

SEMANTIC PRIMITIVES IN LANGUAGE AND VISION

Yorick Wilks
Department of Language and Linguistics
University of Essex
Colchester, England.

The purpose of this brief note is to argue that, whatever the justification of semantic primitives for language understanding may be [see Wilks 1977] there is no reason to believe that it relates to vision in any strong sense.

By "semantic primitives" I mean the general sort of item proposed within Artificial Intelligence (AI) by Wilks (1972, 1977), Schank (1973) and within linguistics by Fodor and Katz (1963), Jackendoff (1975) among many others, in both cases. The generality of these items is essential to my argument, and I shall not count as semantic primitives items used for special tasks, whether or not those tasks are related to vision, as are the visual description primitives of Johnson-Laird (1977).

Spatial versus visual

What follows is highly naive and speculative: it will rest largely upon the opposition of linguistic knowledge to spatial and visual knowledge respectively. I take it for granted that the latter are not necessarily connected, and so to establish that we need spatial knowledge to understand language (to name a task at random) does not establish that we need visual knowledge. The lack of necessary connexion is shown by such hackneyed examples as the person blind from birth, who has no visual, but a great deal of spatial, knowledge.

One initial reason for distinguishing the two is the great deal of argumentation in linguistics in recent years that falls under the general heading localism. This thrust of argumentation has sought to establish the central role of spatial concepts in linguistics, and among its best known proponents are Anderson (1971), Fillmore (1977)

and Jackendoff (1975). One stand in this view is to argue that temporal expressions are in general reducible, in some sense, to spatial ones: that in ten minutes (a time expression) is dependent on the spatial sense of such forms as in five miles. This is a very difficult and general debate: there is contrary evidence from cultures where space is indicated by time (The airport is about ten minutes away), and there is a strong philosophical tradition, centred round Kant, that our sense of time is logically prior to our sense of space. That is to say, we could conceive of structuring our experience without the concept of space, but not without that of time because, if we could not know that one event preceded another, then we could probably not know anything at all; not even mathematics if that consists at bottom in sequences of operations. Michotte's famous experiments on the willingness of subjects to attach the word cause to moving pictures of pairs of "striking billiard balls" is sometimes cited as providing a visual basis for causality (Clarke & Clark 1977), although the notion of causality may well in fact make no sense without the concept of time. We could assert (wrongly, as it happens) that lightning causes thunder without the aid of a spatial concept, but not without a temporal one.

The logical or linguistic priority of space to time is by no means a settled matter, and neither therefore is the thesis of localism. I have argued that the role of the visual in language is not necessarily supported by the need for spatial knowledge, and so the status of the latter need not be discussed. Nonetheless, I have questioned the self-evidential truth of

localism, just in case anyone should think that, if it were true, it would support the centrality of visual knowledge in language understanding.

Let us now, as the brief substance of this paper, look at three arguments that might be put forward to support the dependence, or interdependence, of linguistic and visual knowledge.

Evolutionary arguments

This comes in phylogenetic and ontogenetic forms. The former is the ingenious argument (Gregory 1970) that, since the human race has been able to see for many times more millenia than it has been able to speak or write, then it might seem reasonable to believe, on evolutionary grounds that the brain "took over" the existing visual structures for language understanding and production. This argument may well be true, but at present there is no independent evidence that would count for or against it.

The "ontogenetic form" of the argument - in the individual, that is - is that one first learns words essentially through the visual channel, and so again our linguistic knowledge is essentially dependent upon visual criteria and experience. The best quick answer is to turn to the sort of word often used as a semantic primitive in AI language understanding systems: STUFF (=substance), ATRANS (=changing the ownership of an entity), CAUSE (=preceding and necessitating an event). It is highly dubious that such very general concepts are, or can be, taught by visual/ostensive methods. Can one point at substance as such? One may want, or mean, to, but can one in fact reliably do so?

One structure for many purposes

This is a widespread view in AI that has been argued for explicitly by Minsky (1975) and Rieger (1976), among others. Roughly speaking, it is that implemented systems should use a single knowledge structure for a range of purposes: language understanding, problem solving, etc. It is an additional assumption that human beings do function in this way.

The thesis can be expressed at many levels, and at a sufficiently general level it is almost certainly true. But it might then mean no more than that a single programming language could

express general sub-routines for parsing, noise reduction etc. for a number of input channels. At a more specific level was the thesis, not now widely supported, that language and vision in some sense shared the same "grammar", in the sense of Chomsky's transformational grammar (Clowes 1972). Striking evidence from the parallelism between visual and linguistic ambiguity was found, and the fact that Chomsky's grammars no longer seem such plausible candidates for such a role does not mean that the thesis itself is false at that level.

Let us concentrate for a moment on two more specific levels. First, consider the well-known contrast between such sentences as:

The paper moved

The dog moved

Linguists who differ about much else would want to ascribe a notion of agency to the subject of the second sentence but not the first. Many in AI working on natural language would agree, and add that the notion of agency is essential if other important inferences are to be made. But, surely no one would argue that agency is, in any useful sense ascribed a visual criteria, that could be reduced to the visual differences of paper and dogs. It is in fact a complex theoretical notion dependent upon our beliefs and theories about the world: we do not now attribute agency to trees, though some fellow humans do. But this difference is a theoretical (including linguistic) one, not one of difference of visual perception.

Secondly, we may return to general semantic primitives of the sort already mentioned (and similar inventories may be found in (Bierwisch 1970) and (Leech 1974)).

There are many possible ways in which one might seek to justify such primitives (see Wilks 1977), and Bierwisch (1970) has gone on record as saying that they do denote, and are to that extent dependent upon visually observable entities. I suggested above that that may not be so: one may point at treacle, water or elephant meat, but it is not so clear one can point at SUBSTANCE, yet this notion has a role to play in language understanding for how, without it, can one economically express such axioms as "A quantity of a substance plus a quantity₂ of it

yield a quantity₃ of it". This axiom is not true of physical objects, as distinct from substances.

A well-known confusion must be avoided here: it may well be true, as the model theoretic semanticists like Montague claim, that any contentful notion, primitive or not, refers to a function of sets. In that sense move might be said to refer to a set of entities that move.

However, this point about logical reference has no consequences for the point about whether or not such primitives denote entities in the real world.

Visual and spatial imagery

Finally, it is sometimes argued that the structures underlying language must depend upon those underlying vision if only because natural language is so full of visual imagery. In whatever sense "visual imagery" is taken, this fact is, I believe, irrelevant to any precise assertion under discussion, by which I mean any of

- I) Language understanding processes in humans depend, either as to primitive elements or structure, on visual experience and the mechanisms that interpret it.
- II) The specification of language in humans has no significant overlap, in terms of primitive elements or structure, with that of other faculties, like vision.
- III) Visual processes in humans depend, either as to elements or structure, on linguistic experience and the mechanisms that interpret and produce (sic) it.

For all three theses only anecdotal evidence is available, though I would be strengthened by empirical evidence that the blind from birth were less able to understand the use of visual imagery in language. Those with a predilection for motor theories should be tempted to consider the Whorfian thesis III (Whorf, remember, believed we might perceive, say lightning, as an entity, rather than an activity or process because we denoted it by a member of the theoretical category NOUN, rather than VERB) since, as the structural difference of I and III makes clear, language is an activity in a way vision is not.

Thesis II will be agreeable to those who are impressed by the way in which confusion can arise

when one tries to bring together information on the same topic, but obtained via different channels. As when one refers to two cities whose mutual relation of position one knows from a map; between which one can drive "without thinking"; and also about both of which one has a great deal of textual/factual information. Readers of (Fillmore 1977) will recall his attempt to describe the relation of a text-based frame and an experientially-based scene to the same situation. I think AI workers at this particular interface could profit from considering the extent to which such possible inconsistencies can be matters of theory rather than superficial fact: an observer who is asked whether two sides of a long railway line meet at the furthest point he can see will give an answer not independent of his abstract (possibly linguistically based) theory of parallel lines.

In conclusion, this note has tried to do no more than ward off certain confusions, and to stress how many points of view are still open, since the evidence for and against them is no more than anecdotal, even when the anecdotes come from Psychology labs. The choice between theses I/II/III is a metaphysical one, in the more red-blooded sense of that overtired word: it cannot be made on empirical grounds now, but it can have important practical consequences about where one chooses to look for answers.

References

- Anderson, J. (1971) The Grammar of Case
(London: Cambridge U.P.)
- Bierwisch, M. (1970) "Semantics" in Lyons (ed.)
New Horizons in Linguistics
(London: Penguin)
- Clark, E. & Clark, H. (1977) Psychology and Language
(New York: Harcourt
Brace)
- Clowes, M.B. (1972) "Scene Analysis and Picture
Grammars" in Nake & Rosenfeld
(eds.) Graphic Languages,
(Amsterdam: N. Holland).
- Fillmore, C.J. (1977) "Scenes and Frames
Semantics" in Zampolli (ed)
Linguistic Structures
Processing
(Amsterdam: N. Holland)
- Gregory, R. (1970) "The Grammar of Vision".
The Listener. (London: BBC)
- Jackendoff, R. (1975) "A system of semantic
primitives" in Schank &
Nash-Webber (eds.)
Theoretical Issues in
Natural Language Processing
(Cambridge, Mass.: BBN)
- Johnson-Laird, P. (1977) "Psycholinguistics without
linguistics" in
Sutherland (ed.) Tutorial
Essays in Psychology
(Hillsdale N.J.: Erlbaum)
- Katz, J. & Fodor, J. (1963) "The structure of a
semantic theory".
Language.
- Leech, G. (1974) Semantics (London: Penguin)
- Minsky, M. (1975) "Frame Systems" in Schank &
Nash-Webber (eds.) Theoretical
Issues in Natural Language
Processing. (Cambridge, Mass.: BBN)
- Rieger, C. (1976) Computers and Thought Lecture at
IJCAI4, and published in Artificial
Intelligence.
- Schank, R. (1973) "Identification of Conceptual-
izations underlying Natural
Language". in Schank & Colby (eds)
Computer Models of Thought and
Language. (San Francisco: Freeman)
- Wilks, Y. (1972) Grammar, Meaning and the Machine
Analysis of Language. (London &
Boston: Routledge)
- Wilks, Y. (1977) "Good and bad arguments for
semantic primitives", Memo No.42,
(Edinburgh: Dept. of Artificial
Intelligence).
- Wilks, Y. (1975) "Primitives and words"
in Schank & Nash-Webber (eds)
Theoretical Issues in Natural
Language Processing.
(Cambridge, Mass.: BBN)