WHAT DAVIDSON SHOULD HAVE SAID

Ernest LePORE
Barry LOEWER
Rutgers University

It is natural to think, as many have, that the expressions, both simple and complex, of a language L possess meanings and that a theory of meaning for L should specify these meanings. Such a theory would, it is said, illuminate the semantics of L by showing how the meanings of its expressions are interrelated (for example, how the meanings of complex expressions depend on the meanings of their parts) and also illuminate meaning, at least for L, by specifying the extension of “means” in L. Further, if, as many have claimed, understanding L consists at least in part in knowing, if only implicitly, the extension of the meaning relation for L, such a theory of meaning would be the core of a theory of understanding L. A first proposal concerning a meaning theory for L is that it issue in true theorems of the form:

(M) S means in L that p

for all sentences S of L. If we suppose that meanings are the referents of expressions of the form “that p”, then a theory (TM) which entailed all and only true instances of (M) for sentences of L would fix the extension of “means-in-L” (and presumably, given sufficient metalanguage apparatus, the theory would entail an explicit definition of “means-in-L”). There are, however, a number of difficulties with this way of formulating a theory of meaning.

1. An earlier draft of this paper was presented at a Symposium on Donald Davidson’s philosophy in Gornja Radgona, Yugoslavia, and Bad Radkersburg, Austria, November 12, 1988, and at the University of Minnesota. We would like to thank both groups for the opportunity to present our work. We would also like to thank Michael Root, John Wallace and especially Donald Davidson for their comments.
While instances of (M) specify the meanings of sentences of L, they obviously are inappropriate for specifying the meanings of sub-sentential expressions: “‘and’ means and” is ill-formed. Further, without having some way of specifying the meanings of sub-sentential expressions, it is difficult to see how we would be able to derive the true instances of (M), since the meanings of sentences depend on the meanings of their parts. Another problem is created by the apparent intensionality of “means-in-L that”. A theory which entails instances of (M) presumably will be formulated in an intensional language. But the logics of intensional (as well as extensional) languages for which semantics have been developed contain a principle which allows for the substitution of logically equivalent sentences. So, if (TM) entails “S means that p” it will also entail “S means that p”, where “p” and “p” are logically equivalent. But logically equivalent sentences need not possess the same meaning. Finally, perhaps for the reasons just mentioned, no one has produced a theory which entails all the instances of (M) for even a simple first order language.

Another approach to specifying meanings is “translational semantics”. Instead of instances of (M), a translation theory produces instances of:

**TR)** e in L means the same as e* in L*

A theory entailing instances of (TR) does not face the same problems as (TM) since “means the same as” is purely extensional and its extension is not restricted to sentences. But such a theory may not tell us much about the meanings of the expressions of L. In particular, someone might know all the instances of (TR) for L and yet not understand L. Thus, someone might be able to match Italian expressions with their Russian translations, say, on the basis of lexicographic and syntactic rules, and yet not understand the expressions of either language.²

Davidson has suggested another way of formulating a theory of meaning for L which seems to avoid the pitfalls of both the translational and meaning approaches. Instead of trying to produce a theory which specifies the extension of “means that” or “means the same as” Davidson asks what we want a theory of meaning to do? His answer is:

The theory should specify information such that if someone had this information he would be in a position to understand L, or at least well on his way to understanding L.³

He then claims that a theory of truth for L fits this description. A theory of truth for L issues in theorems of the form:

**T)** S is true-in-L iff p

for each sentence of L. Truth theories are purely extensional, apply via the satisfaction clauses to sub-sentential expressions, and we know, at least for formal languages, how to construct them. Unlike translation theories, the information they specify does seem relevant to understanding L. An instance of (T) says under what conditions S is true and not merely that S has the same truth value as a sentence in some (other) language. As Tarski observed, a true theory for L fixes the extension of “is true-in-L”. But how does it illuminate meaning in L? There are two claims, each related to Davidson’s characterization of the goal of an adequate theory of meaning:

I. The truth theory shows how the truth conditions of complex sentences depend on the truth conditions and other semantic features of constituent expressions. Thus, a truth theory (TT) entails, for example, “S & S’ is true iff p and q” by deriving it from “S is true iff p” and “S’ is true iff q” and an axiom “(∀x)(∀y)(x “&” y is true and y is true)”.

II. A truth theory (TT) issues in true theories of the form (T). Although such “T-sentences” do not say that S means that p, Davidson suggests that in fact an adequate truth theory will entail that “S is true iff p” iff S means that p.

If condition (II) is satisfied by (TT), then we will have succeeded in pairing each sentence of L not only with its truth condition, but with what we might call its meaning condition. We need to

². See LePore and Loewer, 1981, for further discussion and criticism of this approach.

³. See, for example, Davidson 1967, 1973 and 1974.
address the questions whether there are such truth theories for L, and if so, which? This divides into: If (TT) is a true theory for L, will (TT) satisfy:

1. If S means that p, then (TT) will entail “S is true iff p”.

2. If (TT) entails “S is true iff p”, then S means that p.

Tarski stipulated that a theory counts as a truth theory for L only if it entails an instance of (T) for each S of L in which “p” is a translation of S. But, as Davidson observes, we do not want to make this stipulation in formulating a theory of meaning since we would like the fact that “p” translates S to emerge from the conditions placed on the truth theory. Minus the translation condition, the constraints on an adequate truth theory for L are (i) that it entail a T-sentence for each S in L, (ii) that it is true, and (iii) it is finitely axiomatizable. 4

Clearly, not every truth theory for L satisfying (i)-(iii) will satisfy (1). For example, it is plausible that if there is any truth theory for a fragment of English containing “Snow is white” and “Grass is green” and kindred sentences, then there will be a truth theory which entails “Snow is white’ is true iff grass is green” (but not “Snow is white’ is true iff snow is white”). In “Truth and Meaning”, Davidson suggested that the demands on a truth theory — that it be complete (entail a T-sentence for each sentence of L) and that it be true (for example, does not contain “snow refers to grass”) — make it impossible for there to be such a theory. He was wrong.

Suppose that (TT) is a truth theory which satisfies (1). Presumably, there are such theories. From it, though, we can obtain another truth theory (TT’) which satisfies (i)-(iii), but fails to satisfy (1). Just take some predicate clause, for example, “(∀x)(x satisfies ‘is wet’ iff x is wet)” and replace it with “(∀x)(x satisfies ‘is wet’ iff x is wet’),” where “is wet” has as its extension wet things in the actual world and dry things in other worlds. 5

4. We do not want to discuss or motivate condition (iii) in this paper. See LePore and Loewer, 1983, and LePore, 1982.

5. It is not clear that this will work. Suppose the language has counterfactuals? How will an absolute truth theory deal with these? It might be that if we have counterfactuals then our cooked-up clause will yield false T-sentences, violating condition (ii). In any case, there obviously are theories which satisfy (i)-(iii), but fail to satisfy (1). For example, a theory which entails “S is true iff 2+2=4” or “S is true iff 2+2=5” (contingent upon whether “S” is true or false). Among such theories is one which has just true consequences. Of course, one would not know that one had such a theory unless one knew the truth values for all sentences of L.

6. Someone might argue that though both statements are laws only the first is a semantic law. This suggestion is vacuous unless a non-question begging characterization of semantic law is provided.
These predictions concern what sentences the speakers of L hold true under various circumstances. In the way, the theory can be used to predict correct beliefs on the basis of sentences held true. We will imagine that we have identified the sentences of another’s language and those he holds true. We test a theory of truth for his language in the following way: if he holds S true and the truth theory that we are testing entails that S is true iff p, then we will assume he believes that p. Different theories will predict he has different beliefs. Thus, a (TT) which entailed the wrong sentences would predict (together with reasonable accounts of belief formation) that after spending a winter in New York City an English speaker would no longer hold true “Snow is white” but would still hold true “Grass is green”. [It is reversed after spending a summer in Los Angeles during a drought.]

It is not clear to us whether either condition or the two together (they are obviously closely related) will filter out all truth theories except those that satisfy (1). One troubling thought for this second line is that if we do not have access to a person’s beliefs independently of access to his language, as is suggested by some of Davidson’s writings, then there may be alternative truth theories and belief assignments (or accounts of belief formation) which together make exactly the same predictions concerning which sentences the person holds true or would hold true. If so, then this condition will not do as much winnowing as anticipated. However, a defender might respond that it was wrong to expect such.

Although we are not sure whether any non-question begging constraint will yield truth theories which satisfy (1), we are certain that no constraint of any sort will lead to satisfaction of (2). The difficulty is obvious and one we already have mentioned in our discussion of the troubles with any effort to construct a meaning theory. It is the problem of substitutivity of logical equivalences. If (TT) entails that S is true iff p, and if p and p* are logically equivalent, then (TT) will entail S is true iff p*. Indeed, the problem is even worse; if (TT) entails anything R, and it entails that S is true iff p, then it will entail that S is true iff p & R. So, we will derive theorems like “There is a fire in the building” is true iff there is a fire in the building and (q or not q)”. Put simply, as long as we are using normal logical apparatus (the theory is formulated in a first order language or some extension thereof), then satisfying (2) is out of the question. And any tinkering with the logic of the truth theory to prevent it from implying “bad” T-sentences would prevent it from implying “good” T-sentences as well.

How bad is this? Should we then abandon the idea that a theory of truth is the core of a theory of meaning? No! But we should abandon the idea that a theory of truth serves as a theory of meaning by producing T-sentences which simulate M-sentences. Once we return to Davidson’s suggestion that a theory of meaning for L should specify information sufficient for understanding L we should see why this is so. Curiously, neither Davidson, nor others who have advanced or seconded his suggestion (including Davidson’s Oxford followers), have ever quite articulated exactly how knowledge of a truth theory (even one satisfying (1) and (2), if there were one) would yield a theory of understanding of L. Perhaps, it was thought that if a (TT) satisfied (1) and (2), then it would be obvious how since knowing (TT) would amount to knowing the meaning conditions of sentences of L. But what has this got to do with understanding language? To understand a language involves at least this much: to be warranted in drawing (or acting as if one were drawing) inferences characteristic of understanding.

Suppose Donald and Luca are in a hotel in New York waiting for the mail. Mary enters the room and utters “There is a fire in the building”. Donald understands English and Luca does not. On this basis, all other things being equal, we would expect that Donald leaves the building, while Luca just sits there. If in fact that happens, it would be natural to explain Donald’s behavior by noting that upon hearing Mary he acquired the belief that there is a fire in the building. Donald’s belief (and his ensuing behavior) is warranted, in part, because he understands English. The exact psychology of Donald’s belief acquisition is a matter for psychology and we do not intend to address that matter here. Our conception of truth conditional semantics emphasizes its epistemological, rather than its psychological, role. If we want to specify knowledge which if Donald had it,
it, together with other things Donald (and for that matter Luca) knows, would justify his belief, the knowledge that “There is a fire in the building” is true iff there is a fire in the building is the natural candidate. The following reasoning makes this explicit:

Mary’s utterance “There is a fire in the building” is true.

Mary’s utterance is true only if there is a fire in the building.

So, there is a fire in the building.

The aim of producing a theory which satisfies (1) and (2) is not an appropriate aim for a theory of meaning. This aim should not be to define or analyze meaning. Rather, it should be to produce a theory which explains the phenomena for which we invoke the concept of meaning. Here the phenomena is language understanding. Davidson never claimed that the biconditionals issued in meanings. He claimed, rather, that knowing certain biconditionals, those implied by an adequate truth theory, suffices for understanding. Nothing is upset by knowing additional true biconditionals. Someone who knows the above would be in a position to conclude both that there is a fire in the building and also that there is a fire in the building and (q or not q), for any q, from “There is a fire in the building” is true. But there is still a problem.

Even if one knew a truth theory for Mary’s language which satisfies (1) and knew that Mary believes all the T-consequences of this theory (or believes the theory) this would be insufficient for understanding it. When we know that Mary holds true “There is a fire in the building”, say, because she uttered it, we are then in a position to justify our beliefs that there is a fire in the building, assuming she is reliable and sincere. But we cannot conclude that Mary said that there is a fire in the building. The real trouble with a truth theory as a theory of understanding is that its T-sentences are too weak to warrant inferences to what someone said (asked, commanded, etc.) on the basis of what he uttered. Thus, truth theories qua

8. Both Donald and Luca could know that Mary uttered “There is a fire in the building” and even that her utterance is true. Evidence for an utterance being an assertion and for it being true need not be semantic.

theories of understanding become unhinged just where we demand them to be most secure, namely, in accounting for inferences to what others say on the basis of what they utter.

We return to New York. Suppose that when Mary uttered “There is a fire in the building” only Donald was within earshot. Provided Donald understands Mary’s language, then Donald would, in “normal” circumstances, be warranted in believing that Mary said that there is a fire in the building. Indeed, aren’t these sorts of belief constitutive of understanding? Should Donald want to inform Harry of Mary’s utterance, of Mary’s opinion, he can take recourse either to direct or indirect quotation. Either way Donald treats himself as a conduit for information from Mary to Harry. Should Donald utter “Mary said ‘There is a fire in the building’”, then if Harry understands English, minimally he would be warranted in believing that Mary uttered a sentence of her language of the same phonological type as the sentence uttered by Donald. Harry might further conclude that Mary said that there is a fire in the building. How is this possible?

Many writers have argued there can be no answer within the Davidsonian “extensional” framework to this question. Indirect quotation is an “intensional” notion, so the folklore goes, and truth theories are extensional animals. How could knowledge of just the conditions under which sentences are, as a matter of fact, true ever support inferences inside apparently intensional contexts like “Mary said that…?” We quickly restate the objection and answer it. Our answer has important consequences for theories of interpretation.

Foster [1976], Loar [1976], Wallace [1979], Evans and McDowell [1976], among others, each objected to Davidson that the trouble with a truth theory as a theory of meaning is that its T-sentences are too weak. It is this, they argue, that allows theories (or theorems) in which S does not mean that p. Furthermore, it might be thought, it is just this weakness that prevents one from concluding what a speaker says when he produces an utterance. It looks like “A uttered S and S means that p” does entail that “S said that p”. So, there seems to be two distinct, but easy to conflate, issues:

i. Adding further conditions on a truth theory so as to produce one that satisfies the MC, toss out the “bad” ones, and
ii. Determining how to use the truth theory to obtain conclusions like “A said that $p$”.

(i) cannot, as a matter of logic, be solved for the reason we advanced earlier. However, as we also argued, the objection that a biconditional like “‘There is a fire in the building’ is true if and only if there is a fire in the building and (q or not q)” does not provide the meaning of “There is a fire in the building” is misdirected. (ii) is a different objection and it is not misdirected: utterances of “There is a fire in the building” and “There is a fire in the building and (q or not q)” do not say the same thing. Any competent speaker of English knows this. So, defenders of truth theoretic approaches to meaning are left with the problem of discovering a way of representing information which can justify conclusions concerning what is said compatible with the truth theoretic approach. As we will now show, Davidson’s paratactic account, somewhat unexpectedly, provides a solution to this problem.

Whenever Mary utters “There is a fire in the building”, she produces an utterance the same in content (purport, import) as many other utterances of the English sentence “There is a fire in the building”. But then it would follow that there is an utterance $u$ such that Mary produced $u$ and $u$ is the same in content as “There is a fire in the building”. This is, according to paratactic account [Davidson, 1968, 1969], the correct semantic account for an utterance “Mary said there is a fire in the building”, which happens to be our target sentence. On the paratactic account, the fact that Mary uttered “There is a fire in the building” and the fact that her utterance bears the $\text{same as}$ relation to (another) English speaker’s utterance of “There is a fire in the building” implies that Mary said that there is a fire in the building. If we assume, then, “that” refers to an utterance of “There is a fire in the building”, then if the premises are true, the conclusion is true. The conclusion just is “Mary said that”. The exact nature of this inference requires us to discuss the logic of demonstratives. Something that space limitations prevent us from doing here. However, even without this discussion, we have said enough to see how we can overcome perhaps the major objection to Davidson’s theory, namely, that the theory is too weak to support inferences from “Mary uttered ‘There is a fire in the building’” to “Mary said there is a fire in the building”. Sometimes Davidson says that knowledge of a truth theory for a language $L$ is sufficient for understanding $L$ [1967, 1973, 1974]. Later (in his reply to Foster [1976]) he modified this adding that one also must know that the truth theory meets certain empirical and formal constraints. The critics were right inasmuch as neither knowledge of a truth theory nor knowledge that the truth theory is an empirically adequate truth theory can bridge the gap to what is said (even though it effectively bridges the gap to knowledge that there is a fire in the building). What we have just shown is that everyone was looking in the wrong place. It is not (and cannot be) the truth theory, no matter how it is constrained or what is known about it, which bridges the gap between heard utterance and ascribed assertion. What Davidson should have said is that one needs also to know the sameness relation, or some proper subset of this relation, for the language. How much does this add to Davidson’s over all project of stating in extensional terms what it is one knows that enables one to understand a language? Not much. One needs to know the sameness relationship anyway for interpreting indirect discourse, if Davidson’s own extensional account of this sort of discourse is correct.

9. This has little to do with truth theories, per se; rather, it has to do with the truth theory being a theory. Suppose we add a modal operator, both as Foster [1976] and Wallace [1979] suggest, to strengthen the biconditionals which issue from an adequate truth theory. For example, Wallace suggests that we add as a primitive the operator “It a matter of meaning alone”. Therefore, according to him, our target sentences are of the form: It is a matter of meaning alone that $S$ is true if and only if $p$. But how, if his theory has standard logical principles (how can it not?), is he going to prevent the derivation of false theorems of the form: It is a matter of meaning alone that $S$ is true if and only if $(q$ or not $q)$?

10. We need to be careful here. Strictly speaking, the two premises do not logically imply the conclusion, since there are contexts in which the premises are true but the conclusion is false, for example, in which “that” refers to some other utterance.


12. The question naturally arises how one can know when two utterances samessay each other? That is a hard question to answer. We can begin to answer it by considering the point of reports like “Mary said that there is a fire in the building conveying to his listener (a portion of) the information which would have been
It is important to realize that we are not saying that all one needs to know in order to understand a language is the samesay relation for this language. After all, one can know that one sentence S same-says another without understanding S (indeed, without understanding either). Thus, on the paratactic account, one could be in position to conclude "Mary said that p" without understanding what Mary said. But here is exactly where knowledge of the truth theory enters. Our claim is not that speakers need only know which utterances samesay each other; they must also know the truth conditions for these utterances. They can then employ this knowledge together with knowledge of the extension of the samesay relation to conclude, in accordance with patterns of inference now familiar, that Mary said something which is true iff there is a fire in the building.

In conclusion we say that it suffices to understand a language to know:

A) All the biconditionals entailed by an adequate theory, and

B) For any two utterances of sentences of the language, whether or not they samesay each other.

Notice this does not require knowing theories which entail the specific knowledge in (A) and (B). Both kinds of knowledge seem to be infinite, so it may be impossible to specify either without devising a theory. That is not our concern here. Our concern was to show in what sense a truth theory is not too weak to serve as a theory of meaning and in what sense it is.

We argued that both the critical literature and Davidson’s response to this literature have been misguided. Critics have been misguided because they have not been clear about what it is a theory of meaning is supposed to do. Some critics have been misguided because they have thought that a theory of meaning for a language L is supposed to pair each sentence of L with its meaning. No truth theory can do this. Both Davidson and his critics seem to have been misguided in thinking that by adding further conditions on an empirically adequate truth theory we can derive what was asserted. We showed that Davidson had available, though he apparently failed to see so, a reply to his critics in his account of the semantics for indirect discourse reports. On this account "Mary said that there is a fire in the building" and "Mary made an utterance which samesays my utterance of ‘There is a fire in the building’" are equivalent. Everything stands on this and obviously someone might and indeed some have objected here. Whether this account is itself acceptable is the topic of another paper.13

REFERENCES


