Jason Bridges
Phil 340, Winter 2002

Davidson on the idea of a theory of meaning

I. Grice vs. Davidson

<u>Grice's project</u>**:** to give a **reductive analysis** of the notion of linguistic meaning. That is, to show that facts about linguistic meaning *consist in* facts that can be stated without any talk of meaning at all. Since linguistic meaning comes in a variety of forms, there will be many parts to Grice's analysis. Thus, Grice begins with an account of what *an utterer's meaning something by an utterance* consists in. (As we saw, Grice takes it to consist in the utterer's having a certain complicated set of intentions towards her audience.) From there he develops an account of what *a word's having the meaning x* consists in. (We will get to this part of the Gricean story later in the course). And so on.

<u>Davidson's project</u>**:** to explain what is involved in giving a **theory of meaning** for a particular language (e.g., English, Urdu). The *intuitive idea of a theory of meaning for a language* L is the idea of a theory that tells you what all of the words and sentences in L mean. In the course of executing his project, Davidson will sharpen this intuitive idea considerably.

<u>Initial comparison of the projects</u>**:** Even at this initial point, one difference is immediately obvious. A theory of meaning in Davidson's sense is an *empirical* theory of a *particular* language; arriving at the theory, if we don't already know the language in question, will involve considerable fieldwork. A reductive analysis of linguistic meaning, on the other hand, purports to reveal general and necessary truths about the nature of meaning as such. The difference between a reductive analysis of linguistic meaning and a theory of meaning for a particular language is precisely analogous to the difference between a reductive analysis of the notion of causation and a theory of the particular causal relations that hold among certain actual objects—say, among the water molecules in this bucket.

This may lead one to wonder how anyone could think that theories of meaning of particular languages are the business of philosophers, as opposed to empirical researchers. The answer is that Davidson's fundamental concern is not with constructing actual theories of meaning for actual languages (although he has also done work in this regard), but with getting clearer on what *any* such theory is supposed to look like. Davidson does not disagree with Grice that the notion of linguistic meaning could benefit from philosophical elucidation. He just thinks the way to undertake such elucidation is not to attempt a reductive analysis of linguistic meaning (by Davidson's lights, an impossible task), but instead to explain what shape a theory of meaning for a language ought to take.

<u>Davidson's aims, further elaborated</u>**:** What is involved in explaining the shape a theory of meaning for a language ought to take? For Davidson, such an explanation must specify two things:

1. the *form* that the axioms and theorems of the theory should take.
2. the *evidence* that should be used for testing the theory.

For Davidson, a rigorous explanation of both the form and evidence proper to a theory of meaning will provide as much of a philosophical elucidation of linguistic meaning as we need or could reasonably expect.
(The rest of this handout will focus only on the first of these aims.)

Note that Davidson's account, like Grice's, is primarily concerned with **natural languages**—languages that evolved naturally—rather than with **artificial languages**—languages consciously devised by individuals. Natural languages are the languages, with rare exception, that we all speak.

II. First constraint on the form of a theory of meaning: compositionality

Davidson's first constraint on the form of theories of meaning is the following:

> **The compositionality constraint**: A theory of meaning of a natural language L must show how the meanings of sentences of L are determined by properties of the simple expressions composing the sentences, coupled with the order in which the expressions appear.

Davidson states this constraint with the help of a distinction he borrows from mathematical and logical theories: the distinction between **axioms** and **theorems**. The axioms of a theory are its basic postulates; the theorems of a theory are the logical implications of its axioms. According to Davidson, a theory of meaning for a natural language L will respect the compositionality constraint if it consists of the following parts:

1. A set of axioms that assign semantic properties to each of the simple expressions of L.
2. A set of axioms that specify how the meanings of complex expressions are determined on the basis of the semantic properties of the simple expressions composing them.
3. A set of theorems, implied by these axioms, that give the meanings of all possible sentences of L.

<u>Why should a theory of meaning for a natural language respect the compositionality constraint</u>?
Davidson gives two related answers:

1. **The answer from finitude:** A theory of meaning aims to state what every expression in a language means. In the case of a natural language, we can't simply list all the expressions and say what each means, for the simple reason that there are an infinite number of sentences in a natural language. The only feasible solution is to show how, from a finite set of axioms ascribing meanings to a finite number of simple expressions, we can generate theorems that tell us the meaning of every possible sentence in the language.
2. **The interpretive answer:** Unless we view the sentences of a natural language as built up out of a finite stock of smaller parts, it will be impossible to interpret the language. (More on this later.)

Some philosophers have felt that these answers inappropriately portray the motivation for the constraint as *instrumental*, as dependent on the practical goals of the theorist. At least two additional answers have been defended in the literature (and occasionally ascribed to Davidson):

3. **The psychological answer.** A theory of meaning must aim to mirror a speaker's own knowledge of her language, and as a matter of human psychological fact, our knowledge of language exists within us in the form of a compositional theory. (The difficulty with this answer is to say in what sense we 'know' a compositional theory for our own language. Certainly this knowledge is not explicit; otherwise constructing a compositional theory of meaning would be a snap, rather than enormously difficult, as it in fact is.)
4. **The 'social-object' answer.** Natural languages are real objects in the world. A theory of meaning for such a language must aim to spell out how things actually stand with the language in question. And one way in which things stand with a natural language like English is that such a language is compositional. This is a real feature of a real social object, just as being a constitutional democracy is a real feature of the social object that is the United States. "I see language as a social object with a past, a present and a future….What syntax and semantics are answerable to is the state of this language, not the states of the speakers who aspire to speak that language" (Wiggins, "Meaning and Truth Conditions"). (We shall discuss both 2. and 3. at greater length in Part III of the course.)

III. Second constraint on the form of a theory of meaning: Convention T

What are the theorems of a theory of meaning going to look like? What is it to 'give the meaning' of a sentence?

One obvious possibility is that the theorems should be of one of the following forms:

> **S means p *or* S means that p** (where 'S' is replaced by an expression referring to a sentence and 'p' is replaced by an expression that refers to or states the meaning of the sentence in question.)

For example:

> "Snow is white" means that snow is white.

Davidson rejects this proposal, for two reasons:

1. Even a non-reductionist should balk at such an uncritical use of the very notion we're trying to elucidate. Surely we can do better than that.
2. For technical reasons, it is difficult to construct axioms that imply theorems of these forms.

Davidson thinks the proposal is instructive in one respect: it suggests that nothing is in a better position to 'give the meaning' of a sentence than that very sentence itself. Thus we can preserve from the proposal the idea that the theorems of theory of meaning should be of the following form:

> "Snow is white"…snow is white.

But we should fill in the gap with something more perspicuous than 'means' or 'means that'.

With what? What other filling could do the job? According to Davidson, an appealing answer is:

> "Snow is white" is true iff snow is white.

This sentence is certainly true. But can we reasonably view it as 'giving the meaning' of the sentence, "snow is white"? According to Davidson, we can, for doing so fits comfortably with an intuitive idea about meaning:

> **The truth-conditional view of sentence-meaning:** To understand an indicative sentence is to know the condition under which it is true.

Suppose I say, "My house is on fire." In order to understand what I have said, you must know that it is true iff the following conditions holds: my house is on fire.

As Wittgenstein, following Frege, expressed the idea (*Tractatus* 4.024): "To understand a sentence in use means to know what is the case if it is true."

This suggests that one way of giving the meaning of a sentence is to give the condition under which it is true.

Let us try to make the proposal more rigorous. Call the **object language** the language *for* which the theory of meaning is being given, and the **metalanguage** the language *in* which the theory is being given. Then when the object language and metalanguage are the *same*, the theorems of the theory are to have the following form:

> **S is true iff p** (where 'S' is replaced by an expression referring to a sentence in the object language and 'p' is replaced by that very sentence.)

(Terminological note: sentences of this general form are called **T-sentences**.)

The problem with the proposal as it stands is that it only works for the special case in which the object language and metalanguage are the same. We need a more general constraint. Taking a cue from a related suggestion of Tarski's, we might propose the following general condition:

> **Convention T (Tarski-style).** The theorems of a theory of meaning for an object language L must take the following form: **S is true iff p** (where 'S' is replaced by an expression in the metalanguage referring to a sentence in L and 'p' is replaced by a *translation* of that sentence into the metalanguage.)

But Davidson rejects framing a condition in these terms, for doing so relies on the notion of translation, which is too close to the notion of sentence-meaning that he seeks to elucidate.
So he opts for a different tack. He suggests a constraint that is cleansed of any mention of translation:

> **Convention T (Davidson's version).** The theorems of a theory of meaning for an object language L must take the form: **S is true iff p** (where 'S' is replaced by an expression in the metalanguage referring to a sentence in the object language, and 'p' is replaced by a sentence in the metalanguage that is true iff the object-language sentence is true.)

Abandoning the appeal to translation opens Davidson to the objection that his version of Convention T is too weak. All that his version requires is that the meta-language sentence on the right-hand side of the biconditional be true iff the object-language sentence referred to on the left-hand side is true. But consider the following two possible theorems, proposed as parts of theories of meaning for English and French respectively:

> "Snow is white" is true (in English) iff grass is green.
> "La neige est blanche" is true (in French) iff grass is green.

Both of these sentences satisfy Davidson's Convention T! But surely it would be a mistake to claim that they 'give the meaning' of the relevant object-language sentences in any sense, no matter how permissive, of that phrase.

Davidson's answer to this objection is that Convention T is intended to work only in conjunction with the other constraints he imposes on theories of meaning. There are two parts to this reply:

1. We have already discussed one of the additional constraints, the **compositionality constraint**. Taken together, the two constraints require that the same theory of meaning which produced the first T-sentence in the objection also produce T-sentences for the English sentences, "Snow is cold", "Clouds are green", etc., and that it do so on the basis of the semantic properties it assigns to the simple expressions, "snow", "white", "green", etc. It is very difficult to see how this could work. (This will become clearer when we look in detail at how the axioms of a theory of meaning are to be constructed.)
2. Davidson imposes a third constraint on a theory of meaning, the **interpretative constraint**, which requires that theorems of a theory of meaning for a language help to make total sense of the behavior of speakers of that language. This imposes further substantial limitations on the content of these theorems, ruling out, for example, the two T-sentences in the objection. (We will discuss this constraint when we discuss Davidson's account of the evidence proper to a theory of meaning.)

<u>Are sentences like, "'Snow is white' is true iff snow is white" trivially true</u>? No. We will discuss why in class.