Marco SANTAMBROGIO
IDC-Università di Bologna

REFERENCE WITHOUT UNDERSTANDING
A Comment on Diego Marconi's *Understanding and Reference*

A natural-language understanding system which not *really* understands natural language cannot be too far ahead —says Marconi[1], striking a very optimistic note. He even claims that such a machine already exists, in a sense. Before pleading for a more cautious view, let me recite Marconi's main argument.

[1]*See in this issue [Marconi, p. 9-25].*

First of all, Marconi asks us to assume that we do know what it is to perform such complex tasks as summarizing any given text, answering questions concerning the topic a text is about, possibly even translating a text into some language other than the one in which it is written. Let us assume, moreover, that our understanding of these performances is sufficient for us to design systems which can exhibit such highly intelligent behaviour. As a matter of fact, we do already have artificial systems which to a certain extent can carry out some of these tasks. True, their performance suffers from serious limitations at present —e.g. the range of texts which they are able to deal with is limited, the resulting summaries are poor, and so on. But Marconi asks us to assume that such systems can be developed to a much greater degree of perfection, so great in fact as to be virtually undistinguishable from humans in the relevant respects. Of course, much research will be needed before such accomplishments are in sight and one may wonder what new theoretical insights will be needed, of which we have as yet no idea. Marconi seems to think that the fundamental architecture of the existing systems is basically correct so that relatively minor improvements will suffice. Perhaps he is overconfident on this point, but it does not really matter whether he is right or not, since nothing much in his argument is meant to hinge on this assumption —or so he thinks. In fact his contention is precisely that, *even if* we had been very successful in building a sophisticated 'understanding' system of the standard type -so sucessful that nothing is left for the traditional artificialist to hope for-

still we could not say that such a system *really* understands the language it can process.

What is it that the system is unable to do? It is not that it does not know the truth conditions of the sentences, for it can tell us, for any sentence E of some given language, that "E is true (in that language) iff f(p1, p2, ..., pn)" where p1, ..., pn are atomic sentences and f is a function it can compute. This only involves (i) some inferential competence (which the system has, by hypothetis) and (ii) some elementary knowledge of the meaning of the predicate "true in language L" —so elementary in fact that it can be imparted once and for all by the general schema " 'E' is true in L iff E". Similarly, it can tell us that "gold" refers to gold, that "John" refers to John and perhaps, as a consequence of its knowing what gold is, even that something is gold if and only if certain conditions obtained. In fact, it knows the meanings of the basic words of semantics, such as "truth" and "reference", just as well as those of a number of non semantic terms, such as "gold" or "cat". Is this not all there is to know about reference, truth and truth conditions? And knowing the truth conditions of sentences, does not the system know their meaning and the meaning of the words occurring in them? Do *we* know more than that?

Yes, says Marconi, we do and therefore such a system would still be *semantically incompetent* in a crucial respect : for instance, it cannot tell whether a given sentence obtains in any particular situation *which it is not given by linguistic description*. Placed in a room, it cannot look around and tell us how many chairs there are, unless it can infer it from some *sentence* given to it. Similarly, there is no way we can show it an object and elicit a linguistic characterization, e. g. "this is gold" or "this is a cat".

Therefore, although it would be wrong to say that the system does not know the reference of any given word belonging to the language it can process, we have to say it is referentially incompetent in that it cannot discriminate among objects in the real world (telling e.g. cats from cows) and describe them in words. On the other hand, this competence is all we must provide the system with in order to make it *really* understand natural language. The final step in Marconi's argument is the remark that to provide a system with such competence is not really beyond our present capabilities : it only takes a good system of artificial vision and other systems of robotics, which those among us who are only interested in language understanding can safely leave to specialists. In Marconi's own words : "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".

This argument seems to me to be, if not invalid, at least seriously incomplete. In the first few steps, a highly sophisticated, though *blind* system is decreed to be unable to *really understand* natural language. Then, at a further stage in the story, vision is bestowed upon it and, lo and behold, its total behavior becomes virtually indistinguishable from our own, for there is nothing which we can do that it cannot do : it moves around and describes the objects it sees and the situation it finds itself in (besides summarizing, translating, answering questions sensibly, etc.). If any behaviour counts as proof of real understanding of language, our own does. Being virtually undistinguishable from us, the system too must have real understanding. So Marconi concludes that all it needed was vision —more precisely, *referential competence,* a capacity to recognize at least some macroscopic objects or at least to discriminate among them, e.g. of telling cats from cows and of telling whether a given object is gold[2].

But is this kind of referential competence really *sufficient?* In order to see that it cannot be, I shall claim that it is even easier than Marconi himself supposes to provide the system with some such competence. What is even more important, this kind of competence is in itself utterly unrelated to linguistic abilities. It is therefore difficult to see how real understanding of natural language could possibly hinge on it.

Let us consider one of those very sophisticated analysers which, in a matter of seconds, are able to tell us of which substance a given specimen is made. Although I only have very rudimentary notions about such machines, I believe that they are so fast, reliable and can recognize such a wide range of substances that we human beings cannot hope to compete with them. I assume that they are perfectly well able to tell gold from fool's gold, water from tea and much, much more. They can also associate such substances with their names -"gold", "water", "tea", etc. - which they can easily print on a screen together with other relevant information. Although I myself ignore the principles of physics and chemistry on which they operate, I believe that there is nothing essentially mysterious about them. In fact, we are surrounded by a large variety of similar highly competent devices : we have machines to tell the time, others for the date, the weather, the speed of a passing car, the blood pressure and so on. I am pretty sure most of us can think of some moderately efficient device capable of telling a cat from a cow.

Let us focus on a fairly common but not so simple kind of machine which is able to obey verbal orders (admittedly, from a rather restricted range) and which can at the same time tell reliably (displaying the appropriate word and number on a screen) the position in space it finds itself in, at least within the accuracy of one floor : the

2 "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 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)" [Marconi, in this issue, p. 9-25]

lift. On being given the order, say "Go to the second floor", by pushing a button showing (some of) these words in English, it first locates itself in space and, in case it is not already on the second floor, it either climbs up or goes down to it —whichever is appropriate. There is every reason to say that it can recognize the situations it finds itself in, and the floors it encounters, even if they are not given to it through linguistic descriptions. Moreover, there is no doubt (for it shows it in use) that it can reliably associate each floor with the right English word and it is able to verify the correctness of such statements as "I am on the second floor now". With respect to a restricted range of English words, it has therefore full referential competence and, should we adhere to Marconi's dictum that "there is no better evidence of understanding than actual verification" [Marconi, in this issue, p. 19], we could confidently say that it *understands sentences* like "I am on the second floor now". Please note that it is not just English that it can so efficiently process, for we can easily switch it to another language by substituing its buttons with others, displaying for instance *Premier étage*, *Deuxième étage*, etc., or *Primo piano*, *Secondo piano*, etc. Note also that it can use those words (obey those orders) with full generality, for it would exhibit the same appropriate behavior in any building where it could be installed, not just in the one for which it was especially designed.

Would a robot capable of carrying out orders such as those considered by Marconi —e.g. "Bring me the hammer, not the pliers" [Marconi, in this issue, p. 9] be very different from the lift as far its referential competence is concerned? Keeping in mind that the possession of some referential competence in Marconi's sense -i.e. macroscopic recognition and discrimination ability- is a necessary condition of normal semantic competence ; in conjunction with structural and inferential competence, it is also a sufficient condition of semantic competence,

It appears that we must concede that all these machines are well on their way towards semantic competence. And yet they clearly do not know the first thing about language (or about anything else, for that matter).

But how can we resist drawing such obviously absurd conclusions, once we admit that knowing the meaning of a statement is tantamount to knowing its truth conditions, and that knowledge of its truth conditions is best manifested by being capable of actually verifying the given statement in a variety of possible circumstances? It is the very conception of meaning as given by the truth conditions which appears to be reduced to absurdity. A parallel reasoning holds for terms, instead of sentences, and reference in place of truth

conditions.

Of course this argument does not go through ; as a matter of fact, it is not too difficult to spot the fallacy. When one says, as is often said, that a subject's knowing the meaning of a sentence must consist in his knowing its truth conditions, one tacitly assumes that the subject knows what truth is, that he already has the concept of truth. Or, assuming that the subject already knows the meaning of the sentences belonging to a given language, his ability to verify them shows that he knows what truth is. What is not possible is to take the subject's practical ability of reacting in the appropriate way to a certain number of utterances as evidence that he is competent on both scores.

Now, of course we cannot assume that artificial systems of the kind Marconi is envisaging already know the meaning of natural language sentences. But perhaps we can ascribe them at least some knowledge of truth and reference : on the one hand they can use *the words* "truth" and "reference" consistently in dealing with texts (in summarizing and translating them, in answering questions about them, etc.), because this is part of their inferential abilities, and there certainly is nothing more difficult or more mysterious in these words than in others on the same level of abstractness. On the other hand, what else is there to know about truth and reference besides the fact that, for every sentence "E", "E" is true if and only if E and for every term "t", "t" refers to t? But this much the system certainly knows, as Marconi remarks early in the paper.

Should we then give up the doctrine that being able to use a word in all the appropriate ways just *is* understanding the concept corresponding to it? In my view, such a move can be seriously considered, but first one has to examine more carefully what our systems (or the lift, for that matter) can do and what they cannot do.

The obvious absurdity of attributing any kind of semantic competence to the lift stems from the fact that it clearly knows nothing of words, language and sentences being used to describe things. If there is anything it "knows", it is floors and buttons being pushed[3]. Intuitively, to know what truth and reference are, one must have some idea of things in the world forming one realm and words of language forming another, and of the latter as standing with the former in the peculiar relationship of "possibly being used to talk about". Not just the lift, but also the systems Marconi is evisaging know nothing of this, as can clearly be seen, among other things, from the fact that there is at least one crucial circumstance in which they are unable to apply such terms as "reference", "truth" and "description". Here one sees that Marconi has a point in insisting that the objects and the circumstances to be recognized by the system need

[3]Not everybody would shrink from attributing referential competence to a lift. For instance, A. Newell, in "Physical Symbol Systems", in D.A. Norman (ed.), Perspectives in Cognitive Science, Norwood, N. J., Ablex, 1981, considers a definition of the term "designate" that would be applicable even to our lift : a symbol S designates an entity E for an agent x (possibly not animate) just so far as, when x takes S as input, x's behaviour depends on what E is. According to this definition, "First floor" does designate the first floor for our lift. The problem is that it also designates electricity, the mass of the Earth, Newton's laws, the material it is made of and much, much more.

not be given to it in language.

Suppose that the system recognizes that some substance it is confronted with is gold and also that it is appropriate for it to utter e.g. (among infinitely many things) "What you showed me is made of gold". That does not mean however that it knows what it is doing, when describing it, is precisely a case of giving a description, i.e. that the very activity it is engaged in right now is a case of *describing* and its own utterance is *true*. That such a deficiency is in fact crucial is seen from the fact that even for humans this kind of self-recognition is all important. Imagine a person, who is, e.g., riding a bicycle and at the same time explaining what riding a bicycle is, without showing any awareness of the fact that he is engaged precisely in the kind of activity he is trying to describe (suppose he is fumbling trying to remember whether one turns the pedals or does something else in order to propel the bicycle) : surely, we would seriously doubt that he knows what he is talking about. This kind of self-awereness is always important but particularly so when description, truth and reference are concerned. And this is why the systems envisaged by Marconi cannot be fully referentially competent. I am *not* saying that those systems cannot be modified so as to acquire that capability ; all I am saying is that (i) merely being able to tell cats from cows or to apply other concepts to objects in the real world (i.e. referential competence restricted to concepts other than the basic semantic ones) will not help them much in this respect, and (ii) a substantially new kind of reflective ability is required, which is not of the kind required in order to summarize, translate and answer questions about texts.

In fact, the basic semantic concepts present a number of problems which are still far from understanding. Suppose that one is given the task of telling cats from cows, assuming that these are the only kinds of things one can be given —i.e., whatever is not a cat must be a cow. Does one need to have the concepts of cat and of cow? Not necessarily : a pair of scales is all that one needs, given that there is no overlapping of the respective ranges of weight ; one can then arrange the scales in such a way that they display, say, the appropriate flag when either animal is present of course, in order to *invent* the mechanism, one must see that it will work correctly in general, and therefore one must have all the relevant concepts. But the pair of scales need not. Similarly, our lift need not have either the concept of floor or that of number (not to mention those of imperative and obedience). Now, at first sight at least, the same strategy seems always to be applicable to the concepts of truth and reference : in order to evaluate the sentence "A is a true statement" all one has to evaluate is A —the truth predicate has disappeared. Hence the claim that Tarski's Convention T is all that one (or an artificial system) has

to know about truth. Just as, in order to tell cats from cows, there seems to be no need for the concepts of cat and of cow, in the same way the semantic concepts seem to be redundant. But this is absurd, as I have tried to show.

Be this as it may, let me end my comment on a more optimistic note. I too, like Marconi, feel that a system which really understands natural language cannot lie too far ahead. All we have to supply it with is some understanding of such notions as reference, truth and natural language. But first, of course, we must *really* understand these notions ourselves[4].

[4]Marconi's paper [Marconi, in this issue, p.9-25] makes a number of interesting points, besides those I mentioned. On some of them I feel I might have something to contribute but cannot do so because of limited space. Let me only mention that the issue of realism vs anti-realism seems to me to be crucial - much more so, in fact, than Marconi himself seems to believe. In particular he seems to think that for the purposes of understanding natural language, one can safely put off making a choice : perhaps we could have both realist and anti-realist systems. I, on the contrary, now feel that anti-realist systems could not possibly exist -which is not to claim that the former can.