

**Phil 435: Philosophy of Language**

**[Handout 5]**

**Donald Davidson  
Theory of Interpretation**

**Professor JeeLoo Liu**

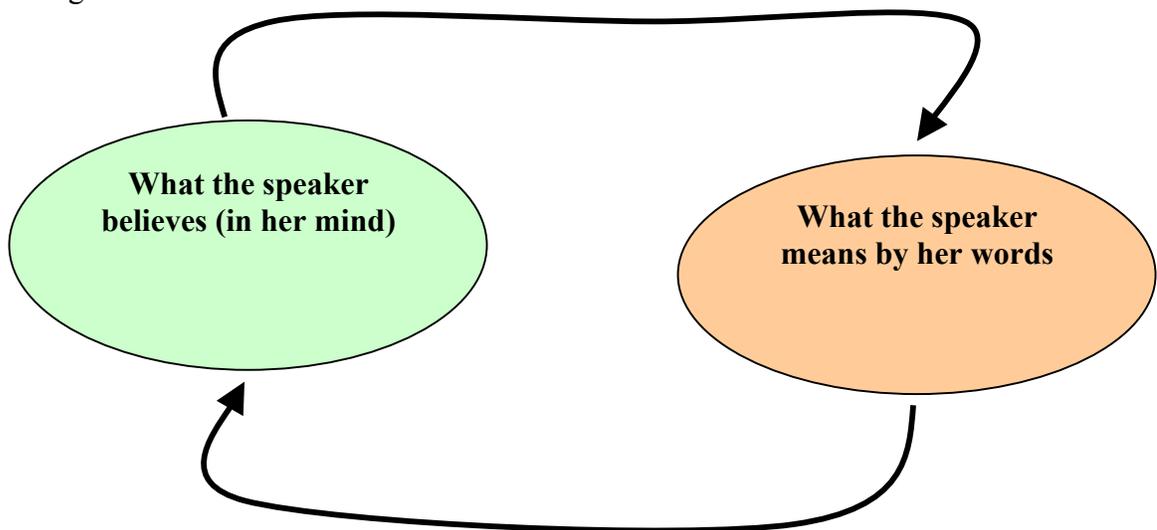
**Donald Davidson: 'Belief and the Basis of Meaning'  
(1974)**

**§ Theory of Interpretation**

1. Meaning and belief play interlocking and complementary roles in the interpretation of speech. We interpret a bit of linguistic behavior when we say what a speaker's words mean on an occasion of use.
2. Since a competent interpreter can interpret any of a potential infinity of utterances, ... what we must do is state a finite theory from which particular interpretations follow.
3. Theory of interpretation is the business jointly of the linguist, psychologist and philosopher. Its subject matter is the behavior of a speaker or speakers, and it tells what certain of their utterances mean.
4. The theory of interpretation can be used to describe what every interpreter knows, namely a specifiable infinite subset of the truths of the theory.

**[The difficulty in interpretation]: → the problem of circularity**

\_\_\_ A central source of trouble is the way beliefs and meanings conspire to account for utterances. A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.



## § The Problem of Circularity:

**Q: As a theory of interpretation, do we start by interpreting the speaker's utterances or by interpreting the speaker's beliefs?**

**Davidson: beliefs, intentions → utterances**

\_\_\_ Perhaps there are some who think it would be possible to establish the correctness of a theory of interpretation without knowing, or establishing, a great deal of beliefs, but it is not easy to imagine how it could be done.

\_\_\_ Far more plausible is the idea of deriving a theory of interpretation from detailed information about the intentions, desires, and beliefs of speakers (or interpreters, or both).

**But:**

\_\_\_ There is a principled, and not merely a practical, obstacle to verifying the existence of detailed, general and abstract beliefs and intentions, while being unable to tell what a speaker's words mean.

**My claim is only that making detailed sense of a person's intentions and beliefs cannot be independent of making sense of his utterances.**

It is not reasonable to suppose we can interpret verbal behavior without fine-grained information about beliefs and intentions, nor is it reasonable to imagine we can justify the attribution of preferences among complex options unless we can interpret speech behavior. A radical theory of decision must include a theory of interpretation and cannot presuppose it.

\* Where do we begin?

**[Radical interpretation]:**

\_\_\_ Since we cannot hope to interpret linguistic activity without knowing what a speaker believes, and cannot found a theory of what he means on a prior discovery of his beliefs and intentions, I conclude that in interpreting utterances from scratch – in *radical* interpretation – **we must somehow deliver simultaneously a theory of belief and a theory of meaning.**

Q: How is this possible?

## § How is radical interpretation possible?

**Basic Assumptions for Radical Interpretation:**

**1. The speaker's attitude of holding certain expressions true, relative to time.**

\_\_\_ We may as well suppose we have available all that could be known of such attitudes, past, present, and future.

**2. Common external circumstances**

\_\_\_ We can describe the external circumstances under which the attitudes hold or fail to hold.

### 3. Reliable judgment on the speaker's attitude.

\_\_\_ We can tell when a speaker holds a sentence to be true without knowing what he means by the sentence, or what beliefs he holds about its unknown subject matter, or what detailed intentions do or might prompt him to utter it.

**The problem, then is this: we suppose we know what sentences a speaker holds true, and when, and we want to know what he means and believes.**

1. Beliefs are best understood in their role of rationalizing choices or preferences.
2. The theory for which we should ultimately strive is one that takes as evidential base preferences between sentences – preferences that one sentence rather than another be true.
3. The theory would then explain individual preferences of this sort by attributing beliefs and values to the agent, and meanings to his words.

### § Background: Holism

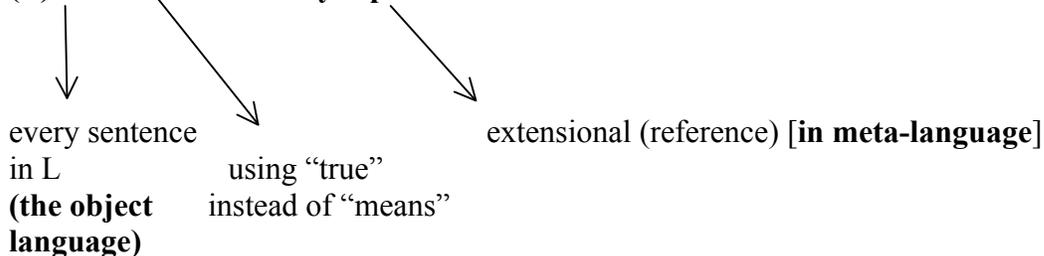
\_\_\_ **One can't interpret an utterance independent of one's knowledge of the speaker's whole set of beliefs and utterances.**

To interpret a particular utterance it is necessary to construct a comprehensive theory for the interpretation of a potential infinity of utterances. The evidence for the interpretation of a particular utterance will therefore have to be evidence for the interpretation of all utterances of a speaker or community.

### § The Method: Tarski's Convention T

\* [Convention T]

(T) *S* is True if and only if *p*.



Example of T-sentences:

\_\_\_ "Snow is white" is true if and only if snow is white.

\_\_\_ "雪是白色的" is true if and only if snow is white. (The T-sentences make a lot more sense when the object language is one with which we are unfamiliar.)

\* The main problem is to find a systematic way of matching predicates of the metalanguage to the primitive predicates of the object language so as to produce acceptable *T*-sentences.

*Q:* How does Convention T help solve the problem of radical interpretation?

\_\_\_ The evidential base for the theory will consist of facts about the circumstances under which speakers hold sentences of their language to be true. **Such evidence, I have urged, is neutral as between meaning and belief and assumes neither.**

\_\_\_ I propose that we take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima-facie evidence that the sentence is true under those circumstances.

Note: Davidson argues that we cannot start from beliefs/intentions to interpret the speaker's utterances, since to know what the speaker means we must know her beliefs and intentions first. With the method introduced here, we can simply derive the speaker's beliefs/intentions from her attitude (of holding sentences true) under our common (between the speaker and the interpreter) circumstances.

***Q:* How do we observe whether the speaker *holds the sentence true* under the given circumstance?**

### **§ How differences in individual speakers are grounded in overall coherence of all speakers:**

Not all the evidence can be expected to point the same way. There will be differences from speaker to speaker, and from time to time for the same speaker, with respect to the circumstances under which a sentence is held true. The general policy, however, is to choose truth conditions that do as well as possible in making speakers hold sentences true when (according to the theory and the theory builder's view of the facts) those sentences are true.

The aim is not the absurd one of making disagreement and error disappear. The point is that widespread agreement is the only possible background against which disputes and mistakes can be interpreted. Making sense of the utterances and behavior of others, even their most aberrant behavior, requires us to find a great deal of reason and truth in them. To see too much unreason on the part of others is simply to undermine our ability to understand what it is they are so unreasonable about. If the vast amount of agreement on plain matters that is assumed in communication escapes notice, it's because the shared truths are too many and too dull to bear mentioning. What we want to talk about is what's new, surprising, or disputed.

### **A further complication: indeterminacy of interpretation**

\_\_\_ A theory for interpreting the utterances of a single speaker, based on nothing but his attitudes towards sentences, would, we may be sure, have **many equally eligible rivals**, for differences in interpretation could be offset by appropriate differences in the beliefs attributed.

\_\_\_ The resulting indeterminacy of interpretation is the semantic counterpart of Quine's indeterminacy of translation, but the degree of indeterminacy will probably be less if my suggestions are followed than if Quine's are.

\_\_\_ My approach introduces formal constraints that do not apply in Quine's radical translation. The formal constraints are that the theory be finitely stated, and that it entails a *T*-sentence for each sentence of the object language.

\_\_\_ \* **Note: For Davidson, the indeterminacy of interpretation does not lead to the possibility of incommensurable (not capable of being compared because of the lack of a common quality for comparisons to be made) conceptual systems.**

**Davidson's optimism:**

1. Given a community of speakers with apparently the same linguistic repertoire, however, the theorist will strive for a single theory of interpretation: this will greatly narrow his practical choice of preliminary theories for each individual speaker.
2. We employ "the principle of charity."

**[the principle of charity]:**

\_\_\_ In cases of translation of someone's malapropism, for example, we should decide in favor of reinterpretation of words in order to preserve a reasonable theory of belief (rather than using a literal translation which makes him a fool).

**§ Common Language and Private Belief Structures (among speakers)**

\_\_\_ What makes a social theory of interpretation possible is what we can construct a plurality of private belief structures: belief is built to take up the slack between sentences held true by individuals and sentences true (or false) by public standards.

\_\_\_ What is private about belief is not that it is accessible to only one person, but that it may be idiosyncratic.

\_\_\_ Attributions of belief are as publicly verifiable as interpretations, being based on the same evidence: if we can understand what a person says, we can know what he believes.

**§ Conclusion:**

Each interpretation and attribution of attitude is a move within a holistic theory, a theory necessarily governed by

- (i) concern for consistency (**the principle of charity**)
- (ii) general coherence with the truth (**holding a sentence true under a given circumstance**)

and it is this that sets these theories forever apart from those that describe mindless objects, or describe objects as mindless.

The idea of the method I suggest is to show how we can have a workable theory of interpretation and a workable way of attributing beliefs without assuming that indeterminacy can be eliminated.

**Donald Davidson: A Nice Derangement of Epitaphs  
(1986)**

**\* The goal:**

\_\_\_ I want to know how people who already have a language manage to apply their skill or knowledge to actual cases of interpretation.

\_\_\_ My problem is to describe what is involved in the idea of 'having a language' or of being at home with the business of linguistic communication.

**§ Main Theses:**

1. **Malapropism is not a rare phenomenon in our language; it is ubiquitous. In all these cases of malapropism, the hearer has no trouble understanding the speaker in the way the speaker intends.**
2. **We want the deep concept to distinguish between what a speaker, on a given occasion, means, and what his words mean. The widespread existence of malapropisms and their kin threatens the distinction, since here the intended meaning seems to take over from the standard meaning.**
3. **We must pry apart what is literal in language from what is conventional or established.**
4. **In interpreting others' utterances, there are no learnable common core of consistent behavior, no shared grammar or rules, no portable interpreting machines set to grind out the meaning of an arbitrary utterance.**
5. **Interpretation comes at the converging points of passing theories (between the speaker and the hearer); linguistic ability is the ability to converge on a passing theory from time to time.**
6. **The theory we actually use to interpret an utterance is geared to the occasion (thus it can change from one occasion to the next).**
7. **We should give up the attempt to illuminate how we communicate by appeal to conventions.**

**[malapropism]:**

\_\_\_ mal·a·prop·ism n

1. the misuse of a word through confusion with another word that sounds similar, especially when the effect is ridiculous
2. an instance of using malapropism.

**Examples:**

\_\_\_ Lead the way and we'll precede.

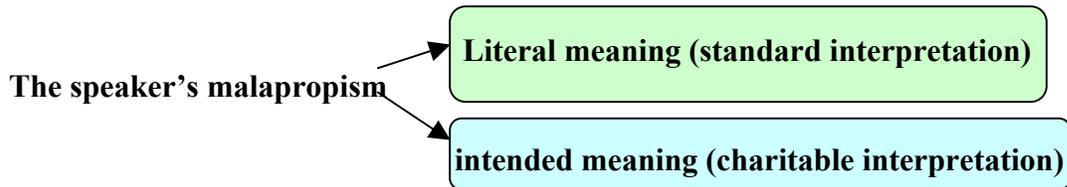
\_\_\_ We need a few laughs to break up the monogamy.

\_\_\_ Familiarity breeds attempt.

\_\_\_ We are all cremated equal.

**What is interesting is the fact that in all these cases the hearer has no trouble understanding the speaker in the way the speaker intends.**

Q: Why? How does the hearer realize that the standard interpretation cannot be the intended interpretation?



## § Literal Meaning

**[first meaning] = literal meaning**

\_\_\_ The concept applies to words and sentences as uttered by a particular speaker on a particular occasion. But if the occasion, the speaker, and the audience are ‘normal’ or ‘standard’, then the first meaning of an utterance will be what should be found by consulting dictionary based on actual usage.

### § Three principles concerning first meaning:

The particular difficulty with which I am concerned in this paper can be brought out by stating three plausible principles concerning first meaning in language: we may label them by saying they require that first meaning be systematic, shared, and prepared.

- (1) **First meaning is systematic.** A competent speaker or interpreter is able to interpret utterances, his own or those of others, on the basis of the semantic properties of the parts, or words, in the utterance, and the structure of the utterance. For this to be possible, there must be systematic relations between the meanings of utterances.
- (2) **First meanings are shared.** For speaker and interpreter to communicate successfully and regularly, they must share a method of interpretation of the sort described in (1).
- (3) **First meanings are governed by learned conventions or regularities.** The systematic knowledge or competence of the speaker or interpreter is learned in advance of occasions of interpretations and is conventional in character.

### \* The problem posed by malapropism

\_\_\_ Malapropisms fall into a different category, one that may include such things as our ability to perceive a well-formed sentence when the actual utterance was incomplete or grammatically garbled, our ability to interpret words we have never heard before, to correct slips of the tongue, or to cope with new idiolects. These phenomena threaten standard descriptions of linguistic competence.

\_\_\_ **How should we understand or modify (1) – (3) to accommodate malapropisms?**

**Davidson: The interpreter has a theory of interpretation.**

\_\_\_ To say that an explicit theory for interpreting a speaker is a model of the interpreter's linguistic competence is not to suggest that the interpreter knows any such theory.

\_\_\_ Claims about what would constitute a satisfactory theory ... are rather claims about what must be said to give a satisfactory description of the competence of the interpreter.

**In other words, we are trying to define conditions under which we can assign *linguistic competence* to the interpreter.**

***Q: Do we all speak the same language?***

It is an enormous convenience that many people speak in similar ways, and therefore can be interpreted in more or less the same way. But in principle communication does not demand that any two people speak the same language. What must be shared is the interpreter's and the speaker's understanding of the speaker's words.

### **§ The Prior Theory vs. the Passing Theory**

An interpreter has, at any moment of a speech transaction, what I persist in calling a theory. (I call it a theory... only because a description of the interpreter's competence requires a recursive account).

I assume that the interpreter's theory has been adjusted to the evidence so far available to him: knowledge of the character, dress, role, sex of the speaker, and whatever else has been gained by observing the speaker's behavior, linguistic or otherwise.

**The prior theory for the hearer** \_\_\_ how he is prepared in advance to interpret an utterance of the speaker.

**The passing theory for the hearer** \_\_\_ how he *does* interpret the utterance.

**The prior theory for the speaker** \_\_\_ what he *believes* the interpreter's prior theory to be.

**The passing theory for the speaker** \_\_\_ the theory he *intends* the interpreter to use.

The speaker says something with the intention that it will be interpreted in a certain way, and the expectation that it will be so interpreted. In fact this way is not provided for by the interpreter's theory. But the speaker is nevertheless understood; the interpreter adjusts his theory so that it yields the speaker's intended interpretation.

### **[The Opponent's View: Theory of Convention]**

\_\_\_ According to that account, each interpreter (**and this includes speakers, since speakers must be interpreters**) comes to a successful linguistic competence, and which he shares with those with whom he communicates. Because each party has such a shared theory and knows that others share his theory, and knows that others know he knows, some would say that the knowledge or abilities that constitute the theory may be called conventions.

### \* Davidson's Rebuttal

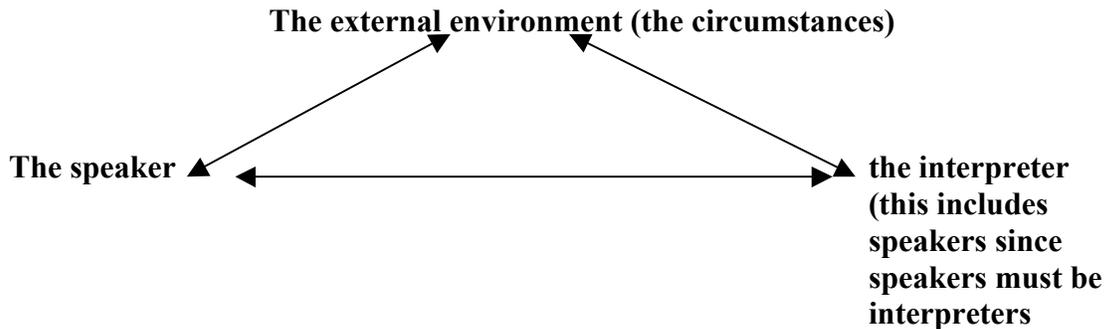
1. What must be shared for communication to succeed is the one the interpreter actually uses to interpret an utterance (the hearer's passing theory), and it is the theory the speaker intends the interpreter to use (the speaker's passing theory).
2. Only if these passing theories coincide is understanding complete.
3. But the passing theory cannot in general correspond to an interpreter's linguistic competence.
4. It is quite clear that in general the prior theory is neither shared by speaker and interpreter nor is it what we would normally call a language.
5. Therefore, the passing theory is not governed by conventions.
6. Therefore, the theory of convention does not explain communication.

### § Davidson's Theory of Interpretation

1. Interpretation is complete only when the speaker's passing theory and the hearer's passing theory coincide.
2. As the speaker speaks his piece the interpreter alters his theory, entering hypotheses about new names, altering the interpretation of familiar predicates, and revising past interpretations of particular utterances in light of new evidence.
3. As speaker and interpreter talk, their prior theories become more alike; so do their passing theories. The asymptote of agreement and understanding is when passing theories coincide.
4. But the passing theory cannot in general correspond to an interpreter's linguistic competence. Not only does it have its changing list of proper names and gerrymandered vocabulary, but it includes every successful – i.e., correctly interpreted – use of any other word or phrase, no matter how far out of the ordinary.
5. Every derivation from ordinary usage, as long as it is agreed on for the moment (knowingly deviant, or not, on one, or both, sides), is in the passing theory as a feature of what the words mean on that occasion.
6. A passing theory is not a theory of what anyone would call an actual natural language. 'Mastery' of such a language would be useless, since knowing a passing theory is only knowing how to interpret a particular utterance **on a particular occasion** (the theory we actually use to interpret an utterance is geared to the occasion).
7. Of course things previously learned were essential to arriving at the passing theory, but what was learned could not have been the passing theory.
8. An interpreter must be expected to have quite different prior theories for different speakers – not as different, usually, as his passing theories; but these are matters that depend on how well the interpreter knows his speaker.
9. The general framework or theory, whatever it is, may be a key ingredient in what is needed for interpretation, but it can't be all that is needed since it fails to provide the interpretation of particular words and sentences as uttered by a particular speaker.

10. A passing theory really is like a theory at least in this, that it is derived by wit, luck, and wisdom from **a private vocabulary and grammar**, knowledge of the ways people get their point across, and rules of thumb for figuring out what derivations from the dictionary are most likely.

### Davidson's Tripartite Theory



### § Conclusion

Most of the time prior theories will not be shared, and there is no reason why they should be. What must be shared for communication to succeed is the one the interpreter *actually* uses to interpret an utterance, and it is the theory the speaker *intends* the interpreter to use. Only if these coincide is understanding complete.

The prior theory has in it all the features special to the idiolect of the speaker and the interpreter is in a position to take into account before the utterance begins. One way to appreciate the difference between the prior theory and our ordinary idea of a person's language is to reflect on the fact that an interpreter must be expected to have quite different prior theories for different speakers – not as different, usually, as his passing theories; but these are matters that depend on how well the interpreter knows his speaker.

**Focus Questions for Essay #4: [This time you'll get the option to answer both or just choose one, because both articles are sophisticated and require careful reading. Of course you are welcome to answer both questions too.]**

- (1) **Putnam:** What does Putnam's *Twin Earth* example show? ) What is his theory of *indexicality* and the *hypothesis of the division of linguistic labor*? How does Putnam use these two theses to argue that 'meaning just ain't in the head'?
- (2) **Burge:** What does Burge's thought experiment demonstrate? How does Burge use this example to argue that 'beliefs are not in the head either'?