METHODOLOGY IN AI AND NATURAL LANGUAGE UNDERSTANDING

Yorick Wilks
Istituto per Gli Studi
Semantici e Cognitivi
Castagnola, Switzerland

Are workers in AI and natural language happy band of brothers marching with their arious systems together towards the romised Land (systems which in the view of ahy well disposed outsiders are only otational variants at bottom), or on the ontrary are there serious methodological ifferences inherent in our various ositions? I think there is in fact one entral difference, and that it is a ethodological reflection of a metaphysical ifference about whether there is, or is ot, a <u>science of language</u>. But it is not asy to tease this serious difference out rom the skein of non-serious methodological iscussions.

By "non-serious methodological etc." I ean such agreed points as that (i) it would e nicer to have an understanding system orking with a vocabulary of Nk words rather han Mk, where N>M, and moreover, that the ocabularies should contain words of aximally different types: so that "house", 'fish", "committee" and "testimonial" would e a better vocabulary than "house", 'cottage", "palace" and "apartment block." nd that, (ii) it would be nicer to have an nderstanding system that correctly nderstood N% of input sentences than one hich understood M%. When I say non-serious ere I do not mean unimportant, but only hat nothing theoretical is in question; so hat, for example, it could be only an rbitrary choice whether or not a system hat understood correctly 95% of sentences rom a 3000 word vocabulary was or was not etter than one which understood 98% from a 1000 word vocabulary.

Indeed, the very sizes of the vocabularies and success rates in the example show that such a choice, however arbitrary, is not one we are likely to be called upon to make in the near future, so let us press a little deeper.

Consider the following three points, which I will name for ease of subsequent reference:

(1) <u>Theory and practice</u>: "Trying hard to make a system work is all very well, but it's too success-oriented, what we need at the moment is more theoretical work".

(2) <u>AI and science</u>: "What we are after is the right set of rules, and expressions of real world knowledge, for understanding natural language: no approximate, 95%, solutions will do, just as they won't do in physics".

(3) <u>Where to start</u>: "Since difficult examples clearly require reasoning to be understood, we cannot even begin without such a theory because, without it, we could not know of even an <u>apparently simple example</u> that it did NOT require reasoning in order to be understood."

The above three positions are not intended to be a parody, and certainly not a parody of anyone in particular's views. I have not in fact heard all three from the same person, even though, in my view, they constitute a coherent position taken together: one which I believe to be not only wrong, and I will come to that, but also harmful. Let me deal with the sociology first, and in the form of a very crude historical generalization.

It is clear that "natural language understanding" has come to occupy a less peripheral place in AI, and much of the credit for this must go to Winograd (1972). The position, expressed in (1), (2) and (3) above, is in some ways a reaction to that, and in my view an excessive one. Behind the positions above lurks the suspicion that the success of Winograd's system was in part due to its oversimplificatons and that we must now be wary, for a while at least, of applications, successful or otherwise: that we must, in short, emphasize how difficult it all is.

Now there is undoubtedly something in this, but it seems to me that the reaction may have the paradoxical effect of causing the study of natural language in AI to be given up altogether. In the last year or two a number of those who seemed to be concerned with the problems of natural language no longer seem to be so. There has been a subtle change: from the analysis of stories, or whatever, to the setting out of systems of plans which now seem to <u>construct</u> stories as they go along. It might then seem natural to move further: from the production of stories about tying one's shoe:laces, shopping in supermarkets, etc. to plans, for robots of course, that will actually shop in supermarkets, tie their own shoe-laces, play diplomacy or whatever. And then of course we are back where we started in AI: back to AI's old central interests, robots, problem-solving and the organization of plans.

All this would be a pity, not only because someone has, as always, to be left holding the baby of natural language analysis, but because it is too soon, and AI has not yet had the beneficial effect it is capable of having, and ought to have, on the study of natural language. There are at least four of these benefits; let me just remind you of them:

(i) emphasis on complex stored structures in a natural language understanding system: frames, if you like (Minsky 1974)

(ii) emphasis on the importance of real world, inductive knowledge, expressed in the structures of (i)

(iii) emphasis on the communicative function of sentences in context, i.e. the finding of the correct-in-context reading for a sentence, as opposed to the standard linguistic view, which is that the task is the finding of a range of possible readings, independent of context

(iv) emphasis on the expression of rules, structures, and information within an operational/procedural/computational environment.

Conventional linguistics has still not appreciated the force of these points, which are of course commonplace in A.I.

Let me now turn to the position sketched out earlier under three headings, and set out some countervailing considerations. It should be made clear that in whaa follows I am making only methodological points aout the assessment of systems in general. No attack on the *content* of anyone´s system is intended.

First, to the *theory and practice* point. It seems to me worth emphasizing again that there can be no other ultimate test of a system for understanding natural language than its success in doing some specific task, and that to pretend otherwise is to introduce enormous confusion. Considerations of logic or psychological plausibility may indeed be suggestive in the construction of AI language systems, but that is quite another matter from their ultimate accountability, which can only be whether or not they work. Suppose some system had all desirable logical properties, and had moreover been declared by every respected psychologist to be consistent with all known experiments on human reactions times and so on. Even so, none of this would matter a jot in its justification as a computational system for natural language.

In a similar vein, it seems to me highly misleading, to say the least, to describe the recent flowering of AI work on natural language inference, or whatever, as *theoretical* work. I would argue that it is on the contrary, as psychologists insist on reminding us, the expression in some more or less agreeable semi-formalism of intuitive, common-sense knowledge, revealed by introspection. I have set out in considerable detail (Wilks 1974) why such an activity can hardly be called "theoretical", in any strong sense, however worthwhile it may be. That it *is* worthwhile is not being questioned here. Nor could it be, since I am engaged in the same activity myself (Wilks 1975b). I am making a meta-, methodological, point that the activity does not become more valuable by being described in value-added terms. The worthwhileness, of course, is shown later by testing, not by the intuitive or aesthetic appeal of the knowledge represented or the formalism adopted.

Let me turn to position (2): *AI and Science*. It seems clear to me that our activity is an engineering, not a scientific, one and that attempts to draw analogies between science and AI work on language are not only overdignifying, as above, but are intellectually misleading. Conduct with me, if you will, the following *Gedankenexperiment*: suppose that tomorrow someone produces what appears to be the complete AI understanding systems, including of course all the right inference rules to resolve all the pronoun references in English. We know in advance that many ingenious and industrious people would immediately sit down and think up examples of perfectly acceptable texts that were not covered by those rules. We know they would be able to do this just as surely as we know that if someone were to show us a boundary line to the universe and say "you cannot step over this", we would promptly do so.

Do not misunderstand my point here: it is not that I would consider the one who offered the rule system as refuted by such a counter-example, particulary if the latter took time and ingenuity to construct. On the contrary, it is the counter-example methodology that is refuted, given that the proffered rules expressed large and interesting generalizations and covered a wide range of examples. For the simple methodology of refutation is the method of idealised science, where one awkward particle can overthrow a theory*. In the study of language such a methodology is no more appropriate than it is to consider the definition of fish as something that swims and has fins as being "overthrown" by the discovery of a whale. Of course it is not, nor does the definition lose its power; we simply have special rules for whales.

The fact of the matter is surely that we cannot have a serious theory of natural language which requires that there be some boundary to the language, outside which utterances are too odd for consideration. Given sufficient context and explanation *anything* can be accommodated and understood: it is this basic human language competence that generative linguistics has systematically ignored and which an AI view of language should be able to deal with. We know in principle (see Wilks 1971 and 1975a) what it would be like to do so, even if no one has any concrete ideas about it at the moment*: it would be a system that could discover that some earlier inference it had made was inconsistent with what it found later in a text, and could return to try again to understand. And here, to be interesting, the backtracking would have to be more than simply the following of some

*The bad influence may not come directly from science, but via "competence theory" in linguistics.

*Winograd´s thesis, of course, had a system for checking inferences and new information against all that it knew already, though it is not clear that such a direct method would extend to a wider world of texts. In (Wilks 1968) there was a very crude program for finding out that an assignment of sense, earlier in a text, had gone wrong, but it was almost certainly an inextensible method.

ranch of a parsing that had been ignored arlier: it would have to be something quivalent to postulatng a new sense of a ord, a new reference of a pronoun, or even new rule of inference itself. It is urely these situations that the "AI aradigm of language understanding", and erhaps it alone, will be capable, in rinciple, of tackling, in the future, and t is these features of language, that equire such maneuvres, that show most learly why the "100%-Scientific Rule" icture does not fit language at all, and hy time spent trying to make it fit may be diversion of attention from really key reas like the heuristics of isunderstanding and contradiction.

Perhaps a moment's further dilation on he role of counter-examples is worthwhile ere. Consider two counter-examples: one roduced against the "expectation as basic echanism of parsing" hypothesis of Riesbeck Riesbeck 1974), and one against my own preference as basic mechanism etc." (Wilks 975c) hypothesis. Riesbeck considers entences such as "John went hunting and hot a buck", where, putting it simply, the oncept of hunting causes the system to xpect more about hunting and so it resolves buck" correctly as the animal and not the ash. One then immediately thinks of "John ent hunting and lost fifty bucks".

Conversely, in my own system I make uch of the preference of concepts for other oncepts to play certain roles, so that for xample in "John tasted the gin", "gin" will e resolved as the drink and not the trap, ecause of the preference of tasting for an dible or potable object like the liquid in. Someone then, plausibly enough, comes p with "He licked the gun all over and the tock tasted good", where the preference on small scale would get the wrong "soup" ense of "stock", and not the "gun part".

It should be clear that these ounter-examples are to what appear to be, uperficially, opposed theories of parsing. y point is that in neither case do the xamples succeed in showing a theory seless, i.e. neither "preference is no ood" nor "expectation is no good" follow rom the production of the counter-examples. hat is needed of course, and what in fact oth parties are trying for, is some uitable mixture of the approaches. But, nd here is the key point, there will not be ny magic right mixture either. There can nly be a combination that will itself go rong with sufficiently ingenious examples. nly a recovery mechanism will save us, just s it saves people, who misunderstand all he time. There will never be, nor could here be, a RIGHT combination, in the way hat $F = \frac{km_1m_2}{r^n}$ gives a right theory of ravitation when, and only when, n = 2 .

Finally, let me turn to the third spects of the initial position, which I alled where to start. This brings up the ery difficult question about the relation f reasoning to natural language, and I have ade some remarks on that in the paper in ection 2 on "Primitives". Here I just want to try and counter, in a brief and inadequate manner, what I see as the bad effects of the where to start view.

The view is an alternative to a more simple-minded view which goes as follows: "we should now concentrate on difficult examples, requiring reasoning, when studying natural language understanding, because the basic semantics and syntax have been done, and we are therefore right to focus on the remainder". This view is simply historically false about what has been done, so let us leave that and turn to the much subtler where to start view which holds that, on the contrary, the basic semantics of natural language understanding have not been done and cannot even be started without a full theory of reasoning capable of tackling the most difficult examples, because, without such a theory, we can't know that it isn't needed, even in the apparently simplest cases. The argument is like that against the employment of paramedical staff as a front line in community medicine: we cannot have a half-trained doctor treating even influenza, because unless he's fully trained he can't be sure it isn't pneumonia.

One obvious trouble with the argument, in both its linguistic and medical forms, is its openness to reductio ad absurdum replies. It follows from that position, if taken seriously as a theory of human understanding, that no one understands anything until they are capable at least of understanding everything. So, for example, a child could never properly be said to understand anything at all, nor perhaps could the overwhelming majority of the human race. There is clearly something untrue to our experience and common-sense there.

I am not treating this position with the seriousness it deserves in the space available here. In a weaker form it might draw universal agreement. If, for example, it were put in the weaker form that it was not really worth starting machine translation in the way they did in the 1950's, because they knew they had no semantic mechanisms, and so without some ability to go further, it was not even worth starting there. In that weaker form the argument looks far more plausible.

What I am questioning here is its stronger form: and again the reply is the same, namely that the position is another version of the 100%-rule fallacy: that in science you have to have a complete theory to have any worthwhile theory at all. This is untrue to language and diverts our attention from application and from an extensible system that could misunderstand and recover.

Let me summarise the position paper: it is an attack on what I have called the 100%-rule fallacy, alias the use of scientific methodology and assessment in work on AI and natural language. In my view this position has four unfortunate aspects:

1. It requires holding, usually implicitly, the false metaphysical position that there is some boundary to natural language over which one cannot step.

2. It has a false view of the role of counter-examples as *rejectors*.

3. It encourages talk of theoretical advance in a non-theoretical area, and downgrades the engineering aspects of AI, and thus the notions of tests and application, which are the only criteria of assessment we have or could have.

4. It distracts attention from the heuristics of misunderstanding which should be the key to further advance.

REFERENCES

Minsky, M., "A framework for representing knowledge", *MIT AI Memo No. 306*, 1974.

Riesbeck, C., "Computational understanding", *Memo from ISSCO No. 4*, 1974.

Wilks, Y., "Computable Semantic Derivations", Systems Development Corp., Memo, 1968.

Wilks, Y., "Decidability and Natural Language", *Mind*, 1971.

Wilks, Y., "One Small Head", *Foundations of Language*, 1974.

Wilks, Y., "Philosophy of Language" in *Notes for the Tutorial on Computational Semantics*, ISSCO, Castagnola, 1975a.

Wilks, Y., "A preferential Pattern-matching Semantics for Natural Language Inference", *Artificial Intelligence*, 1975b.

Wilks, Y., "An intelligent analyzer and understander of English", *Comm. A.C.M.*, 1975c.

Winograd, T., *Understanding Natural Language*, Edinburgh, 1972.