Diego MARCONI
Università di Torino

UNDERSTANDING AND REFERENCE

The title will remind some readers of a well-known article by Hilary Putnam [Putnam, 1979]. In that work, he tried to dissociate the two concepts of the title -reference and understanding ; he wanted to show that the theory of language understanding and the "theory of reference and truth have much less to do with one another than many philosophers have assumed" (p.199). I tend to agree with him as far as current theories of reference are concerned. But, on the other hand, I would like to show that any plausible account of language understanding must deal with what I shall call "referential competence" (or the referential aspect of semantic competence). By this I mean the ability to use words (and sentences) to discriminate, and have other people discriminate among objects in the *real world* : the ability to tell cats from cows by calling the former 'cats' and the latter 'cows', to describe a man as *walking* rather than *running*, to pick up the appropriate tool if one is requested (and willing) to obey the order "Bring me the hammer, not the pliers". I believe that referential competence is a crucial factor of what Putnam himself has called "the contribution of our linguistic behavior to the success of our total behavior" (p.202). So, to this extent, I believe that understanding should not be kept apart from reference.

In a later, even more famous article on "Brains in a Vat" [Putnam, 1981], Putnam argued that reference -the ability to refer to objects, such as trees- requires some sort of causal interaction with the domain of reference. Part of the following should be read as an elaboration on his thesis.

Why machines do not understand natural language

We do not really know what it is to understand natural language, but we know there are performances which stand a *criterial* relation to understanding, in Wittgenstein's sense. For example, if a person can summarize a text we say that he has understood it (whereas if he cannot, we doubt that he understood). If a person can answer questions concerning the topics a text is about, and his (her) answers appear to be based on the information contained in the text, we say that person has understood the text- whereas if he (she) cannot answer, it is legitimate to raise doubts about his or her understanding. If she can correctly translate the text into another language, we say she understood (but if she cannot and yet does know the second language, we are inclined to say that she did not understand). Such are the "paradigmatic" cases in which we say that somebody understands a text in a natural language : Wittgenstein would say that our use of such words as 'understanding' and 'to understand' is intertwined with such performances and the ability to carry them out. Of course, understanding is not identical with summarizing, or answering questions, or translating (or at any rate, it would be highly unnatural to say so). However, we probably learn how to use the concept of understanding by learning how to assess such performances.

As we all know, today we have artificial systems which can carry out such tasks, with different degrees of success. They are called "natural-language understanding systems" just because they are capable of one or the other among such performances[1]. But, in spite of the fact that these systems can carry out the very performances on the basis of which we normally say of a human being that she understands a language, many would say that such systems do not *really* understand natural language[2].

It is fair to acknowledge that the present systems are certainly not as good as human beings at carrying out such tasks : their translations are often clumsy, their summaries unintelligent, the questions they can aswer, relatively few in number. Moreover, the existing systems can (usually) carry out *one or the other* among such tasks : in contrast with human beings, they are *either* translators *or* question-answering devices *or* automatic abstractors. Finally, the range of texts that each system can process is strongly restricted, lexically at any rate. These are technologically important limitations, sharply discriminating between the systems' performance and human performance on the same terrains. In order to overcome such limitations, more is required than just building huge lexical databases (which perhaps we would not know how to manage) or integrating complex systems into one

[1]Since the repudiation of Turing's test, the stress has tended to be placed on the way such performances are carried out; i.e. on the structure of the programs and the kind of data they have access to. Here I am neither implying that such features are irrelevant to whether a system can legitimately be said to understand language, nor restoring some version of Turing's test as having definitional import. I am simply suggesting that no one would think we were dealing with (prospect) natural-language understanding systems if they were not capable of this kind of performances.

[2]Prominent among such critics is of course John Searle. The view which I put forth in the present article bears some similarity to an objection that Searle discusses under the name of the "robot's answer" [see Searle, 1980]. I believe that Searle's discussion of it is totally unfair even to his own formulation of the

big system (which would risk being inefficient and uncontrollable) : we need to solve problems which have not even been formulated clearly so far, from metaphoric language to pragmatic competence and "contextual" knowledge.

objection : hovewer, I will not resume the particular discussion, as this is not intended as an examination of Searle's views on the matter.

Even though the AI community is concentrating on this kind of limitations, I surmise that it is not essentially because of them that natural-language processing systems are said not to *really* understand natural language. To realize this, imagine we have been successful in building a very sophisticated *understanding* system of the standard type. Such a system would have a perfect syntactic analyzer, a vast lexical database, and a semantic interpreter capable of compositionally constructing fully analytic semantic representations : they would be as explicit as we need them to be in order for the system to carry out -thanks to a reasoning module- all the inferences that could be plausibly attributed to a competent, or even a very competent speaker. From 'There are four elephants in the living-room' our system would infer that there are four large animals in the living-room, that there are four elephants in the house, that there is an even number of elephants in the living-room, that there are higher mammals (to be more precise, proboscideans) in the living-room ; it could even infer that the living-room's furniture is likely to be badly spoiled. We suppose that our system can answer questions and summarize texts, thanks to a powerful text-generator. In short, we imagine that our system is the traditional artificialist's dream come true.

Why would we say, even of such a system, that it really does not understand the language it can process? What is it that the system cannot do? One could raise the suspicion (which would be fatal, of course, from the standpoint of model-theoretic semantics[3]) that the system does not know the *truth-conditions* of the sentences it processes. Is such a suspicion well-founded? It depends on what is meant by 'knowing the truth-conditions'. For instance, if knowing the truth-conditions of an English sentence E is to know a true biconditional of the form 'E is true (in English) iff f(p1, p2, ..., pn)', where p1, ..., pn are atomic sentences of English[4] and f is a function which the system can compute, then it can be argued that the system does know the truth-conditions of E and the other sentences it can process. For in this sense knowledge of the truth conditions is a kind of inferential competence ; but we assumed that our system is as competent inferentially as a very competent human speaker. Therefore, in this sense the system knows the sentences' truth-conditions better than most of us.

[3]I am referring e.g. to D. Lewis' classical criticism of decompositional semantics "à la Katz" [Lewis, 1970]. See [Fodor 1976, p.120-22], for interesting counterarguments.

[4]Or of any other language the system can command.

Would it be right to say that the system doesn't know a sentence's truth-conditions in the sense that it cannot establish, for each situation

S, whether the sentence is true or false in S? That would not be correct either. In fact, if situation S is *described in language*, the system *can* indeed determine whether a sentence is true or false in S. This is exactly what systems do which (like our system) can aswer questions relative to a text's topic : such systems determine whether certain sentences (corresponding to the questions) are true or false in the situations described by the texts they have processed. As we assumed that there are no limitations -either lexical or syntactic or discursive or of any other kind- to the texts the system can process, we conclude that our system can indeed determine whether a given sentence is true or false in any situation which can be described in the language the system can interpret.

But, on the other hand, our system cannot establish whether a sentence is true or false in a situation which is not given through language. For example, it cannot determine whether a sentence is true or false in the real world : it cannot verify the sentence, unless the *real world* is given to it through a linguistic description. If you place the system in a room and require it to evaluate the sentence 'There are at least four chairs in this room', the system won't do it.

The same kind of inability -or more precisely, a strictly related inability- can be highlighted by focusing on the reference of single words rather than on the sentences' truth-conditions. We assumed that the system has remarkable inferential ability : for example, it can draw many inferences concerning elephants, i.e. inferences involving sentences where the word 'elephant' occurs. Still, one could claim - perhaps Searle would claim[5]- that such inferences as the system can carry out are not *about elephants* at all : strictly speaking, it would not even be correct to say that they involve sentences in which the *word 'elephant'* occurs. The system can indeed manipulate strings of symbols including a symbol which materially coincides with the English word 'elephant'. Such a symbol, however, is devoid of meaning for the system : emphatically, it does not mean "elephant" (i.e. it does not mean what the English word 'elephant' means). Whatever conclusions the system can infer are not in themselves about elephants : they are strings of symbols, meaningless for the system, which *we* (the system's users) interpret as pertaining to elephants.

There is much that is wrong with the familiar argument : still, it does point, though confusedly, to real inadequacies of the system and its understanding of language. First of all, it is certainly wrong to *oppose* knowledge of meaning and the ability to manipulate symbols, as if *genuine* knowledge of meaning were forever *something else* with respect to symbol-manipulating ability. Even a little acquaintance with Wittgenstein suffices to persuade most of us that

[5]Searle believes indeed that symbol manipulation by computers never involves meanings. It is surprising that recently, while reiterating his view to that effect, he also claimed that "any information which could be expressed in a language can be encoded (in machine-readable notation)" [Le Scienze, n°259, marzo 1990, p.17]. Supposedly, information involves meaning : a mere string of sign (be they Os, 1s, or of any other shape) is not a piece of information, nor does it encode a piece of information.

knowing the meaning is, essentially, being able to use. It is still true that "The account according to which understanding a language *consists in* being able to use it... is the only account now in the field" [Putnam, 1979, p.199]. The problem is not whether knowledge of meaning can be *reduced* to symbol-manipulation, but what *kind* of symbol-manipulating abilities count as knowledge of meaning -as it is obvious that many such abilities would not be regarded as adequate. Secondly, it is wrong to say that the system does not know the meaning of "elephant" for it ignores the word's *reference*. There is a (Davidsonian?) sense in which the system does know the reference of 'elephant' : it knows that the word refers to elephants, i.e. to large mammals, proboscideans, living in Africa or India (or zoos), etc. For example, it would certainly be incorrect to say that, for all the system knows, its conclusions might be about flamingos rather than elephants. The system can very well tell elephants (mammals, proboscideans, etc...) from flamingos, which are birds, waders, pink or white (not grey like elephants), etc.

Notice the analogy with knowledge of the sentences' truth-conditions. The analogy extends to the conclusion, concerning the system's actual inability : the system cannot *recognize* elephants in the real world or in a photograph, as little as it can *verify* a sentence about elephants. Whatever the exact nature and importance of such incompetence, I think it underlies our feeling that natural-language understanding systems are only metaphorically such : they do not *really* understand natural language.

In saying this, I am committing myself to two theses. First, if one seriously believes that "we never get out of language", one has no reason to deny that understanding systems understand, aside from their present imperfection (lack of coverage, lack of analytic precision, low speed, etc.). I regard this as a kind of *reductio ad absurdum* of such views as Richard Rorty's, according to which the connection between language and the world does not fall within the scope of semantics. Secondly, a system which made up for this kind of incompetence would indeed understand natural language. Or at any rate, it does not seem that the discussion has revealed *other* reasons for which an artificial system does not understand natural language, beside the system's referential incompetence, as I shall call it.

Semantic competence and recognition

But, what kind of inability is this, and how is it related to semantic competence (or incompetence)? Indeed, is it *at all* related to it? True, the system cannot recognize elephants in the real world or in a photograph ; but neither can I recognize molibneum, or X rays, or a Chippendale chair, which is not to say that I totally ignore the

meanings of such expressions. This objection was used by Yorick Wilks [Wilks, 1982], among others, to reject the idea that knowledge of truth-conditions could be reduced to recognition abilities : "I can surely know enough of the meaning of 'uranium' -Wilks says- to use the word effectively even though I cannot give the performance referred to and know no one who can" -the performance being the recognition of something, X, as uranium. Here, Wilks is claiming that recognitional ability is not a necessary condition of (lexical) semantic competence. His claim can be construed in two different ways, *weak* and *strong*. In the *weak* interpretation Wilks' thesis is true, but ultimately irrelevant to our problem. In its *strong* interpretation, on the other hand, the thesis is false (or so I think), and its falsity is shown by the widespread belief that natural-language understanding systems do not understand natural language.

In the *weak* interpretation, the thesis claims that one can have *some* competence relative to a word although one lacks *full* referential competence *relative to that same word* : I can use the word 'uranium' well enough even though I cannot recognize uranium. In this sense, the thesis is true, I believe. In a couple of articles [Marconi, 1987, 1989], I have tried to argue for a somewhat related view of lexical competence : I have tried to show that lexical semantic competence has two aspects or components -inferential and referential- interacting with and supporting, but nor fully determining each other. Particularly, inferential competence does not by itself engender recognitional ability. No doubt, I may know a lot about uranium without being able to recognize uranium in the real world : if such is the case, I will not denied *some* competence relative the word 'uranium'. However, two remarks are in order at this point. The first and more obvious is the following : inability to recognize does not mean a complete lack of referential competence. I cannot recognize uranium ; but if I am presented -on a tabletop, say- with a fruit I don't know, an animal I never saw before and a bit of uranium and asked to pick the uranium, I will easily do it. As far as uranium is concerned, I no doubt lack full recognitial ability but I do possess some ability to discriminate. It makes sense to suppose that such ability to discriminate is part of referential competence, i.e. of that aspect of competence which underlies the application of language to the world.

Secondly, I would like to point out that in such cases our judgment of the relative importance of referential competence (or even just recognitional ability) for competence as such is highly sensitive to social norms and the distribution of competences and skills throughout the linguistic community. If I cannot recognize uranium but I know that it is an element with heavy atomic weight, radioactive under certain circumstances and so forth, few would say that I don't

know what 'uranium' means. If I know that dolphins are sea mammals frequently spotted even in the Mediterranean etc ; but I do not have the faintest idea of how a dolphin looks like, the linguistic community may have doubts concerning my competence. And finally, if I cannot recognize a cat, people in the community will tend to say that I do not know what 'cat' means, whatever my zoological competence about cats.

Anyway, Wilks' thesis in the weak version just holds that some semantic competence is compatible with the lack of a full recognitional ability. Of course, this does not entail that recognitional ability is *irrelevant* to semantic competence. Neither does it entail that semantic competence is compatible with *complete* recognitional inability with respect to *every* word in a language. What has been granted so far is just the following : a speaker who cannot identify uranium is not thereby disqualified as a competent speaker, not even with respect to the word 'uranium'. We also saw, on the other hand, that words are not all on a par in this respect, and that there is no reason to conclude that generalized recognitional inability would be equally harmless.

In the strong version, Wilks' thesis holds that it is possible to have *full* semantic competence relative to a word although one lacks referential competence even indirectly connected with that word. This is, first of all, very hard to confirm or disconfirm as a thesis about the competence of human beings : as a matter of fact, if one has even limited (not to say full) competence relative to a word one usually has some ability to discriminate in its application, and therefore some referential competence. If I know anything at all about opals I know they are precious stones, so I can tell opals from cats or books. If the word 'pangolin' is not totally foreign to me, I know that pangolins are animals (not planets or Indian military men) ; and so on. But it is essentially our intuitive judgment of *language-understanding* systems which should lead us to reject the strong thesis. These systems have - or may be supposed to have- a very rich inferential competence ; it is precisely because of their lack of referential competence that we tend to deny them genuine semantic competence (we tend to say they don't really understand natural language)[6].

[6] *Of course, here I am assuming as facts that (1) we do deny such systems real competence, and (2) we do so because they lack referential abilities. A supporter of Wilk's "strong" thesis would retort that the reason we (rightly or wrongly) say they don't understand has nothing to do with recognition abilities.*

Reference and recognition

Thus I take it for granted that some recognition ability is an aspect of semantic competence, at least in the case of those words for which recognition is at all relevant. There is, however, a radical objection against using such words and phrases as 'reference' and 'referential

competence' in this context : particularly, against regarding 'recognition (and discrimination) ability' as quasi-synonymous with 'referential competence'. Philosophers of a realist bent think of reference as an objective relation, thoroughly independent of the methods by which we (e.g.) establish that a predicate P applies (or does not apply) to some objects x. The methods by which it is determined that the predicate 'is gold' applies to an object x have changed through history, but the reference of 'gold' has not changed : 'gold' refers, and always reffered, to gold. In some cases, the methods in question may be precarious, yield variable results, vary from person to person or from community to community, or even be unknown : but reference is what it is. Therefore, a philosopher of realist leaning will look with suscipion at my use of the expression 'referential competence' to mean the possession of methods of recognition or discrimination. If one means by 'referential competence' the knowledge of reference -the realist will say- then to have such methods *is not* to know the reference. Knowing the reference of 'gold' is knowing that the word refers to gold : which does not *require* that one can recognize gold by the chemical analyst's or the jeweler's methods. On the other hand, such methods do not *guarantee* that one has access to the reference of 'gold' (witness Putnam's science-fictional examples : we might discover that such methods are and always were defective, that they fall to pick up gold and just gold).

Of course, the methods one usually hints at in these discussions have little in common with the contents of a normal speaker's referential competence. A normal speaker's application of word such as 'cat', or 'water', or even 'gold' is based on rough, macroscopic identification criteria, close to those underlying pattern recognition : not on DNA, or chemical, or spectrographic analysis, i.e. not on the kind of methods that are involved in the realists' discussions. However, one objection from the realist's side applies *a fortiori* to the contents of a normal referential competence : macroscopic recognition criteria are even more conspicuously fallible and unreliable than *scientific* methods. They make us identify hydrogen peroxide as water, iron pyrites as gold, plastic imitation wood as wood ; under certain conditions, even porcelain cats as cats. But of course, the realists say, 'cat' does not refer to porcelain 'cats' nor 'gold' to iron pyrites. It is thus more than ever incorrect to label 'referential competence' a recognition ability which is so far removed from actually identifying reference.

This is not the proper place for a thorough discussion of the realist's conception of reference. I am, however, personally convinced of its many merits. For example, I grant that there are good reasons

for maintaining that the reference of 'gold' has not changed (for the last twenty centuries, say) whereas our criteria for determining whether something is gold have changed : thus there may be good reasons to keep reference and recognition separate. Therefore, one can freely regard the discussions that follow as irrelevant to the issue of reference proper, and of its determination. Here, I am not trying to refute the realist theory of reference but rather to argue for the following thesis : the possession of some referential competence in my sense -i.e. macroscopic recognition and discrimination ability- is a necessary condition of normal semantic competence ; in conjunction with structural[7] and inferential competence, it is also a *sufficient* condition of semantic competence. If a person (or a system) has good inferential competence and good referential competence in my sense, it is hard to deny her semantic competence (at the lexical level). The relation between the content of referential competence in this sense and linguistic reference in the realists' sense remains an open problem ; but the plausibility of my thesis does not hinge on the solution of this problem. Its depends on our intuitive assessment of the actual competence of a system, human or artificial, that would have the aforementioned properties.

[7]By "structural competence" I mean the ability to compute the meaning of a complex expression from its syntactic structure and the meanings of its constituents. B. Partee [Partee, 1980, p.61-2] has meant by "the structural part of semantics" the study of the compositional semantic rules, such as it is carried out in Montague grammar. See also [Marconi, 1987].

Understanding and verification

The idea that to build a system endowed with genuine semantic competence is to enrich a structurally and inferentially competent system with the additional ability to apply words in the real world, and consequently to verify sentences in the real world, may be seen as an attempt at rehabilitating *verificationism,* i.e. the thesis according to which to understand a sentence is to know how to go about verifying it. But today, verificationism is minoritarian, not to say discredited.

Notice, however, that it is not my intention to *identify* semantic competence with the ability to verify. The question is, at most, whether verification abilities are relevant to understanding. I believe they are -indeed, I believe that (recognition and) verification abilities are a necessary condition of understanding- in the limited sense that has been stressed above but needs perhaps to be made clear again. As far as words such as 'cat', 'yellow' or 'walk' are concerned, the inability to verify (under normal circumstances) simple sentences in which they occur would be regarded as evidence of semantic incompetence. Which of course does not mean that the same should be said of such words as 'although', 'eight', or 'function' (thus Goldbach's conjecture or Fermat's last theorem do not come into the

picture). Nor does it mean that recognition (and verification) abilities are a *sufficient* condition of semantic competence.

Still, it could be objected that, even within such limitations, the ability to verify is at most a *symptom* of understanding ; it cannot be a necessary condition. The argument runs as follows. Most cases of understanding are cases of understanding *in absentia*[8] : in most cases, the texts and speeches we understand -daily newspapers, novels, our friends' accounts of their own feats- are not about the scene we have under our eyes at the moment of understanding. In all such cases, verification is simply impossible. There are, indeed, exceptions : there are cases of understanding *in praesentia*. Examples are : reading the instructions for a household appliance while looking at the machine itself and its controls ; obeying an order such as 'Take the book in front of you and read the first line on p. 28' ; listening to a person who is telling us about his medical condition. But such cases, though frequent, are not the most frequent. To account for natural-language understanding is essentially to account for understanding *in absentia* : verification simply does not come into the picture.

[8]More on this in [Marconi, 1990].

Moreover, it has been plausibly argued [Johnson-Laird, 1983, p. 246] that the understanding of fictitious discourse is not essentially different from the understanding of true discourse ; the distinction, all-important as it is in other respects, is irrelevant from the standpoint of language processing. *A fortiori,* one could say, *in absentia* understanding cannot differ in kind from understanding *in praesentia*[9]. So, even in the case of understanding *in praesentia* the possibility of verification cannot be crucial.

[9]If understanding in absentia were qualitatively different from understanding in praesentia , then the understanding of fictitious discourse - which is mostly in absentia- would differ in most cases from the understanding of true discourse, which can be in praesentia.

However, the argument as it stands fails to draw the (obvious) distinction between *not being in a position to* verify a sentence and being factually unable to verify it. Right now, I am not in a position to verify the sentence 'There are six people sitting in the next room', but it would clearly be inappropriate to say that I am factually unable to verify it, or that I don't know how to verify it. The clearest cases of understanding *in absentia* seem to be of this type : they are cases in which one is not in a position to verify whatever is asserted, but would know how to do it (of course, one is usually unwilling to). The same purpose would be served by a distinction between the ability to verify a sentence and the *possibility* of verifying it : I may have the ability without there being the objective possibility, or *vice-versa.* What we lack in the case of *in absentia* understanding is the possibility of verification : which proves nothing concerning our possessing the ability to verify or the role it plays in understanding.

But if what matters (when it does matter) is not actual verification but the ability to verify, why should we want our system to carry out actual verifications? The answer is simple : that is the only way to

effectively show that the system does possess the required abilities. As long as we do not face the problem of actual verification, we shall tend to have systems construct semantic representations (of single sentences or whole texts) which are nothing but formulas of a more or less formal language, themselves in need of interpretation. The only way to build a system to which we would prepared to grant genuine semantic competence is to build a system that can actually verify natural-language sentences. Of course understanding -even understanding *in praesentia*- does not consist in or require actual verification, but there is no better evidence of understanding than actual verification.

A referential machine

A referentially competent system must be able to perceive (typically, to see) the real world, just like us. Therefore, for an artificial system the beginning of referential competence is to be found in artificial vision. However, a possible misunderstanding must be avoided. There is a naive picture of the relation of perception to semantic competence which keeps coming back, in spite of Wittgenstein's attempts at dispelling it and of Putnam's more recent criticism [Putnam, 1981]. In this naive view, part of semantic competence is represented by a certain store of mental *images* associated with words, such as the image of a dog, of a table, of a running man. Thanks to these images we can apply to the real world words such as 'dog', 'table' or 'run' : this is done by *comparing* our images with the output of perception (particularly, of vision). Today, this picture may be somehow supported by reference to prototype theory (although the theory does not license it). Now, the point is not that we do not have mental images : perhaps there are good reasons to believe that we do have something of the kind. The point is that, in the naive picture, the images' *use* in relation to the real world or the perceptual scene is left undescribed. In Putnam's words, "one could possess any system of images you please and not possess the *ability* to use the sentences in situationally appropriate ways... For the image, if not accompanied by the ability to act in a certain way, is just a *picture*, and acting in accordance with a picture is itself an ability that one may or may not have" [Putnam, 1981, p.19]. In other words, in the naive picture the whole explanatory burden is carried by the relation of *comparison* between an image and the perceptual scene, which relation (or process, or whatever it is) is itself unexplained.

Anyway, systems of artificial vision are not organized like that :

there is no store of images to be compared with the perceptual scene. Classes of objects the system can recognize (e.g. tables or cubes) are identified with classes of shapes which are themselves interpreted as *relational structures*, i.e. labelled graphs where the nodes represent object parts and the arcs represent relations between parts : a node is labelled with an ideal property value or a set of constraints on such a value, whereas an arc is labelled with a relation value, or a set of constraints on such a value. For example, a table is identified with a class of shapes expressed by a relational structure, whose nodes represent parts of the table (top, legs) whereas the arcs represent relations between two parts. Node and arc labels are not absolute values, but constraints on possible values. The problem of recognizing a table in a scene is then the problem of "finding subgraphs of the scene graph that are close matches to the object graph, or that satisfy the constraints defined by the object graph" [Rosenfeld, 1988, p. 286]. The scene graph is the result of a sequence of processing stages. In the first stage, the image provided by a sensor is digitalized, i.e. converted into an array of numbers "representing brightness or color values at a discrete grid of points in the image plane" [Rosenfeld, 1988, p.266], or average values in the neighborhoods of such points (elements of the array are called *pixels).* In the second stage *(segmentation),* pixels are classified according to several criteria, such as brightness, or belonging to the same local pattern (e.g. a vertical stroke). In the third stage *(resegmentation),* parts of the image such as rectilinear strokes, curves, angles, etc. are explicitly recognized and labelled. In the fourth stage, properties and relations of such local patterns are identified : both their geometric properties and relations, and (e.g.) the distribution of grey levels through a given local pattern, color relations between two patterns etc. The scene graph's nodes are the local patterns with their properties, and its arcs are the relations among local patterns, with their values. To recognize a table in a scene is thus -as we saw- to find a subgraph of the scene graph which satisfies the constraints associated with the table-graph. In practice, recognition is complicated by several factors : it is hard to make it invariant with respect to different illumination conditions, and 3D vision raises many additional problems. In what follows I shall disregard this kind of problems (though they are of course far from trivial) to focus on others.

From our viewpoint, the relational structure associated with the class of tables, together with the matching algorithm which applies it to the analyzed scene represents the content (or part of the content) of the system's referential competence relative to the word 'table'. If a system were endowed with this kind of competence, plus a minimal

amount of structural semantic competence (to repeat : the ability to determine the meaning of a complex expression from its syntactic structure and the meanings of its constituents) and inferential competence, it could verify sentences such as 'There is a vase on a table', 'There is a vase on *the* table', 'There are two small chairs in front of the table', etc.

A system of this kind is, obviously, a *pattern-recognition* system. If it can recognize telephones at all, it will identify chocolate telephones as telephones. But chocolate telephones are not telephones, for they are imitation telephones, fake telephones. Therefore -it could be objected- a system like ours would not possess genuine referential competence, since the word 'telephone' does not refer to chocolate telephones : to be a telephone is not to have the *shape* of a telephone, and to be a cat is not to look like a cat.

Actually, I am not so sure that a chocolate telephone is not a telephone. But anyway : the point is not whether the system is capable of *foolproof* verifications, but whether its use of language manifests the kind of competence which is characteristic of a human speaker. If one of us -wrongly, perhaps- calls a chocolate telephone a 'telephone', she will not be regarded as *linguistically* incompetent[10]. And as far as shape labels (such as 'cube', 'sphere', 'pyramid') or color words are concerned, it would be hard to deny that our system has the semantic competence of a normal speaker.

[10]*But if, even once told that the telephone in question is a chocolate one, she insisted that, for all she knows, that is a telephone all right, then the question would perhaps arise of her linguistic competence. Such is indeed the system's position : but that is only because its perceptual abilities are supposed to be very limited.*

Beyond pattern recognition

The worst problem is not that just mentioned. The part of the lexicon whose application is essentially governed by pattern recognition is strongly limited. Even the application of a word like 'box' is not based merely on the identification of a shape : and the reason is not simply that there are prism -like boxes, cylindrical boxes, cubic boxes and more, but that it is essential to a box to be a container. A parallelepiped of solid wood, size 25 x 10 x 5 cms, is not a box. A parallelepiped of the same size that has a groove parallel to its basis is not a box either. That an object is recognized (correctly in normal cases) as a box depends on a large amount of knowledge, most of which is not available to a mere pattern recognizer : it depends on the social nature and function of the place where it is located, on the function the object itself can be presumed to play, etc. Or think of 'snuff-box'. A snuff-box has a characteristic shape, but not any object of that shape is likely to be a snuff-box. Here we are dealing with a word whose recognition-procedure must use an

object's presumable *function*. There are many common words which raise the same problem : 'desk', 'ball' (as opposed to 'sphere'), 'dish', 'lever', aerial'.

Besides, there are words to which a definite shape is indeed associated, but their application is (further) restricted by constraints on the *material* they are made of : you can't say 'a plate of cardboard' (one ought to say 'a sheet of cardboard') or 'a plate of wood' (but rather a 'plank' or a 'board'). In cases like this, or that of words having a functional component, the information needed for their correct application can be extracted from the scene, but not by the means that are available to a pattern-recognition system.

The case of Wittgenstein's family-words is different. Their recognition-algorithm is so to speak essentially disjunctive : the application of words such as 'toy' is based on an object's belonging to one of several disjoint sets. Something is called a toy if it is *either* a doll *or* a ball *or* a teddy-bear *or* ... There is not characteristic aspect of a toy, nor is the word's application based only on the identification of a function. A baby may play with a spoon, but this does not make it a toy. Here it seems clear that the term's application depends on the previous application of other words -or anyway, it looks hard for a system to achieve it on any other grounds. In other words, in order to verify the sentence 'There is a toy on the table' the system must either verify at least one sentence in a finite list constructed from the inferential meaning of 'toy', i.e. it must verify either 'There is a doll on the table' or 'There is a ball on the table' etc. —the *top down* method ; or it must infer (more plausibly) 'There is a toy on the table' from (say) 'There is a ball on the table', having verified the latter sentence from recognizing a ball in the appropriate location (the *bottom up* method). Notice that, as of today, the latter method is not really available to artificial vision : systems of vision can (under certain conditions) recognize objects in a scene starting with the objects, not starting with the scene [Rosenfeld, 1988, p. 287-88]. They can determine whether and where in a scene a given object is located starting with the object's definition, but they cannot determine which objects are present starting with a scene's analysis. Thus the more plausible *bottom up* solution is not technically possible, as a matter of fact.

In a sense, the competence associated with family-names can be regarded as indirect : we cannot identify toys, but we can identify (e.g.) dolls and we know that dolls are toys. Such a necessarily (or *grammatically)* indirect referential ability should be distinguished from *contingently* indirect referential ability. This is relevant to the many cases in which we are able to apply a word thanks to a *description*. I may not know the characteristic look of an amanita, but

as I know that an amanita is a mushroom with convex, gilled cap, cylindrical, elongated stem with a ring at about half its height and a kind of small sack (the volva) at its foot, I am (luckily) capable of recognizing amanitas. Of course, that *I* must go through a description doesn't mean that amanitas have no characteristic shape, that there can't be direct referential competence on the subject. By the way, I myself am not a good example for I do have direct competence : I recognize amanitas by the way they look, not by checking the values in a list of properties. An artificial system could go either way. What seems important to remark is that it need not have one direct recognition procedure for each word of the language (for each word, i.e., that has referential content) : like us, it can exploit its inferential competence (in our case, the description of an amanita) to construct an indirect procedure.

The case of relational nouns such as 'uncle' or 'owner' is totally different. These words seem to have referential content, as they are applied to objects in the world. It is no less obvious that their application is not based on form-identification : there is no characteristic aspect of an uncle, or a bachelor. Much of the information thanks to which we recognize an uncle in a photograph does not come from the photograph. We recognize an uncle by recognizing a man of whom we *know* that he is the uncle of X. Similarly, we recognize a bachelor by recognizing a man and knowing -from an independent source- than that man is not married. We recognize the owner of the red car by recognizing a person and knowing that she owns the red car. In all such cases, recognition depends on the interplay between *bona fide* object-recognition and some knowledge base.

On the other hand, these nouns are in a class of their own with respect to all other nouns we mentioned, for the ability to apply them to objects in the real world does not seem to be part of semantic competence. Of one who could not answer the question 'Is there an uncle in this photograph?' we would not want to say he doesn't know what 'uncle' means (whereas we would say so if one could not infer 'X is her uncle' from 'X is her mother's brother'). Thus, this is a case where the notion of referential competence and the classical notion of reference part company. 'Uncle' does have reference -the set of the first elements of the pairs (x, y) such that x is uncle of y. But speaking of "knowing the reference of 'uncle' " as other than knowing the definition of 'uncle', or the inferential meaning of 'uncle', does not seem to make much sense. So, this is perhaps a good observation-point to start reflecting on the relation between reference in the model-theoretic sense and referential competence (which problem I left aside in this paper, as I already stressed).

Obviously, the examples I discussed cover a tiny fraction of the lexicon : I have said nothing of adjectives, for instance, or verbs, or proper names. But I hope to have made my main point clear : in a sense, the referential machine already exists. It is the coupling of a (traditional) natural-language understanding system with a system of artificial vision. In principle, I don't see any special difficulties in making such a system capable of understanding *in absentia* : the recognition procedures can be turned into procedures for the construction of "mental models". If we can verify the sentence 'There is a vase on a table', we ought to be able to construct a model of it. Philosophically, the important point is the following : such a model could be said to be a genuinely *mental* model -not just a translation into some formalism which is itself in need of interpretation- *precisely because* the model's construction procedures are the same as would be activated in order to verify the sentence in the real world[11].

Difficulties lie elsewhere. To endow the system with competence on more than a tiny fraction of the lexicon is very hard, both technically and conceptually. It is also very difficult to disentangle the relation between reference in the classical sense and referential competence. This would require, in the one hand, a clarification of the relation between an individual speaker's competence and the community's competence, or -as I would rather have it- to clarify the *norm-governed* character of semantic competences ; on the other hand, it would require a re-examination of the *vexata quaestio* of the relation between 'X is P' and 'X is (appropriately) called "P" ', i.e. the relation between the alleged objectivity of reference and the intersubjectivity of the communal norms regulating the application of words.

[11] In his article [Woods, 1981], W.A. Woods defined the "meanings of terms" as "abstract procedures built upon a basic set of perceptual primitives that are essentially those of our own direct perceptions... but are treated as if these primitives could be applied in arbitrary contexts of time, space and perceiver". Upon such procedures more abstract (and complex) meaning functions are built : Woods regards them as having "an intensional structure that permits the intelligent system not only to execute them against the external world in particular situations of time and place (with the system itself as perceiver) but also to simulate them in hypothetical situations" (p. 329-330). What I am trying to stress here is that the fact that the model-constructing procedures are essentially the same as those which are "executed against the external world" makes it hard to regard the constructed models as "mere translations" into some uninterpreted formalism, meaning by this that the model -constructing procedures themselves are merely syntactic devices embodying no genuinely semantic knowledge.

Research leading to this paper was partly supported by C.N.R. grant n° 88.01280.08.
I wish to thank Paolo Leonardi and Kevin Mulligan for their comments on a previous version of the present article.

Bibliography

FODOR (J.)
1976, *The language of Thought*, Harvester.

JOHNSON-LAIRD (Ph.)
1983, *Mental models*, Cambridge Univ. Press.

LEWIS (D.)
1970, "General Semantics", *Synthese*, p. 18-77.

MARCONI (D.)
1987, "Two Aspects of Lexical Competence", *Lingua e Stile*, p. 385-395.

1989, "Rappresentare il significato lessicale", in R. Viale (a cura di), *Mente umana, mente artificiale*, Feltrinelli, Milano, p.76-83.

1990, "Alcune riflessioni su senso e condizioni di verità", *Problemi fondazionali nella teoria del significato*.

PARTEE (B.)
1980, "Montague grammar, Mental Representations and Reality", in S. Oelham, S. Kanger, eds., *Philosophy and Grammar*, Reidel, p. 59-78.

PUTNAM (H.)
1979, "Reference and Understanding", in A. Margalit, ed., *Meaning and Use*, Reidel, p. 199-217.

1981, "Brains in a Vat", in *Reason, Truth and History*, Cambridge Univ. Press, p. 1-21.

ROSENFELD (A.)
1988, "Computer Vision", *Advances in Computers*, vol. 27, 1988, p. 265-308.

SEARLE (J.)
1980, "Minds, Brains and Programs", *Behavioural and Brain Sciences*.

WILKS (Y.)
1982, "Some Thoughts on Procedural Semantics", in W.G. Lehnert, M.H. Ringle, eds., *Strategies for Natural Language Processing*, Erlbaum.

WOODS (W.A.)
1981, "Procedural Semantics as a Theory of Meaning", in A.K. Joshi, B.L. Webber, I.A. Sag, eds., *Elements of Discourse Understanding*, Cambridge Univ. Press, p. 300-334.