College London PhD thesis.
- Carruthers, P. (1996) *Language, Thought and Consciousness : An Essay in Philosophical Psychology*, Cambridge : Cambridge University Press.
- Carruthers, P. and Boucher, J. (eds.) (1998) *Language and Thought*, Cambridge : Cambridge University Press.
- Carruthers P. and Smith P. (eds.) (1996) *Theories of Theories of Mind*, Cambridge : Cambridge University Press.
- Cavalli-Sforza, L.L. e Feldman, M.W. (1981) *Cultural Transmission and Evolution : A Quantitative Approach*, Princeton, Princeton University Press.
- Chomsky, N. (1980) *Rules and representations*. New York: Columbia University Press.
- Dawkins, R. (1976) *The Selfish Gene*, Oxford : Oxford University Press.
- Dawkins, R. (1982) *The Extended Phenotype*, Oxford: Oxford University Press.

- Dawkins, R. and Krebs, J. R. (1978) Animal signals : Information or manipulation? In J. R. Krebs & N. B. Davies (eds.) *Behavioural Ecology*, pp. 282-309, Oxford : Basil Blackwell Scientific Publications.
- Dennett, D. (1991) *Consciousness Explained*, New York : Little Brown and Co.
- Dennett, D. (1995) *Darwin's Dangerous Idea: Evolution and the Meaning of Life*, New York: Simon and Schuster.
- Dennett, D. (1998) Reflections on language and mind. In P. Carruthers, J. Boucher (eds.) *Language and Thought*, pp. 284-294, Cambridge : Cambridge University Press.
- Dunbar, R. I. M. (1996) *Grooming, Gossip and the Evolution of Language*, London : Faber & Faber.
- Durham, W. H. (1991) *Coevolution : Genes, Culture and Human Diversity*, Stanford : Stanford University Press.
- Frith, U. (1989) *Autism: Explaining the Enigma*, Oxford: Blackwell.
- Gergely, G. Nadasdy, Z., Csibra, G, and Biro, S. (1995) Taking the intentional stance at 12 months of age. *Cognition* 56 (2) 165-173.
- Gomez, J-C. (1998) Some thoughts about the evolution of LAD, with special reference to TOM and SAM. In P. Carruthers and J. Boucher (eds.) *Language and Thought*, pp. 76-93. Cambridge : Cambridge University Press.
- Goody, E. N. (1997) Social intelligence and language : Another Rubicon? in A. Whiten & R. Byrne (eds.) *Machiavellian Intelligence II*. pp. 365-396. Cambridge : Cambridge University Press.
- Grice, H.P.(1957) Meaning, *Philosophical Review*, 66, 377-388.
- Grice, H.P. (19) *Studies in the way of words*. Cambridge, Mass: Harvard University Press.
- Happé, F. (1994). *Autism: An introduction to psychological theory*. London: UCL Press.
- Hauser, Marc D. (1996) *The Evolution of Communication*. Cambridge, Mass: MIT Press, Bradford Books.
- Humphrey, Nicholas K. (1976). The Social Function of Intellect. In P.P.G. Bateson and R.A. Hinde (eds.) *Growing Points in Ethology*. pp. 303-317. Cambridge : Cambridge University Press.
- Hurford, J. R., Studdert-Kennedy, M., Knight, C. (eds.) (1998) *Evolution of Language*, Cambridge : Cambridge University Press.

- Krebs, J.R. & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (eds.) *Behavioural Ecology*, pp. 380-402. Sunderland, MA: Sinauer Associates.
- Lumsden Charles J. & E.O. Wilson (1981). *Genes, mind and culture*. Cambridge, Mass: Harvard University Press.
- Millikan, Ruth (1984) *Language, Thought and Other Biological Categories*, Cambridge, Mass : MIT Press.
- Millikan, R. (1993) *White Queen Psychology and Other Essays for Alice*, Cambridge, Mass : MIT Press.
- Millikan, R. (1998a) Language conventions made simple. *The Journal of Philosophy*, XCV, 4, pp. 161-180
- Millikan, R. (1998b) A common structure for concepts of individuals, stuffs, and real kinds : More mama, more milk, more mouse. *Behavioural and Brain Sciences*, 9 (1), pp. 55-100.
- Pinker, S. (1994) *The Language Instinct*, New York : Morrow.
- Pustejovsky, J. (1996) *The Generative Lexicon*, Cambridge, Mass : MIT Press.
- Schiffer, S. (1972) *Meaning*. Oxford: Clarendon Press.
- Sober, E. (1984). *The nature of selection*. Cambridge Mass.: MIT Press.
- Sperber, Dan (1990). The evolution of the language faculty: A paradox and its solution. *Behavioral and Brain Sciences* 13 (4), 756-758.
- Sperber, D. (1994) Understanding verbal understanding. In J. Khalfa (ed.) *What is Intelligence?* pp. 179-198, Cambridge : Cambridge University Press
- Sperber, D. (1996) *Explaining Culture : A Naturalistic Approach*, Oxford : Basil Blackwell.
- Sperber (forthcoming) Metarepresentations in an Evolutionary Perspective. In D. Sperber (ed.) *Metarepresentations*. Oxford: Oxford University Press.
- Sperber, D, and Wilson, D. (1995) *Relevance: Communication and cognition*. Second Edition. Oxford : Basil Blackwell. (First edition 1986)
- Sperber, D. and Wilson, D. (1996) Spontaneous deduction and mutual knowledge. *Behavioural and Brain Sciences* 110:4, 179-184
- Sperber, D, and Wilson, D. (1998). The mapping between the mental and the public lexicon. In Peter Carruthers and Jill Boucher (eds.) *Language and Thought*, 184-200. Cambridge, Cambridge University Press

Stainton, Robert J. (1994) Using non-sentences: An application of Relevance Theory. *Pragmatics and Cognition*, 2 (2): 269-284.

Strawson, P. (1964) Intention and convention in speech acts. *Philosophical Review* 73: 439-460.

Wilson, D. & Sperber, D. (forthcoming) Truthfulness and relevance.