Truth theories, translation manuals, and theories of meaning

Jeff Speaks

January 20, 2006

Versions of the Davidsonian program in semantics aim to use a Tarskian truth theory to accomplish the tasks of a theory of meaning. In "Truth and Meaning," Davidson identified two related aims for the theory of meaning: (i) to give the meanings of expressions of the language, and (ii) to explain the semantic competence of speakers of the language, by stating information knowledge of which would be sufficient to understand the language.

From the start, the attempt to use a truth theory to accomplish these tasks faced a number of fundamental objections. In the last thirty years, though, a number of variants of Davidson's original framework have been proposed to avoid these objections. The question I wish to raise in this paper is: do any of the proposed modifications of Davidson's original theory validate the idea that, as Davidson put it, we find in Tarskian truth theories "the sophisticated and powerful foundation of a competent theory of meaning"?[1]

I will argue that they do not. To show this, it will be useful to begin with a discussion of two traditional problems for the Davidsonian program, and the solution to these problems offered by Max Kölbel in his "Two Dogmas of Davidsonian Semantics."[2]

I. Davidsonian semantics without the two dogmas

It is uncontroversial that one of the aims of a semantic theory for a language is to, in some sense, "give the meaning" of expressions of that language. What is not so uncontroversial is that this should be accomplished by designing a theory to entail, for each expression in the language, a *meaning theorem* of the form

S means in L that p,

instances of which are obtained by replacing S with the name of a sentence, p with a sentence, and L with the name of a language. Davidson, for example, designed his

---

[1] "Truth and Meaning," reprinted in his *Inquiries into Truth & Interpretation* (Oxford: Clarendon Press, 1999), p. 24.
[2] *Journal of Philosophy* XCVIII, 12 (December 2001): pp. 613-635.

theory to yield, not meaning theorems, but *T-sentences* of the form

S is T in L iff p,

instances of which are obtained in the same way, with a truth predicate for the language named by the value of L replacing T. Kölbel calls the view that the target theorems of a theory of meaning should be biconditional T-sentences the 'biconditional doctrine.'

This biconditional doctrine gave rise to two fundamental early criticisms.[3] The first problem stems from the fact that it is not enough for a semantic theory whose theorems are T-sentences to yield true theorems; the T-sentence

"Snow is white" is T in English iff grass is green

is true, but tells us hardly anything about the meaning of "Snow is white." Rather, we want a semantic theory to entail, for each sentence of the object language, exactly one *interpretive* T-sentence: a T-sentence such that the sentence used on its right-hand side gives the meaning of the sentence mentioned on its left-hand side. Our theory must entail *at least one* such T-sentence for each sentence in the object language because the aim is to give the meaning of each sentence in the language; and it must entail *no more than one* because, if the theory had as theorems more than one T-sentence for a single sentence S of the object language, an agent who knew all the theorems of the theory would not yet understand S, since such an agent would not know which of the T-sentences which mention S was interpretive. Following Larson and Segal, I shall call this the *extension problem*.[4]

The first half of the extension problem — designing a theory to entail at least one interpretive T-sentence for each sentence of the object language — is a problem which can only be solved by detailed examination of natural language constructions. But the second part of the extension problem is different in character. Given that Davidsonian theories yield T-sentences as theorems, the problem arises that, for any sentences p, q, if the theory entails a T-sentence

S is T in L iff p,

then, since p is logically equivalent to p & $\neg(q$ & $\neg q)$, the theory will also entail the T-sentence

S is T in L iff p & $\neg(q$ & $\neg q)$,

which, if the first is interpretive, won't be.

As Kölbel notes, the standard solution to this second half of the extension problem is the specification of a set of *canonical inference rules*, which have the following characteristic: any T-sentence derivable via these rules from the axioms of the theory will be interpretive. The idea is that the above example shows that the rules of

---

[3]Early forms of both criticisms may be found in J. A. Foster, "Meaning and Truth Theory," in *Truth and Meaning: Essays in Semantics*, Gareth Evans & John McDowell, eds. (Oxford: Oxford University Press, 1976), pp. 1-32.

[4]Richard Larson & Gabriel Segal, *Knowledge of Meaning* (Cambridge, MA: MIT Press, 1995), ch. 2.

inference for our semantic theory cannot be as permissive as the rules of inference of classical logic; we need to restrict the sorts of inferences which can be employed to arrive at theorems of the theory in order to block the derivation of uninterpretive theorems like the above. Several different ways of doing this have been suggested.[5] Suppose that one of these is successful. We've then solved the extension problem: the theorems of our semantic theory will be just those T-sentences derivable via the canonical rules of inference from the axioms, and so each of these theorems will be an interpretive T-sentence.

The second, and more fundamental, traditional problem facing Davidsonian theories is the *information problem*. Even if our semantic theory entails all and only interpretive T-sentences, it is not the case that knowledge of what is said by these theorems would suffice for understanding the object language. For, it seems, I can know what is said by a series of interpretive T-sentences without knowing that they are interpretive. I may, for example, know what is said by the interpretive T-sentence

"Londres est jolie" is T in French iff London is pretty

but still not know the meaning of the sentence mentioned on the left-hand side of the T-sentence. The truth of what is said by this sentence, after all, is compatible with the sentence used on the right-hand side being materially equivalent to, but different in meaning from, the sentence mentioned on the left. This seems to indicate that knowing what is said by a truth theory of the relevant kind is not, after all, sufficient for understanding a language.

Kölbel's central idea is that, once we have defined a set of canonical inference rules, we can specify one further inference rule which takes us from theorems of the form

S is T in L iff p

to meaning theorems of the form

S means in L that p.[6]

Because every theorem of the theory is the result of a canonical derivation from the axioms of the theory, and so is guaranteed to be interpretive, each of the meaning theorems derived by using this inference rule will be true.

This additional rule of inference is attractive for two reasons. First, as Kölbel says, it avoids the "suspect" result that "a semantic theory . . . does not, on its own, provide information on what sentences mean."[7] Second, it promises to solve the information problem; someone who knows what is said by the theorems of a Kölbel-style semantic theory will not be in doubt about whether these theorems are interpretive, since they directly state information about the meaning of the sentence mentioned on the left-hand side.

The inclusion of such a rule of inference, of course, amounts to the abandonment of the biconditional doctrine – the first of Kölbel's two dogmas of Davidsonian semantics.

---

[5]See, for example, Larson & Segal, *op. cit.*

[6]Kölbel gives two different ways of specifying this rule, depending upon how the canonical inference rules are specified, in §III.

[7]Kölbel, *op. cit.*, 618.

The second dogma is what Kölbel calls the 'truth doctrine': the view that truth is, in some sense, explanatorily prior to meaning. Once we reject the biconditional doctrine, Kölbel says,

> "it becomes easier to take an instrumentalist view of the predicate 'is T' in some theorems of a theory of meaning: it enables the recursive machinery to generate interpretive T-theorems — and ultimately theorems of the form 's means that p.' The important function of the predicate is that it allows us to generate theorems that pair object-language sentences with their metalanguage interpretations."[8]

In the sort of theory outlined above, the predicate 'is T' is a placeholder, and nothing more; *any* predicate could be substituted in its place and, so long as the canonical rules of inference of the theory were rephrased so as to apply to this new predicate rather than to 'is T,' there would be no difference at all in the theorems of the theory. In this sense, giving up the biconditional doctrine requires also giving up the truth doctrine. The meaning of the predicate 'is T' plays no explanatory role in the theory.

Now that we've arrived at a Davidsonian semantics free of the two dogmas, we can ask: does this sort of theory accomplish the two aims of theories of meaning we listed at the outset?

II. Semantic theories as giving the meanings of expressions

Whether or not a semantic theory "gives the meaning" of sentences of a language is a matter of degree; a theory may give more or less information about meaning. We could design a theory which told us, for each sentence of a language, a few things that the sentence did not mean; and, while this would surely give us some information about meaning, it would not give us as much as we should expect from a semantic theory. There is no doubt that a semantic theory of the sort envisaged by Kölbel gives us some information about meaning; the question is whether it gives us enough.

One way to decide how much information we ought to expect from a semantic theory is by attention to the actual practice of semantic theorists. If, for example, there are two opposed views about the semantic content of some expression-type, then the existence of the dispute gives us a (admittedly defeasible) criterion for judging whether a semantic theory for that expression-type gives us enough information about meaning as we should want: the theory in question should be inconsistent with at least one of the rival views. For example, classical descriptivists claim, while Millians deny, that the semantic content of an ordinary simple proper name is equivalent to the meaning of some definite description; no semantic theory for proper names worthy of the name could be consistent with both views. In some cases, we may want to say that the dispute is not a substantive disagreement about semantic fact, and so discard the dispute rather than the theory; but if a semantic theory routinely fails to decide disputes within semantics about the proper treatment of some expression-type which it purports to explain, it would be hard to avoid the conclusion that that semantic theory fails to provide sufficient information about meaning.

---

[8] *Ibid.*, 623.

4

I will argue that the sort of neo-Davidsonian theory proposed by Kölbel fails this test. My method will be to describe a semantic theory which (I will argue) clearly fails this test, and to show that it provides as much information about meaning as does the kind of semantic theory which Kölbel describes.

Suppose that we have, instead of a truth theory for a language, a translation manual from one language to another which yields *translation theorems* of the form

S in L means the same as S' in L'

instances of which are obtained by replacing S and S' with the names of sentences, and L and L' with the names of languages. Ordinarily, we would not take such a theory to give us much information about the meanings of sentences; intuitively, it tells us which sentences are equivalent in meaning without telling us what any of them mean. This intuitive judgement is borne out by applying the test described above, for the existence of this sort of translation manual connecting, say, English and French sentences would not decide disputes about the right treatment of expressions of either language. To use the example above, knowledge of how to translate English sentences involving proper names into French sentences involving proper names would clearly not resolve the dispute between competing views of the semantics of names, such as that between Millians and classical descriptivists; and the fact that we can translate English belief sentences into their French counterparts hardly means that we can give an adequate semantic theory for belief sentences.[9]

Given that a simple translation manual of this sort does not give us enough information about meaning, we can, following Gilbert Harman, turn our attention to a kind of expanded translation manual:

> "Suppose . . . that we have a recursive procedure for recognizing the relevant instances of 'S (in L) translates into our language as T.' Then we can easily formulate a recursive procedure for recognizing relevant instances of 'S (in L) means P' or 'S (in L) is true if and only if P' (where what replaced 'S' is the same name of a sentence as that which replaced 'S' in the previous schema and what replaces 'P' is the sentence named by what replaced 'T' in the previous schema). Then we can treat each of the instances of one of the latter schemas as axioms in a formal theory of meaning or a formal theory of truth . . . "[10]

If we add such an inference rule to our translation manual, then, supposing that we have a translation theorem for each sentence of L and that L' is the language of the theory, we can generate a meaning theorem for each sentence of L. Does this mean that such an expanded translation manual gives us as much information about meaning as we should want from a semantic theory?

Again, intuitively, it seems not: given the manifest failure of the un-expanded translation manual to give us enough information about meaning, it would be surprising if we could get all the information about meaning we want from a semantic

---

[9]Thanks to an anonymous reviewer for the latter example. The claim here is not, of course, that cross-linguistic comparisons are never relevant to semantic theory. The point is only that, for some expression e, to produce expressions in other languages which are synonymous with e is not, by itself, to give a semantics for e.

[10]"Meaning and Semantics," reprinted in his *Reasoning, Meaning, and Mind* (Oxford: Clarendon Press, 1999), p. 197.

theory simply by adding to the translation manual a rule of inference which tells us, in effect, to remove the quotes from the sentence on the right-hand side of the translation theorem and to replace the 'means the same as' with 'means that.' And, as above, this intuitive judgement can be supported by reference to the rough criterion of relevance to semantic disputes discussed above. For, despite the fact that the theorems of our expanded translation manual will be meaning theorems rather than translation theorems, our expanded translation manual would not decide many of the disputes in which semanticists engage. Consider again the dispute between descriptivists and Millians over the meanings of proper names. Despite their disagreement over the meaning of the name 'Benjamin Franklin,' *both* descriptivists and Millians will agree that the following meaning theorem is true:

> "Benjamin Franklin was an American" means that Benjamin Franklin was an American.

So this meaning theorem does not give us enough information about the meanings of the expressions mentioned on its left-hand side to decide debates in semantics about the correct treatment of those expressions. Since our expanded translation manual consists only of such meaning theorems along with the original translation manual, we can conclude that it too falls short of giving us enough information about meaning.[11]

Now consider a second expansion of our translation manual. We have been supposing that the original translation manual generates translation theorems for sentences on the basis of axioms which say what proper names of L and L' designate the same thing, what predicates are satisfied by the same things, etc. Let's now suppose that, in addition to the extra rule of inference which permits the derivation of meaning theorems from translation theorems, we add to the theory analogous rules of inference for the translation theorems for subsentential expressions. So, for example, such an extra rule of inference might permit the inference from a translation axiom for proper names like

> "Londres" designates (in French) what "London" designates (in English).

to a statement of what the proper name mentioned on the left-hand side designates, namely

> "Londres" designates (in French) London.

This extra rule of inference could be stated in much the same way as the rules of inference discussed by Kölbel or Harman in the passages mentioned above, and we can imagine adding analogous rules for predicates and other sub-sentential expression-types.[12]

---

[11]And, of course, this example is just that; we might as well have used disagreements over the meanings of belief ascriptions, or sentences containing generics, or any number of other cases.

[12]Again, we rely on the fact that one of the languages connected by the translation manual is the language of the theory. More on this below, in §III. This type of expansion can provide analogous theorems for every type of sub-sentential expression, including complex predicates and singular terms. After all, the translation manual will have to deliver theorems which say how such complex sub-sentential expressions are to be translated; and rules can be added, analogous to the above, which take any such translation theorem to a theorem which says (for example) what a given complex singular term designates.

Does our doubly expanded translation manual give us enough information about meaning? It is hard to see why matters should be any different than with our once-expanded translation manual. After all, just as semantic theorists who differ on the semantics of proper names agree on the truth of meaning theorems like

> "Benjamin Franklin was an American" means that Benjamin Franklin was an American.

so they agree on 'designation theorems' like

> "Londres" designates (in French) London.

But the latter are all that our doubly-expanded translation manual has to add to the sorts of translation manuals which were argued above not to give us enough information about meaning. As above, it may be worth remarking how unsurprising this conclusion is. If we begin with a theory as far removed from semantics as a theory which pairs synonymous expressions without saying what either means, then it would be surprising if we could answer our questions about semantics by the addition of rules of inference which permit the removal of quotation marks, the replacement of 'means the same as' with 'means that', the replacement of 'designates the same as' with 'designates,' etc.

With this conclusion in hand, we can now ask: does the failure of our doubly-expanded translation manual to provide enough information about meaning show that Kölbel-style Davidsonian theories, too, fail in this regard? I think so; below I will defend the following argument:

1. Doubly-expanded translation manuals do not provide enough information about meaning.
2. Kölbel-style semantic theories do not provide any more information about meaning than do doubly-expanded translation manuals.
   _____
   C. Kölbel-style semantic theories do not provide enough information about meaning.

I've argued above for (1); now we turn to the argument for (2).

Doubly expanded translation manuals and neo-Davidsonian semantic theories have the same theorems, so there is no difference in informativeness be found there. The axioms of the two theories are different; the axioms of the Davidsonian theory will include, for example, claims about what singular terms in the language designate and about what objects satisfy the predicates in the language, whereas the axioms of the translation manual will include claims about which singular terms in one language designate the same thing as which singular terms in the other language, and which predicates in one language are satisfied by the same objects as which predicates in the other language. But, given the rules of inference which were included in the doubly-expanded translation manual, this can hardly justify the claim that a neo-Davidsonian theory gives more information about meaning than this translation manual; after all, given these rules of inference, the axioms of the doubly-expanded translation manual entail the axioms of the theory outlined by Kölbel. And a Davidsonian who adopts Kölbel's solution to the information problem can hardly reject the

inclusion of these rules of inference, since they are precisely analogous to those which license the derivation of meaning theorems from T-sentences.

It is natural to object at this point that a Kölbel-style Davidsonian theory passes a crucial test which even the doubly-expanded translation manual does not. This is the test Davidson had in mind when he wrote,

> "An acceptable theory should, as we have said, account for the meaning (or conditions of truth) of every sentence by analysing it as composed, in truth-relevant ways, of elements drawn from a finite stock."[13]

That is, it is natural to think that Davidsonian theories, but not the expanded translation manuals, meet some sort of compositionality requirement. This attempt to distinguish the Davidson theory from our expanded translation manuals is open to two objections. A first response is that the translation manuals we have been discussing *do* account for the meanings of sentences in terms of a finite number of elements of those sentences, combined in truth-relevant ways: after all, they deliver meaning theorems on the basis of translation theorems, which in turn follow from a finite number of axioms about which proper names designate the same thing, which predicates apply to the same things, etc. So, while a Kölbel-style Davidson theory passes this compositionality test, it seems that our expanded translation manuals will pass it as well.

But I think that it would be fair for the proponent of this compositionality-based objection to press his point one step further. For even if there is one clear sense in which our expanded translation manuals account for the meanings of sentences in terms of their parts, there is also a good sense in which they do not. And that is just that, if we take only the axioms of the translation manuals, we cannot, by logic alone, derive much at all about the meanings of the expressions in either language. We can only arrive at meaning theorems with the addition of rules of inference which go well beyond logic in allowing us to derive meaning theorems from translation theorems. So there is a sense in which, although the translation manual does arrive at meaning theorems for an infinite number of sentences via axioms concerning finitely many elements, the meaning theorems aren't obtainable solely (i.e., by logic alone) from these axioms. And this, one might think, is surely a sort of compositionality which a satisfactory semantic theory should exhibit.

But by this stage in the argument it is clear that this argument from compositionality has become less a vindication of Kölbel-style Davidsonian semantics than a diagnosis of its error. Just as the expanded translation manuals permit the derivation of meaning theorems only with a rule of inference which goes substantially beyond logic in allowing the derivation of meaning theorems from translation theorems, so Kölbel-style Davidsonian theories permit the derivation of meaning theorems only with the help of a rule of inference which goes substantially beyond logic in allowing us to derive meaning theorems from T-sentences. So, as was the case with expanded translation manuals, there is a sense in which, although a Kölbel-style semantic theory does arrive at meaning theorems for an infinite number of sentences via axioms concerning finitely many elements, the meaning theorems aren't obtainable solely

---

[13] "Semantics for Natural Languages," reprinted in his *Inquiries into Truth and Interpretation*, p. 56. Thanks to an anonymous reviewer for pressing this point.

(i.e., by logic alone) from these axioms. And if this is a kind of failure of compositionality in the case of the translation manuals, it is here as well.[14]

So either compositionality is a test which Kölbel-style theories fail, or it is a test which they pass in the bad company of doubly-expanded translation manuals. Either way, it provides no help for the neo-Davidsonian.

III. SEMANTIC THEORIES AS EXPLAINING UNDERSTANDING

The second goal of semantic theorizing endorsed by Kölbel and other Davidsonians is that a semantic theory for a language should explain the competence of speakers of the language by stating information knowledge of which would be sufficient for understanding the language in question.

In the present context, the point of invoking the aim of explaining semantic competence is the justification of neo-Davidsonian approaches to semantics. The idea is that, to the extent that neo-Davidsonian theories of the sort defended by Kölbel can explain linguistic competence, we are justified in adopting such theories. However, the requirement that semantic theories explain semantic competence admits of a stronger and a weaker interpretation. The key question for separating these two interpretations of this requirement is: Does the explanation of competence offered by the theory presuppose knowledge of another language, or not? The weaker version of the requirement that semantic theories explain competence with the object language is that knowledge of the semantic theory, plus knowledge of some other language, would be sufficient for competence; the stronger version is that knowledge of what is said by the semantic theory alone would suffice for competence.

---

[14]One might press the objection by emphasizing the qualification 'truth-relevant' in the quote from Davidson above. After all, while the doubly-expanded translation manual does generate meaning theorems on the basis of axioms concerning sub-sentential expressions, those axioms may not seem 'truth-relevant'; after all, they concern which expressions of another language designate the same thing, or are satisfied by the same things, as the sub-sentential expression in question, and not the contribution made by that expression to the determination of truth conditions. The doubly-expanded translation manual does have 'truth-relevant' theorems which state what sub-sentential expressions designate, are satisfied by, etc.; but one might object that these theorems are not used in the derivation of the meaning theorems.

But, I suggest, this objection only makes sense in the context of Davidson's original theory, which did genuinely attempt to explain meaning in terms of truth. In later versions of the Davidsonian program, as noted by Kölbel in the passage quoted above, it is hard to see the predicate 'is T' as anything other than a placeholder whose only role is to "generate theorems that pair object-language sentences with their metalanguage interpretations" (623). And if this is true of the predicate which figures in T-sentences, it's hard to see how it could fail to be true of the predicates – 'designates', 'is satisfied by' – which appear in the axioms whose role it is to generate those T-sentences. But, if this is true, then it's hard to see the axioms of a Kölbel-style Davidsonian theory as picking out truth-relevant properties of sub-sentential expressions any more than do the axioms of the doubly-expanded translation manual. Both are such that, given the rules of inference of their respective theories, they permit the derivation of truth-conditions (in the form of meaning theorems) for sentences in which they figure; but neither are any more 'truth-relevant' than that.

(Another, less interesting response to the objection is simply to modify the translation manual in such a way that the 'designation theorems' and 'satisfaction theorems' delivered by the doubly-expanded translation manual do play a role in the derivation of the meaning theorems for whole sentences. This could be done by leaving the axioms of the translation manual be as above, derive the designation and satisfaction theorems as in the doubly-expanded translation manual, and then let the derivation of translation theorems for whole sentences proceed via the derivation of T-sentences from these designation and satisfaction theorems.)

Thanks to an anonymous referee for *Linguistics & Philosophy* for suggesting this extension of the compositionality objection in the text.

I will argue that this distinction can be used to present a dilemma for the neo-Davidsonian. On the weaker interpretation, Kölbel-style semantic theories do meet the requirement of explaining competence; however, it turns out that this interpretation is so weak that many other sorts of theories, which are not plausibly regarded as theories of meaning for a language, also meet the requirement. Hence, the fact that it satisfies the weaker interpretation of the requirement that theories explain competence does not do much to justify Kölbel-style semantic theories. And, I will argue, Kölbel-style semantic theories do not meet the stronger version of the requirement.

To show the problems with the weaker interpretation, recall our original, unexpanded translation manual which, for each sentence of L, entails a translation theorem of the form

S in L means the same as S' in L'

This theory clearly meets the weaker interpretation of the requirement that a semantic theory explain competence. For surely if one knew such a translation manual, and understood the sentences of L', one could thereby arrive at an understanding of each of the sentences of L. But this fact should not convince anyone that translation manual semantics is a promising research program. A translation mapping from the sentences of English to the sentences of French simply does not explain the semantics of English; the fact that possession of such a manual, plus knowledge of French, would suffice for an understanding of any given sentence of English doesn't change this. I take this much to be uncontroversial. But, if this is uncontroversial, then it should also be uncontroversial that satisfaction of the weaker version of the requirement that a theory explain semantic competence does not do much to bolster a theory's semantic credentials.

So, if we are to make a case for neo-Davidsonian semantic theories based on their explanation of semantic competence, it will have to turn on the stronger interpretation of 'explaining competence.' Do neo-Davidsonian theories explain competence, in this sense?

To answer this question, it will be useful to return to our doubly-expanded translation manual. It might be thought that this sophisticated translation manual, unlike the un-expanded translation manual which has translation theorems as its output, meets the strong interpretation of the requirement that semantic theories explain competence. After all, the doubly-expanded translation manual has meaning theorems as its output; and meaning theorems, unlike translation theorems, make no reference to sentences of a language other than the object language.

That would be a mistake. Even though the theorems of the theory make no reference to another language, the explanation of how speakers are supposed to arrive at those theorems does. Speakers begin with knowledge of axioms which translate expressions of the object language into expressions of some other language; using these axioms, let us suppose that they are able to arrive at translation theorems for each sentence of the object language. At this point, the speaker knows what is said by a number of sentences of the form,

S in L means the same as S' in L'

It is crucial to note that, at this point in the derivation, we do *not* need to suppose that the speaker is competent with either L or L'; given a set of translation axioms,

I could derive translation theorems for two languages, neither of which I understand. But the speaker is, at this point, supposed to move from a translation theorem like

> "Londres est jolie" means the same in French as "London is pretty" in English.

to a meaning theorem like

> "Londres est jolie" means in French that London is pretty.

Now, as noted above, a speaker's derivation of the former needn't rely on knowledge of either French or English; indeed, there is no problem in the idea that a speaker can know the propositions expressed by both the translation theorem and the meaning theorem without knowledge of either French of English. The problems begin when we ask how speakers could master the rule of inference which takes them from translation theorems to meaning theorems.

How can a speaker know which meaning theorem is derivable from the translation theorem quoted above? We can think of the rule of inference which is supposed to guide the derivation as follows:

> If $\ulcorner S$ in L means the same as S' in $L'\urcorner$ is a theorem, then, if the language named by 'L'' is the language of the theory and 'p' is the sentence named by 'S'', the following is also a theorem: $\ulcorner S$ in L means that $p\urcorner$.

It is important to be clear about what this rule of inference gets you: the idea is that if you know that a certain translation theorem (i.e., a certain sentence) is a theorem of the theory, then the rule permits you to conclude that a certain meaning theorem (another sentence) is a theorem as well. But suppose that you know that a certain meaning theorem, say

> 'Londres est jolie' means that London is pretty.

is a theorem of the theory, and hence true. Knowledge that this sentence is true is not enough for you to know what 'Londres est jolie' means, just as knowing that 'The sky is blue' is true does not, by itself, tell you what color the sky is. In addition to knowing that these sentences are true, one must also know what these sentences mean. So any explanation of competence which makes use of this rule of inference will also make use of knowledge of what the theorems of the theory mean; and, if we are relying on what knowledge of the theorems of the theory mean, we are relying on competence with the language of the theory. This is enough to disqualify it from meeting the strong interpretation of the requirement that semantic theories explain semantic competence.[15]

---

[15]The proponent of explanation of linguistic competence offered by the doubly-expanded translation manual might object at this point that we have illicitly further strengthened the strong interpretation of the requirement that theories explain competence. Such a theorist might object that mastery of the relevant rule of inference just amounts to having a reliable ability to move from the translation theorems to the meaning theorems paired with them by the theory; in general, we don't need to suppose that this ability is underwritten by any particular sort of theoretical knowledge, and so in particular we do not have to suppose that this ability is underwritten by knowledge of the language of the theory. It is important to see that this move weakens the requirement of explaining

This might indicate that a more promising route would be to specify the relevant rule of inference from translation theorems to meaning theorems in terms which do not explicitly presuppose that there is a particular language in which the speaker represents the expanded translation manual. Along these lines, we might try something like the following:

> If S means in L the same thing as S' means in L', then, if p is the proposition expressed by S' in L', p is the meaning of S in L.[16]

This rule delivers knowledge of what is said by the relevant meaning theorems, rather than just knowledge of which meaning theorems express truths; but it wears its reliance on mastery of L' (the language of the theory) on its sleeve. In order to use this rule to infer meaning theorems from translation theorems, one has to know which propositions are expressed by which sentences of L'; and to presuppose this knowledge is to presuppose competence with L'.

The doubly-expanded translation manual is thus an example of a theory which, though its theorems are free from reference to sentences of any language other than the object language, makes covert appeal to knowledge of another language in explaining how speakers could arrive at knowledge of the meanings of the sentences of the object language. Hence it does not meet the stronger version of the requirement that semantic theories explain competence with the object language. Given the assumption which we have adopted for the purposes of argument — that semantic theories ought to meet the strong interpretation of the requirement that they explain competence — this counts against translation manual semantics. This, again, is no big surprise. The more interesting question is whether similar considerations count against the kind of neo-Davidsonian theory defended by Kölbel.

Like the doubly-expanded translation manual, a Kölbel-style semantic theory has meaning theorems as its output. As with the translation manual, we should ask how a speaker is to arrive at those theorems. Let us suppose that the speaker knows

---

competence to the point of triviality. If we allow the theories in question to include extra-logical rules of inference, and if we do not require any explanation of how speakers could come to master these rules of inference, then a great number of theories which have no plausible claim to being adequate semantic theories will count as explaining the competence of speakers. The doubly-expanded translation manual, we've seen, is one such theory. But there are more obvious examples. Imagine a theory which consists of an infinite number of axioms of the form 'S is a sentence of English' — one axiom for each sentence of English — and an infinite number of rules of inference (one for each axiom) which permit inferences from these axioms to the appropriate meaning theorems. Such a theory should not count as explaining the competence of speakers of English because we have no grip on how a speaker of English could master such a theory. This kind of example seems to indicate that knowledge of a theory (including mastery of its rules of inference) can only explain competence if we have some kind of grip on how someone could know that theory. The point above is just that the only way we can get a grip one how someone can master an expanded translation manual for a language L is by crediting them with knowledge of the language L' in which the translation manual is stated.

So a refusal to explain mastery of the rules of inference on the part of the proponent of translation manual semantics is a mistake, for the same reason as adoption of the weaker version of the requirement of explaining competence is a mistake. The point of the requirement that semantic theories explain competence is to justify Kölbel-style neo-Davidsonian theories. If we trivialize this requirement to the point that obviously non-semantic theories can meet it, any such justification that satisfaction of this requirement might have provided is lost.

[16]Here 'S' and 'S'' are variables over sentences, 'L' and 'L'' are variables over languages, and 'p' is a variable over propositions.

the axioms of a theory of truth for the object language L, and is able to derive interpretive T-sentences for each sentence of L. At this stage, the speaker knows, for each sentence of L, what is said by an interpretive T-sentence, for example

'Londres est jolie' is T in French iff London is pretty.

From this T-sentence, the speaker is supposed to employ Kölbel's rule of inference to arrive at the meaning theorem,

'Londres est jolie' means in French that London is pretty.

As above, we are licensed to ask whether mastery of this rule of inference relies on the speaker's competence with another language — in this case, the language of the theory.

We can begin by focusing on a formulation of the extra rule of inference which follows Kölbel's own:

If $\ulcorner S$ is true in L iff $p \urcorner$ is a theorem of the theory, then $\ulcorner S$ means that $p \urcorner$ is a theorem of the theory.[17]

It is clear, however, that for this rule of inference to yield knowledge of the meanings of sentences in L, the speaker must have prior knowledge of the language of the theory. After all, if the theory is to explain competence with L, what is needed is not an explanation of how a speaker could come to know that certain sentences are meaning theorems; what is needed is an explanation of how a speaker could come to know the propositions which are expressed by those meaning theorems. And to move from the former to the latter, the speaker has to understand the meaning theorems, which are stated in the language of the theory. So the use of this rule of inference to explain competence with L fails to meet the strong interpretation of our requirement for essentially the same reason that the first formulation of the extra rule of inference for the expanded translation manual does.[18]

As above, we could try to avoid the problem by formulating the rule of inference in terms of what is said by the theorems of the theory, without mentioning expressions in the language of the theory, as follows:

If it follows from the theory that S is true iff p, then S means that p.

But there is a problem with understanding this formula. 'S' can be understood as a universally quantified variable over sentences; but how is 'p' to be understood? One idea is that 'p' is a sentence letter, and that the above is a schema; on this formulation, to master this rule of inference is to know that every instance of the schema is true.

---

[17] See the discussion in Kölbel, pp. 619-623. Kölbel's formulation is complicated by his attention to the fact that, as he rightly notes, the theorem in question must be canonically derived from the axioms (because of the extension problem, discussed in §I above). I ignore this point here for simplicity.

[18] Kölbel is well aware that his extra rule of inference involves a meta-theoretic language which mentions expressions in the language of the theory; indeed, he argues convincingly that this does not make the rule of inference illegitimate (pp. 619-620). But what is at issue in the present discussion is not whether the rule of inference is legitimate, but whether one could master this rule of inference without already having knowledge of the language of the theory.

But an instance of a schema is a sentence; and if what this rule of inference gets us is that a certain sentence is true, then we are back in the problem we were trying to avoid. Knowledge that a certain meaning theorem is true yields knowledge of the meaning of the sentence mentioned in the theorem only if one understands the theorem; but, if we are making use of competence with the language in which the meaning theorems are stated, we have already given up hope of meeting the stronger version of the requirement that a theory explain competence.

So it seems that no formulation of the rule of inference which Kölbel adds to Davidsonian meaning theories can play a role in an explanation of competence with the object language which does not make use of prior competence with another language.[19]

One might be inclined to press the following objection to the above argument: any theory has to be stated in some language or other; hence any explanation of semantic competence based on knowledge of a semantic theory will presuppose understanding of the language of the theory. To argue that knowledge of a neo-Davidsonian theory for L cannot explain competence with L without the help of an understanding of the language of the theory is thus to argue against a straw man: *no theory* can meet the requirement of explaining competence, on the strong interpretation.[20]

There are two points to make in response to this objection. First, there is an important sense in which it does not matter if the objector is correct. If no theory can meet the strong interpretation of the requirement that semantic theories cannot explain competence, then, in particular, neo-Davidsonian theories cannot. But, given the failure of the weaker interpretation of that requirement (based on the fact that it is satisfied even by an un-expanded translation manual), this indicates that we lack an interpretation of the requirement that semantic theories explain competence which is both non-trivial and satisfied by Kölbel-style theories. This does not imply that there is no such requirement; perhaps there is some formulation of the requirement that semantic theories explain competence which is intermediate between the strong and weak interpretations, non-trivial, and satisfied by Davidsonian theories. I, for one, cannot see what such an interpretation could be. If, as I suggest, we conclude that there is no such interpretation, then we can only assess Davidsonian theories on the basis of how well they meet the requirement of giving the meanings of expressions of the object language, and we have already seen (in §II) that they do not fare well by that measure.

---

[19]This line of argument against Davidsonian explanations of semantic competence is inspired by the argument in Harman's "Language, Thought, and Communication", reprinted in Harman, *op. cit.*, pp. 166-182.

A second, related understanding of the above formulation of this rule of inference interprets 'p' as a substitutional variable. This fails for much the same reasons as the interpretation of 'p' as a sentence letter: knowledge of what is said by a sentence which employs substitutional quantification is parasitic on understanding the language in which the sentence is framed. (This is not, of course, true of ordinary objectual quantification; one can know the proposition expressed by 'Someone is happy' without understanding English.)

Yet a third attempt would be to try to understand 'p' as an objectual variable ranging over propositions. Then the rule of inference could be reformulated as follows:

If it follows from the theory that S is true iff p is true, then S means p.

The problem here is that a truth theory does not yield theorems of the form $\ulcorner S$ is true iff p is true$\urcorner$, where 'p' is replaced by a term which designates a proposition.

[20]Thanks to two anonymous referees for raising different versions of this objection.

The second reply is that it is not obvious that any explanation of competence in terms of knowledge of a semantic theory would presuppose competence with a language in which the theory is represented. Imagine, for example, a very simple language consisting only of atomic sentences of the form ⌜a is F⌝, where 'a' is a simple name and 'F' a simple predicate. Suppose now that we give a simple Russellian semantic theory for the language, according to which the meaning of every name is the object for which it stands, the meaning of every predicate is a property, and the meaning of every atomic sentence is the ordered pair consisting of the object which is the meaning of the name in the sentence, and the property which is the meaning of the predicate in the sentence; such propositions are true iff the object has the property.

This is obviously an extremely simple language. I introduce it only to point out that it is not obvious that knowledge of what is said by the axioms must involve competence with a language in which the axioms are statable, and that it is not obvious that knowledge of the meaning of each sentence — i.e., according to this theory, knowledge of which Russellian proposition is associated with each sentence — is only derivable from the propositions expressed by the axioms via the representation of those axioms in some language which the speaker in question understands. (There is, after all, no obvious general requirement that knowledge of a proposition entails competence with some sentence expressing that proposition.) This is enough to cast some doubt on the assumption that explanations of semantic competence based on knowledge of a semantic theory *must* always presuppose competence with a language in which the theory is stated. It is sometimes possible to know what is said by a theory without being competent with a language expressing the theory.

But this is a side point. The more important conclusion of this section for the purposes of the argument of this essay is that, whether or not explanations of semantic competence via a semantic theory inevitably presuppose prior competence with a language, we have found no reason to think that there is a non-trivial formulation of the requirement that semantic theories for a language should explain competence with that language which is met by neo-Davidsonian theories of the sort outlined by Kölbel.

IV. THE INFORMATION PROBLEM AND THE DAVIDSONIAN PROGRAM

So far I have argued that the response to the information problem suggested by Kölbel is inadequate: the theory which results from that response neither gives the meanings of sentences, nor explains semantic competence in any non-trivial sense. This raises the question of whether the objections raised against Kölbel generalize to other neo-Davidsonian attempts to solve the information problem. The purpose of this section is to argue that they do, by examination of two solutions to the information problem which are apparently quite different from Kölbel's: that offered by Davidson himself in his "Reply to Foster," and that defended by Larson and Segal in *Knowledge of Meaning.*

Davidson's own response to the information problem was to, in effect, add a rule of inference to the truth theory which allows the derivation from an interpretive T-sentence

S is T in L iff p

of a sentence of the form

> A T-theory, meeting such-and-such formal and empirical constraints, states that S is T in L iff p[21]

How might knowledge of what is said by an instance of the latter explain a speaker's understanding of S? Suppose that any T-theory meeting certain formal and empirical constraints has as theorems only interpretive T-sentences. Then suppose that our speaker knows this, and knows what is said by an instance of the above schema. In this case, she can infer that the sentence S mentioned on the left-hand side of the T-sentence entailed by the T-theory meeting these constraints means that p. And this knowledge is presumably sufficient for understanding S.[22]

The first thing to notice is the structural similarity to Kölbel's theory: both, in response to the information problem, posit an extra rule of inference which goes from T-sentences to another sort of sentence — in Kölbel's case, the latter is a meaning theorem, whereas in Davidson's, it is a claim about what follows from a theory of a certain description. We might formulate the rule of inference suggested by Davidson as follows:

> If $\ulcorner S$ is true in L iff $p \urcorner$ is a theorem of the theory, then $\ulcorner$A T-theory, meeting such and such formal and empirical constraints, states that S is T in L iff $p \urcorner$ is a theorem.

Given this rule of inference, it is not difficult to show that the objections raised above against Kölbel's solution to the information problem apply also to Davidson's. First, recall the argument of §II: expanded translation manuals do not give enough information about meaning to serve as semantic theories; Kölbel's semantic theory gives no more information about meaning than an expanded translation manual; hence Kölbel's theory does not give enough information about meaning to serve as a semantic theory. To turn this argument into an argument against Davidson's revised theory rather than Kölbel's, all that is needed is to change the characterization of our expanded translation manual. Rather than supplementing our original translation manual with the rule of inference

> If $\ulcorner S$ in L means the same as S' in $L' \urcorner$ is a theorem, then, if the language named by 'L'' is the language of the theory and 'p' is the sentence named by 'S'', the following is also a theorem: $\ulcorner S$ in L means that $p \urcorner$.

we can instead add to it the rule of inference

> If $\ulcorner S$ in L means the same as S' in $L' \urcorner$ is a theorem, then, if the language named by 'L'' is the language of the theory and 'p' is the sentence named by 'S'', the following is also a theorem: $\ulcorner$A T-theory, meeting such and such formal and empirical constraints, states that S is T in L iff $p \urcorner$.

---

[21] See Donald Davidson, "Reply to Foster," in *Truth and Meaning: Essays in Semantics*, *op. cit.*, pp. 33-41.

[22] Strictly, as Scott Soames has pointed out, this chain of inferences is invalid; it relies on a further premise connecting relations of entailment between propositions to relations of logical consequence between sentences. I ignore this point here. See Scott Soames, "Truth, Meaning, and Understanding," *Philosophical Studies* 65 (1992), pp. 17-35.

The rest of the argument of §II can then run as above. The argument of §III — that Kölbel's theory does not meet the strong interpretation of the requirement that semantic theories explain competence — does not need even this modification. The rule of inference added to Davidson's theory, like the rule of inference added to Kölbel's, delivers in the first instance knowledge of which theorems of the theory are true rather than knowledge of the propositions expressed by those theorems, and so does not, by itself, explain semantic competence with the object language. The only way to fill this gap is by requiring that speakers know the meaning of the theorems of the theory, which requires competence with the language of the theory; and this is exactly the kind of appeal to knowledge of another language which the strong interpretation of the competence requirement prohibits.[23]

Larson and Segal propose a solution to the information problem which is different in kind from the solutions offered by Davidson and Kölbel. In *Knowledge of Meaning*, they view semantics as the study of the unconscious possession of a truth theory by the semantic module of our language faculty.[24] Given this view of semantics, they can solve the information problem not by positing extra rules of inference, but rather by claiming that we simply *treat* the theorems of our internalized and interpretive T-theory as interpretive; that is, we interpret sentences by taking their meaning to be given by the sentence used on the right-hand side of the relevant T-sentence produced by the semantic module, even though that T-sentence does not entail a meaning theorem and does not say anything explicitly about meaning.[25]

Despite its important differences from the views of both Kölbel and Davidson, this position is open to essentially the same objections. Just as the argument of §II could be restated to apply to Davidson's revised position via a change in the additional rule of inference by which the manual is expanded, so too we can, in this case, imagine a translation manual expanded by the rule of inference

> If ⌜S in L means the same as S' in L'⌝ is a theorem, then, if the language named by 'L'' is the language of the theory and 'p' is the sentence named by 'S'', the following is also a theorem: ⌜S is true L iff p⌝.

By parallel argument to that laid out in detail in §II, it follows that a theory of the sort laid out by Larson and Segal gives no more information about meaning than this sort of expanded translation manual. The argument of §III also applies: in order for a speaker to be able to take a T-sentence as interpretive — i.e., take the sentence mentioned on the left-hand side of the T-sentence to mean the same thing as the sentence used on the right-hand side — the speaker must represent the T-sentence to himself in some language in which he is competent. Hence the theory fails to meet the strong interpretation of the requirement that semantic theories explain competence with the object language.[26]

---

[23] See §III and especially footnote 19 for the reasons why only the strong interpretation of the competence requirement can do any work, and why rules of inference like Davidson's and Kölbel's cannot avoid appeal to knowledge of the language of the theory by stating the relevant rules of inference using either substitutional quantification or objectual quantification over propositions.

[24] Larson & Segal, *op. cit.*, p. 22.

[25] *Ibid.*, pp. 38-9.

[26] One caveat: given the differences in the views of the nature of semantics between that implicit in Davidson's early work and that of *Knowledge of Meaning*, it may well be that Larson and Segal would disavow the traditional aims of Davidsonian semantic theories stated at the outset, and endorse the

The moral of the story, I think, is this. Davidson's original idea in "Truth and Meaning" embodied a daring and bold claim: that meaning was to be explicated in terms of the conceptually more fundamental notion of truth. The original form which this conjecture took — the idea that a truth theory could do the work of a theory of meaning — fails because of the information problem. Attempts have been made to save the form of the original idea, even if not its explanatory ambitions to elucidate meaning in terms of truth, by supplementing the truth theory with extra rules of inference which are supposed to provide the information which the truth theory does not. But these rules of inference cannot do the work for which they were enlisted: they do not enable a truth theory to give enough information about the meanings of expressions, and they do not provide a non-trivial explanation of semantic competence. I suggest that it is difficult to imagine other modifications of the original Davidsonian program which would fare any better.

If this is correct, we may be able to draw a more general lesson as well. One of the sources of interest in the Davidsonian program is that it provides a way to think systematically about the semantics of a language while avoiding commitment to, in Davidson's phrase "meanings as entities." If one thinks of Davidsonian semantics as the main alternative to semantic frameworks like Fregeanism and Russellianism which are straightforwardly committed to the existence of language-independent propositions, then the lesson is this: Davidsonian semantics does not provide a viable way to avoid commitment to the reality of content.[27]

---

conclusion that their preferred semantic theory does nothing that an expanded translation manual does not. This possibility raises meta-semantic questions which, while important, go beyond the scope of this essay.

[27]Thanks to Scott Soames, Brendan Gillon, and two anonymous referees for *Linguistics & Philosophy* for helpful comments on earlier versions of this paper.