Why we do not have to worry about speaking the language of the computer

Hubert L. Dreyfus University of California, Berkeley, California, USA

Communication is not the exchange of information

If we are to believe the most optimistic predictions from the field of artificial intelligence, computers like HAL in 2001 will speak our language so we will not have to speak theirs. That would be an ideal state of affairs. Since we have to deal with computers, it would be convenient if we could talk to them in ordinary language. But that is not the current state of the art, and, although it is not in principle impossible, there are good reasons to doubt that, using current techniques, such communication will ever be achieved or even approached.

To understand why there may never be good natural language interfaces we have to review the hopes and failures surrounding machine translation. Around 1950 researchers thought that to achieve fully automatic high quality translation one had merely to put a lexicon and the rules of grammar of two languages into a computer. But by 1966 the effort had utterly collapsed. A government report concluded: "there is no immediate or predictable prospect of useful machine translation" (National Academy of Sciences, 1966, p. 32).

This attempt at machine translation failed because, as Bar-Hillel (1964) pointed out, language is highly ambiguous. People, however, are not bothered by this ambiguity. Bar-Hillel thought human beings resolve linguistic ambiguity by using a huge encyclopedia of facts. He gave as an example the sentence: "The box was in the pen", from the context:

Little John was looking for his toy box. Finally he found it. The box was in the pen.

And argued that:

What makes an intelligent human reader grasp this meaning so unhesitatingly is ... his knowledge that the relative sizes of pens, in the sense of writing implements, toy boxes, and pens, in the sense of playpens, are such that when someone writes under ordinary circumstances and in something like the given context, "The box was in the pen", he almost certainly refers to a playpen and most certainly not to a writing pen (Bar-Hillel, 1964, p. 94).

It might seem from Bar-Hillel's example that the computer need only check the immediate verbal context for words such as "toy" to determine that "playpen" is the relevant reading. But a little modification of the example will show that contextual analysis cannot get around Bar-Hillel's objection: "Little John was playing on the floor besides his pen. He was drawing a picture with a red pen on some green paper. When the drawing was finished, he looked for his toy box, and found it inside the pen". It is conceivable that the first two occurrences of • MCB University Press, 0959-3845

Information Technology & People

Speaking the language of the computer

281

"pen" could be disambiguated with information from the surrounding words. But it is clear that since the clues to both meanings of pen are in the immediate verbal context of the last sentence (indeed, in the sentence itself), the disambiguation of the last occurrence of "pen" requires the "common knowledge" that Bar-Hillel has in mind.

A few years later I took the argument one step further, suggesting than an appeal to context is more fundamental than an appeal to facts, for the context determines the significance of the facts. Thus in spite of our general knowledge about the relative size of pens and boxes, we might interpret "The box is in the pen" when whispered in a James Bond movie, as meaning just the opposite of what it normally would mean. But even this formulation is not quite right. Now I would say that thanks to the way human beings are always in a context the ambiguity is not resolved at all. It simply does not arise. That is why the possible ambiguity of syntax and semantics was not noticed until much time and money had been wasted trying to achieve machine translation.

The basic error behind the idea of machine translation was already pointed out by Heidegger in 1927. It was the notion accepted from Descartes to Husserl that language corresponds to or represents facts about the world and is used to communicate this factual information from one user to another. Heidegger denies both premises of this information-conveying account. Natural language is not a mathesis-like representation in which discrete elements represent discrete features of the world. It contains metaphors, pronouns, indefinite articles, etc., as well as latent ambiguity. Thus natural language is not at all the unambiguous representational system required to convey information from one subject to another subject[1]. This, however, presents no problem in ordinary language use, for, as Heidegger points out in *Being and Time*, language is not normally used for the exchange of information. Rather language calls attention to some aspect of a world the language users already share.

Communication is never anything like a conveying of experiences, such as opinions or wishes, from the interior of one subject into the interior of another ... In discourse Being-with becomes "explicitly" *shared*, that is to say, it *is* already, but it is unshared as something that has not been taken hold of and appropriated (Heidegger, 1962 [1937], p. 205).

Programming keeps one honest, so those who were trying to get computers to translate natural language were forced to recognize this phenomenon. They saw that if one is to resolve ambiguity, one must establish an agreed on context. So around 1970 researchers in artificial intelligence developed ways of representing isolated contexts or micro-worlds in which each word had only one meaning. The most successful example was Winograd's blocks world, SHURDLU. The program was designed to carry on a dialog with a person (via teletype) concerning the activity of a simulated robot arm in a tabletop world of toy objects (often referred to as the "blocks world"). The program could answer questions, carry out commands, and incorporate new facts about its world. Researchers soon developed micro-worlds which were stereotyped versions of more everyday contexts. Roger Schank called these context scripts and used restaurant behaviour as his example; Marvin Minsky called them frames, using

282

ITP

11,4

as his example a birthday party. Each showed that you could view events in the everyday world as taking place in many such stereotypical situations.

But at this point the fundamental problem of establishing a context reemerged. How was the computer to determine in any actual case of translation or communication which context was relevant, i.e. which script or frame to use? It became clear that to do this one needed a meta-frame which contained all the contexts of everyday life – a formal substitute for what Heidegger called the phenomenon of world. The shared world is the everyday shared background that is not communicated but which makes communication possible. If this broadest context could be made explicit and fed into the computer then, since much of our behaviour is, indeed, stereotyped-articulated by *Das Man* Heidegger would say – the computer could call up the appropriate frame at each point, so that for computers, just like for people, the problem of ambiguity would rarely arise.

Heidegger, however, denied that the shared world presupposed in communication could be represented as an explicit and formalized set of facts. As he put it:

The context of ... references, which, as significance, is constitutive for worldhood, can be taken formally in the sense of a system of relations. But ... [t]he phenomenal content of these "relations" and "relata" ... is such that they resist any sort of mathematical functionalization; nor are they merely something thought, first posited in an "act of thinking". They are rather relationships in which concernful circumspection as such already dwells (Heidegger, 1962 [1937], pp. 121-2).

Artificial intelligence researchers try to avoid the problem raised by Heidegger by making a move that Husserl had already made in response to *Being and Time*. He claimed that the world, the background of significance, the everyday context, was merely a very complex system of beliefs. Thus one could, in principle, suspend one's dwelling in the world and achieve a detached, description of the human belief system. As Husserl put it in *Crisis:*

The life which effects world-validity in natural world-life does not permit of being studied from within the attitude of natural world-life. What is required then, is a *total* transformation of attitude, a *completely unique, universal epoche* (Husserl, 1970, p. 148).

Thus, in the ideal limit the transcendental phenomenologist could complete the task that had been implicit in philosophy since Socrates, *viz.* one could make explicit the beliefs and principles underlying all intelligent behaviour. As Husserl put it:

[W]e move in a current of ever new experiences, judgments, valuations, decisions. ... None of these acts, and none of the validities involved in them is isolated: in their intentions they necessarily imply an infinite horizon of inactive validities which function with them in flowing mobility. The manifold acquisitions or earlier active life are not dead sediments; even the background ... of which we are always concurrently conscious but which is momentarily irrelevant and remains completely unnoticed, still functions according to its implicit validities (Husserl, 1970, p. 149).

[These implicit validities] are included as *sedimented history* in the currently constituted intentional unity ... – a history *that one can always uncover by following a strict method* (Husserl, 1969, p. 245).

Speaking the language of the computer

ITP 11,4	Since he firmly believed that the shared background could be made explicit as a belief system Husserl was ahead of his time in looking toward the possibility of artificial intelligence. He notes:
284	The pressing question is whether within the eidetic domain of (phenomenologically) reduced phenomena there could not also be, <i>alongside</i> the descriptive, an idealizing procedure that substitutes pure and strict ideals for intuited data and that would serve as the basic medium for a mathesis of experience (Husserl, 1913, No. 75, p. 174).
	Of course, the task of writing out a complete account of everyday life turns out to be much harder than initially expected. During his 25 years of trying to spell out the components of the subjects' representations (the <i>noema</i>), Husserl found that he had to spell out more and more of the subject's common-sense understanding of the everyday world. As he laments:
	To be sure, even the tasks that present themselves when we take single types of objects as restricted clues prove to be extremely complicated and always lead to extensive disciplines when we penetrate more deeply. That is the case, for example, with a transcendental theory of the constitution of a spatial object (to say nothing of a Nature) as such, of psycho-physical being and humanity as such, culture as such (Husserl, 1960, pp. 54-5).
	He spoke of the <i>noema</i> 's "huge concreteness" (Husserl, 1969, p. 244) and of its "tremendous complication" (p. 246) and he sadly concluded at the age of 75 that he was a perpetual beginner and that phenomenology was an "infinite task" (Husserl, 1970, p. 291). There are hints in his frame paper that Minsky has embarked on the same "infinite task" that eventually overwholmed Husserl. He too complains:
	Just constructing a knowledge base is a major intellectual research problem We still know far too little about the contents and structure of common-sense knowledge. A "minimal" common-sense system must "know" something about cause-effect, time, purpose, locality, process, and types of knowledge We need a serious epistemological research effort in this area (Minsky, 1981, p. 124).
	Minsky's naïveté and faith are astonishing. Transcendental phenomenology is just such a research effort. Indeed, philosophers from Socrates to Husserl, who uncovered all these problems and more, have carried on serious epistemological research in this area for 2,000 years without notable success. The common knowledge problem has blocked all progress in theoretical artificial intelligence for the past 25 years (Dreyfus, 1992). Winograd was one of the first to see the limitations of SHURDLU and all scripts and frames generalizations of the micro-worlds approach. Having "lost faith" in artificial intelligence, he now teaches Heidegger and Gadamer in his computer science courses at Stanford, and points out "the difficulty of formalizing the commonsense background that determines which scripts, goals and strategies are relevant and how they interact". Since natural language understanding requires that human beings and the computer share a common-sense context it requires the solution of the common- sense knowledge problem. Consequently, Winograd is skeptical about the possibility of giving a computer natural language. "The limitations on the formalization of contextual meaning", he says, "make it impossible at present –

and conceivably forever – to design computer programs that come close to full mimicry of human language understanding" (Winograd, 1984, p. 142).

Heidegger's view, which I share, is that the common-sense knowledge problem is not only incredibly difficult, it is misformulated from the ground up. As Husserl saw, making background knowledge explicit only makes sense if it is already implicit knowledge. But the everyday context which forms the background of communications is not a belief system or a set of rules and principles at all, not even a highly complex, implicit one, but is rather a set of social skills, a kind of know-how, any aspect of which makes sense only on the rest of the shared social background[2]. Heidegger also calls this background "familiarity" of our shared "understanding of being". Making this background explicit in terms of a set of beliefs – of knowing that – which makes no appeal to this background is not an infinite task but a task one cannot even begin.

The incredible difficulty faced by artificial intelligence researchers in spelling out our supposed implicit belief system is an artifact produced by supposing that the phenomenon of world is in principle representable, i.e. that our everyday understanding is common-sense knowledge. Artificial intelligence researchers (as well as philosophers who postulate an ideal communication situation in which all the shared background would be exhaustively represented or at least be representable when necessary for understanding) are examples of the perennial philosophical failure to distinguish the world from the sort of things – facts, assumptions, beliefs, principles – which show up on the background of the world. In Heideggger's terms this is the latest and most extreme example of metaphysics - the failure to appreciate the ontological difference – the difference between being and beings. In sum, the artificial intelligence goal of communication with the computer in natural language, by means of the exchange of information on a background which is in principle analysable as merely more information, is a chimera. The misguided nature of this goal remains hidden from us because it is the culmination of our own philosophical tradition.

Practical consequences

We have seen that natural language is not the precise representational medium required for the pure exchange of information, but that, given their shared background practices, the ambiguity, etc. of natural language presents no problem for human language users. We have followed the failure of attempts to enable the computer to understand natural language by giving it a theory which makes explicit the background understanding shared by human beings. If one still wants to exchange information with a computer only one alternative remains. One must create an artificial language which is specified completely and unambiguously. All the symbols in this language will have to refer uniquely to specific decontextualized features of a specific domain. The programmer will have to perform this decontextualization and the user will have to learn just which decontextualized features of the environment the symbols in the computer language represent.

Speaking the language of the computer In current computer user interfaces this artificial information exchange is implemented in one of three ways:

- (1) One can simply list the decontextualized symbols and operations that have been chosen by the programmer. The user can then learn them and by using them appropriately get the computer to do what he wishes. This has been the common mode of interaction up to now. But people want more user-friendly interfaces, i.e. interfaces which give the illusion of being cases of linguistic communication.
- (2) An honest way of facing this demand has been pioneered by Texas Instruments in their recent Naturallink program. In this system the user communicates with the computer by selecting syntax and vocabulary from menus that the computer presents. Thus the user uses an artificial approximation to natural language provided by the programmer. He does not have to memorize this code since he is constantly presented with his options on the screen.
- (3) A less straightforward but more seductive alternative is for the computer to accept natural language input. It then uses artificial intelligence techniques to parse the sentences and a lexicon to assign meaning. A typical version of this approach is Q & A developed by Symanteck. In such a system, if the user is lucky enough to use syntax and vocabulary already in the machine and if he manages to avoid semantic ambiguity, metaphor, ambiguous pronoun reference, etc., after a minute or so of processing the machine will respond appropriately. If the user is not lucky enough to stay within the limits of the code, the machine will type out that it cannot understand and the user will have to modify his input and try again. Approach three thus turns out to be a disguised version of approach one. The user who does not want to waste his time will soon learn the limited syntax and dictionary programmed into the machine. As Winograd and Flores put it:

[T]here are cases in which the appearance of natural language can make a computer system seem less formidable, encouraging use by people who would resist a more visibly formal approach. [But] it may well be more efficient for a person to learn a specialized formal language designed for that purpose, rather than learning through experience just which English sentences will and will not be handled. When interacting in natural language it is easy to fall into assuming that the range of sentences that can be appropriately processed will approximate what would be understood by a human being ... Since this is not true, the user ends up adapting to a collection of idioms – fixed patterns that experience has shown will work. Once the advantage of flexibility has been removed, it is not clear that the additional costs of natural language (verbosity, redundancy, ambiguity, etc.) are worth paying in place of a more streamlined formal system (Winograd and Flores, 1986, p. 129).

Social implications

The inevitable artificiality of computer languages leaves us with two social problems: one trivial and one slightly more serious. Since we cannot use our

286

ITP

11,4

natural language to communicate with computers, and there is no advantage in using computer language to communicate with human beings who share our background practices, we are entering a situation in which we may need to learn both a natural language and some special computer-communicating skills. For example, since artificial intelligence researchers have failed to make computers understand cursive writing, we may all have to learn how to write our checks in machine readable script.

As for the view that someone who learns a rigid, impoverished, unambiguous contextless and inhuman computer code will be tempted to use it in his or her daily interactions, this seems to me as unlikely as that mathematicians and physicists would talk to each other in formulas when they were deciding where to eat lunch. We already possess many context-free technical languages but they have not affected the way people talk in their homes, in the market place, or even in the laboratory. To ask to what degree computer language will impoverish our everyday use of language is to suppose computer code and human language exist on the same dimension. Such a supposition succumbs to the traditional view of language as communication between subjects and forgets the marvellous power of our shared background practices which allow us to use language to point out aspects of our shared world.

Another more plausible but ultimately mistaken view is to think of the introduction of computer codes as comparable to the advent of writing. Writing did decontextualize speech. Speech points out aspects of a shared situation, but a written text can be read in any situation whatsoever. But this analogy, while interesting, is nonetheless misleading. A natural language is understandable only in the context of the shared human world. Thus the move from local context to human world context which takes place in passing from speaking to writing is radically different than the move from natural language to machine code which eliminates our shared world altogether.

The real danger

Computers process "information", context-free facts, according to procedural rules. That means computers must reason things out rather like inexperienced persons, whereas for human beings experience leads to know-how – a far superior holistic, intuitive wisdom that cannot be imitated by rule-following computers (Dreyfus and Dreyfus, 1986).

The increasing bureaucratization of society heightens the danger that in the future intuitive wisdom will be lost through overreliance on rationality. Today, as always, individual decision-makers understand and respond to their situation intuitively. When time permits, they further validate and fine-tune their intuitions using deliberative rationality. But increasingly the success of science and the availability of computers have led more and more toward that explicit, detached mode of problem description and evaluation of alternatives sometimes called calculative rationality.

This movement towards self-conscious, rational justification sounds enlightened and progressive until one realizes that genuine know-how, wisdom, Speaking the language of the computer

287

and good judgment are sacrificed in the process. Any attempt to be explicit and logical, so that rational discussion might be directed toward the relevance and validity of isolated elements (factors, attributes, cues, etc.) and to abstract principles, limits "judgments" to the choice of those elements and principles. But with experience comes a decreasing concern with accurate assessment of isolated elements and abstract principles, so when experts are asked to judge isolated factors and justify these conclusions they are shorn of their expertise.

It is often desirable that experts defend their recommendations against other experts, or in some way be cross-examined so that those affected can question the expert's presuppositions. If this is taken to mean that the expert must articulate his values, rules, factual assumptions, and principles, examining becomes a futile exercise in rationalization in which expertise is forfeited and time is wasted. But the alternative need not be the imposition of unquestioned authority. In Japan, consensus seems to be reached through discussion without reducing intuition to rationalization. The cross-examination of competing experts in an intuitive culture seems to take the form of a conflict of interpretations in which each expert is required to produce and defend a coherent narrative which leads naturally to seeing things intuitively from his point of view.

The question for our society now is whether we are going to accept the view of man as an information processing device, or whether we are still enough in touch with our pre-Cartesian essence to realize the limits of the computer metaphor. With our mechanical contrivances now able to solve certain problems more effectively than we can, we are being forced to rethink some very old and by now very basic aspects of our self-image. Let us hope that the rethinking will lead to a new definition of who we are, one that values our capacity for involved, intuitive wisdom more than our ability to be rational animals.

What we do now will determine what sort of society and what sort of human beings we are to become. If we think of ourselves only as repositories of factual knowledge and of information processing procedures, then we understand ourselves as someday to be surpassed by bigger and faster machines running bigger and more sophisticated programs. We will still have a natural language because we will still have a shared background of skills, feelings, bodies, and social artifacts, but if we have educated ourselves to use calculative rationality on all occasion we will have no content to talk about but the facts and procedures we share with our machines. All use of language to convey convictions and intuitions or to move people to see things in a new way will be disparaged for not conveying any information. In such a rationalized culture everything important will, indeed, be convertible into code.

Notes

- 1. Here subject means any self-contained entity with an internal representation of the world and so includes human beings as understood since Descartes as well as computers programmed for conventional artificial intelligence.
- 2. For the argument that the background is best understood as shared social skills, and that there is no reason to believe skills can be captured in theory, see Dreyfus and Dreyfus (1986).

288

ITP

11,4

References	Speaking the
Bar-Hillel, Y. (1964), "The present status of automatic translation of language", in Alt, F. (Ed.) <i>Advances in Computers,</i> Academic Press, New York, NY.	language of the
Dreyfus, H.L. (1992), <i>What Computers Still Can't Do: A Critique of Artificial Reason</i> , The MIT Press, Cambridge, MA.	computer
Dreyfus, H.L. and Dreyfus, S.E. (1986), <i>Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer,</i> (Paperback edition, with Athanasiou, T.) The Free Press, New York, NY.	289
Heidegger, M. (1962 [1937]), <i>Being and Time</i> (trans. John Macquarrie and Edward Robinson), Basil Blackwell, Oxford.	
Husserl, E. (1913), <i>Ideen zu einer reinen Phanomenologie und phenomenologischen Philosophie</i> , M. Nijhoff, The Hague.	
Husserl, E. (1960), Cartesian Meditations, M. Nijhoff, The Hague.	
Husserl, E. (1969), Formal and Transcendental Logic, M. Nijhoff, The Hague.	
Husserl, E. (1970), <i>The Crisis of European Sciences</i> , Northwestern University Press, Evanston, IL.	
Minsky, M. (1981), "A framework for representing knowledge", in Haugeland, J. (Ed.), <i>Mind Design</i> , The MIT Press, Cambridge, MA.	
National Academy of Sciences (1966), <i>Language and Machines</i> , National Academy of Sciences, Washington, DC.	
Winograd, T. (1984), "Computer software for working with language", <i>Scientific American</i> , September.	
Winograd, T. and Flores, F. (1986), <i>Understanding Computers and Cognition: A New Foundation for Design</i> , Addison Wesley, Reading, MA.	