Radical Interpretation

by Donald Davidson

Kurt utters the words "Es regnet" and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. We could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we *do* know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation. The second question, how we could come to have knowledge that would serve to yield interpretations, does not, of course, concern the actual history of language acquisition. It is thus a doubly hypothetical question: given a theory that would make interpretation possible, what evidence plausibly available to a potential interpreter would support the theory to a reasonable degree? In what follows I shall try to sharpen these questions and suggest answers.

The problem of interpretation is domestic as well as foreign: it surfaces for speakers of the same language in the form of the question, how can it be determined that the language is the same? Speakers of the same language can go on the assumption that for them the same expressions are to be interpreted in the same way, but this does not indicate what justifies the assumption. All understanding of the speech of another involves radical interpretation. But it will help keep assumptions from going unnoticed to focus on cases where interpretation is most clearly called for: interpretation in one idiom of talk in another. [1]

What knowledge would serve for interpretation? A short answer would be, knowledge of what each meaningful expression means. In German, those words Kurt spoke mean that it is raining and Kurt was speaking German. So in uttering the words "Es regnet", Kurt said that it was raining. This reply does not, as might first be thought, merely restate the problem. For it suggests that in passing from a description that does not interpret (his uttering of the words "Es regnet") to interpreting description (his saying that it is raining) we must introduce a machinery of words and expressions (which may or may not be exemplified in actual utterances), and this suggestion is important. But the reply is no further help, for it does not say what it is to know what an expression means.

There is indeed also the hint that corresponding to each meaningful expression there is an entity, its meaning. This idea, even if not wrong, has proven to be very little help: at best it hypostasizes the problem.

Disenchantment with meanings as implementing a viable account of communication or interpretation helps explain why some philosophers have tried to get along without, not only meanings, but any serious theory at all. It is tempting, when the concepts we summon up to try to explain interpretation turn out to be more baffling than the explanandum, to reflect that after all verbal communication consists in nothing more than elaborate disturbances in the air which form a causal link between the non-linguistic activities of human agents. But although interpretable speeches are nothing but (that is, identical with) actions performed with assorted non-linguistic intentions (to warn, control, amuse, distract, insult), and these actions are in turn nothing but (identical with) intentional movements of the lips and larynx, this observation takes us no distance towards an intelligible general account of what we might know that would allow us to redescribe uninterpreted utterances as the right interpreted ones.

Appeal to meanings leaves us stranded further than we started from the non-linguistic goings on that must supply the evidential base for interpretation; the "nothing but" attitude provides no clue as to how the evidence is related to what it surely is evidence for.

Other proposals for bridging the gap fall short in various ways. The "causal" theories of Ogden and Richards and of Charles Morris attempted to analyze the meaning of sentences, taken one at a time, on the basis of behaviouristic data. Even if these theories had worked

for the simplest sentences (which they clearly did not), they did not touch the problem of extending the method to sentences of greater complexity and abstractness. Theories of another kind start by trying to connect words rather than sentences with non-linguistic facts. This is promising because words are finite in number while sentences are not, and yet each sentence is no more than a concatenation of words: this offers the chance of a theory that interprets each of an infinity of sentences using only finite resources. But such theories fail to reach the evidence, for it seems clear that the semantic features of words cannot be explained directly on the basis of non-linguistic phenomena. The reason is simple. The phenomena to which we must turn are the extra-linguistic interests and activities that language serves, and these are served by words only in so far as the words are incorporated in (or on occasion happen to be) sentences. But then there is no chance of giving a foundational account of words before giving one of sentences.

For quite different reasons, radical interpretation cannot hope to take as evidence for the meaning of a sentence an account of the complex and delicately discriminated intentions with which the sentence is typically uttered. It is not easy to see how such an approach can deal with the structural, recursive feature of language that is essential to explaining how new sentences can be understood. But the central difficulty is that we cannot hope to attach a sense to the attribution of finely discriminated intentions independently of interpreting speech. The reason is not that we cannot ask necessary questions, but that interpreting an agent's intentions, his beliefs and his words are parts of a single project, no part of which can be assumed to be complete before the rest is. If this is right, we cannot make the full panoply of intentions and beliefs the evidential base for a theory of radical interpretation.

We are now in a position to say something more about what would serve to make interpretation possible. The interpreter must be able to understand any of the infinity of sentences the speaker might utter. If were are to state explicitly what the interpreter might know that would enable him to do this, we must put it in finite form.[2] If this requirement is to be met, any hope of a universal method of interpretation must be abandoned. The most that can be expected is to explain how an interpreter could interpret the utterances of speakers of a single language (or a finite number of languages): it makes no

sense to ask for a theory that would yield an explicit interpretation for any utterance in any (possible) language.

It is still not clear, of course, what it is for a theory to yield an explicit interpretation of any utterance. The formulation of the problem seems to invite us to think of the theory as the specification of a function taking utterances as arguments and having interpretations as values. But then interpretations would be no better than meanings and just as surely entities of some mysterious kind. So it seems wise to describe what is wanted of the theory without apparent reference to meanings or interpretations: someone who knows the theory can interpret the utterances to which the theory applies.

The second general requirement on a theory of interpretation is that it can be supported or verified by evidence plausibly available to an interpreter. Since the theory is general—it must apply to a potential infinity of utterances—it would be natural to think of evidence in its behalf as instances of particular interpretations recognised as correct. And this case does, of course, arise for the interpreter dealing with a language he already knows. The speaker of a language normally cannot produce an explicit finite theory for his own language, but he can test a proposed theory since he can tell whether it yields correct interpretations when applied to particular utterances.

In radical interpretation, however, the theory is supposed to supply an understanding of particular utterances that is not given in advance, so the ultimate evidence for the theory cannot be correct sample interpretations. To deal with the general case, the evidence must be of a sort that would be available to someone who does not already know how to interpret utterances the theory is designed to cover: it must be evidence that can be stated without essential use of such linguistic concepts as meaning, interpretation, synonymy and the like.

Before saying what kind of theory I think will do the trick, I want to discuss a last alternative suggestion, namely that a method of translation, from the language to be interpreted into the language of the interpreter, is all the theory that is needed. Such a theory would consist in the statement of an effective method for going from an arbitrary sentence of the alien tongue to a sentence of a familiar language; thus it would satisfy the demand for a finitely stated method applicable to any sentence.[3] But I do not think a translation manual is the best form for a theory of translation to take.

When interpretation is our aim, a method of translation deals with a wrong topic, a relation between two languages, where what is wanted is an interpretation of one (in another, of course, but that goes without saying since any theory is in some language). We cannot without confusion count the language used in stating the theory as part of the subject matter of the theory unless we explicitly make it so. In the general case, a theory of translation involves three languages: the object language, the subject language, and the metalanguage (the languages from and into which translation proceeds, and the language of the theory, which says what expressions of the subject language translate which expressions of the object language). And in this general case, we can know which sentences of the subject language translate which sentences of the object language without knowing what any of the sentences of either language mean (in any sense, anyway, that would let someone who understood the theory interpret sentences of the object language). If the subject language happens to be identical with the language of the theory, then someone who understands the theory can no doubt use the translation manual to interpret alien utterances; but this is because he brings to bear two things he knows and that the theory does not state: the fact that the subject language is his own, and his knowledge of how to interpret utterances in his own language.

It is awkward to try to make explicit the assumption that a mentioned sentence belongs to one's own language. We could try, for example, "'Es regnet' in Kurt's language is translated as 'It is raining' in mine", but the indexical self reference is out of place in a theory that ought to work for any interpreter. If we decide to accept this difficulty, there remains the fact that the method of translation leaves tacit and beyond the reach of theory what we need to know that allows us to interpret our own language. A theory of translation must read some sort of structure into sentences, but there is no reason to expect that it will provide any insight into how the meanings of sentences depend on their structure.

A satisfactory theory for interpreting the utterances of any language, our own included, will reveal significant semantic structure: the interpretation of utterances of complex sentences will systematically depend on the interpretation of utterances of simpler sentences, for example. Suppose we were to add to a theory of translation a satisfactory theory of interpretation for our own language. Then we would have exactly what we want, but in an unnecessarily bulky form. The translation

manual churns out, for each sentence of the language to be translated, a sentence of the translator's language; the theory of interpretation then gives the interpretation of these familiar sentences. Clearly the reference to the home language is superfluous; it is an unneeded intermediary between interpretation and alien idiom. The only expressions a theory of interpretation has to mention are those belonging to the language to be interpreted.

A theory of interpretation for an (unknown) object language may then be viewed as the result of the merger of a structurally revealing theory of interpretation for a known language, and a system of translation from the unknown language into the known. The merger makes all reference to the known language otiose; when this reference is dropped, what is left is a structurally revealing theory of interpretation for the unknown language—couched, of course, in familiar words. We have such theories, I suggest, in theories of truth of the kind Tarski first showed how to give [4].

What characterises a theory of truth in Tarski's style is that it entails, for every sentence s of the object language, a sentence of the form:

s is true (in the object language) if and only if p

Instances of the form (which we shall call T-sentences) are obtained by replacing "s" by a canonical description of s, and "p" by a translation of s. The important undefined semantical notion in the theory is that of *satisfaction* which relates sentences, open or closed, to infinite sequences of objects, which may be taken to belong to the range of the variables of the object language. The axioms, which are finite in number, are of two kinds: some give the conditions under which a sequence satisfies a complex sentence on the basis of the conditions of satisfaction of simpler sentences, others give the conditions under which the simplest (open) sentences are satisfied. Truth is defined for closed sentences in terms of the notion of satisfaction. A recursive theory like this can be turned into an explicit definition along familiar lines, as Tarski shows, provided the language of the theory contains enough set theory; but we shall not be concerned with this extra step.

Further complexities enter if proper names and functional expressions are irreducible features of the object language. A trickier matter concerns indexical devices. Tarski was interested in formalized languages containing no indexical or demonstrative aspects. He could

therefore treat sentences as vehicles of truth; the extension of the theory to utterances is in this case trivial. But natural languages are indispensably replete with indexical features, like tense, and so their sentences may vary in truth according to time and speaker. The remedy is to characterize truth for a language relative to a time and a speaker. The extension to utterances is again straightforward. [5]

What follows is a defence of the claim that a theory of truth, modified to apply to a natural language, can be used as a theory of interpretation. The defence will consist in attempts to answer three questions:

1. Is it reasonable to think that a theory of truth of the sort described can be given for a natural language?
2. Would it be possible to tell that such a theory was correct on the basis of evidence plausibly available to an interpreter with no prior knowledge of the language to be interpreted?
3. If the theory were known to be true, would it be possible to interpret utterances of speakers of the language?

The first question is addressed to the assumption that a theory of truth can be given for a natural language; the second and third questions ask whether such a theory would satisfy the demands we have made on a theory of interpretation.

1. *Can a theory of truth be given for a natural language?*

It will help us to appreciate the problem to consider briefly the case where a significant fragment of a language (plus one or two semantical predicates) is used to state its own theory of truth. According to Tarski's Convention T, it is a test of the adequacy of a theory that it entails all the T-sentences. This test apparently cannot be met without assigning something very much like a standard quantificational form to the sentences of the language, and appealing, in the theory, to a relational notion of satisfaction [6]. But the striking thing about T-sentences is that whatever machinery must operate to produce them, and whatever ontological wheels must turn, in the end a T-sentence states the truth conditions of a sentence using resources no richer than, because the same as, those of the sentence itself. Unless the original sentence mentions possible worlds, intensional entities, properties or propositions, the statement of its truth conditions does not.

There is no equally simple way to make the analogous point about an alien language without appealing, as Tarski does, to an unanalysed notion of translation. But what we can do for our own language we ought to be able to do for another; the problem, it will turn out, will be to know that we are doing it.

The restriction imposed by demanding a theory that satisfies Convention T seems to be considerable: there is no generally accepted method now known for dealing, within the restriction, with a host of problems, for example, sentences that attribute attitudes, modalities, general causal statements, counterfactuals, attributive adjectives, quantifiers like "most", and so on. On the other hand, there is what seems to me to be fairly impressive progress. To mention some examples, there is the work of Tyler Burge on proper names [7], Gilbert Harman on "ought" [8], John Wallace on mass terms and comparatives [9], and there is my own work on attributions of attitudes and performatives [10], on adverbs, events and singular causal statements [11], and on quotation [12].

If were are inclined to be pessimistic about what remains to be done (or some of what has been done!), we should think of Frege's magnificent accomplishment in bringing what Dummett calls "multiple generality" under control [13]. Frege did not have a theory of truth in Tarski's sense in mind, but it is obvious that he sought, and found, structures of a kind for which a theory of truth can be given.

The work of applying a theory of truth in detail to a natural language will in practice almost certainly divide into two stages. In the first stage, truth will be characterized, not for the whole language, but for a carefully gerrymandered part of the language. This part, though no doubt clumsy grammatically, will contain an infinity of sentences which exhaust the expressive power of the whole language. The second part will match each of the remaining sentences to one or (in the case of ambiguity) more than one of the sentences for which truth has been characterized. We may think of the sentences to which the first stage of the theory applies as giving the logical form, or deep structure, of all sentences.

2. *Can a theory of truth be verified by appeal to evidence available before interpretation has begun?*

Convention T says that a theory of truth is satisfactory if it generates a T-sentence for each sentence of the object language. It is

enough to demonstrate that a theory of truth is empirically correct, then, to verify that the T-sentences are true (in practice, an adequate sample will confirm the theory to a reasonable degree). T-sentences mention only the closed sentences of the language, so the relevant evidence can consist entirely of facts about the behaviour and attitudes of speakers in relation to sentences (no doubt by way of utterances). A workable theory must, of course, treat sentences as concatenations of expressions of less than sentential length, it must introduce semantical notions like satisfaction and reference, and it must appeal to an ontology of sequences and the objects ordered by the sequences. All this apparatus is properly viewed as theoretical construction, beyond the reach of direct verification. It has done its work provided only it entails testable results in the form of T-sentences, and these make no mention of the machinery. A theory of truth thus reconciles the demand for a theory that articulates grammatical structure with the demand for a theory that can be tested only by that it says about sentences.

In Tarski's work, T-sentences are taken to be true because the right branch of the biconditional is assumed to be a translation of the sentence truth conditions for which are being given. But we cannot assume in advance that correct translation can be recognised without preempting the point of radical interpretation; in empirical applications, we must abandon the assumption. What I propose is to reverse the direction of explanation: assuming translation, Tarski was able to define truth; the present idea is to take truth as basic and to extract an account of translation or interpretation. The advantages, from the point of view of radical interpretation, are obvious. Truth is a single property which attaches, or fails to attach, to utterances, while each utterance has its own interpretation; and truth is more apt to connect with fairly simple attitudes of speakers.

There is no difficulty in rephrasing Convention T without appeal to the concept of translation: an acceptable theory of truth must entail, for every sentence s of the object language, a sentence of the form: s is true if and only if p, where "p" is replaced by any sentence that is true if and only if s is. Given this formulation, the theory is tested by evidence that T-sentences are simply true; we have given up the idea that we must also tell whether what replaces "p" translates s. It might seem that there is no chance that if we demand so little of T-sentences, a theory of interpretation will emerge. And of course this would be so if we took the T-sentences in isolation. But the hope is

that by putting appropriate formal and empirical restrictions on the theory as a whole, individual T-sentences will in fact serve to yield interpretations.

We have still to say what evidence is available to an interpreter—evidence, we now see, that T-sentences are true. The evidence cannot consist in detailed descriptions of the speaker's beliefs and intentions, since attributions of attitudes, at least where subtlety is required, demand a theory that must rest on much the same evidence as interpretation. The interdependance of belief and meaning is evident in this way: a speaker holds a sentence to be true because of what the sentence (in his language) means, and because of what he believes. Knowing that he holds the sentence to be true, and knowing the meaning, we can infer his belief; given enough information about his beliefs, we could perhaps infer the meaning. But radical interpretation should rest on evidence that does not assume knowledge of meanings or detailed knowledge of beliefs.

A good place to begin is with the attitude of holding a sentence true, of accepting it as true. This is, of course, a belief, but it is a single attitude applicable to all sentences, and so does not ask us to be able to make finely discriminated distinctions among beliefs. It is an attitude an interpreter may plausibly be taken to be able to identify before he can interpret, since he may know that a person intends to express a truth in uttering a sentence without having any idea *what* truth. Not that sincere assertion is the only reason to suppose that a person holds a sentence to be true. Lies, commands, stories, irony, if they are detected as attitudes, can reveal whether a speaker holds his sentences to be true. There is no reason to ruel out other attitudes towards sentences, such as wishing true, wanting to make true, believing one is going to make true, and so on, but I am inclined to think that all evidence of this kind may be summed up in terms of holding sentences to be true.

Suppose, then, that the evidence available is just that speakers of the language to be interpreted hold various sentences to be true at certain times and under specified circumstances. How can this evidence be used to support a theory of truth? On the one hand, we have T-sentences, in the form:

(T) "Es regnet" is true-in-German when spoken by x at time t if and only if it is raining near x at t.

On the other hand, we have the evidence, in the form :

(E) Kurt belongs to the German speech community and Kurt holds true "Es regnet" on Saturday at noon and it is raining near Kurt on Saturday at noon.

We should, I think, consider (E) as evidence that (T) is true. Since (T) is a universally quantified conditional, the first step would be to gather more evidence to support the claim that:

(GE) $(x)(t)$(if x belongs to the German speech community then (x holds true "Es regnet" at t if and only if it is raining near x at t))

The appeal to a speech community cuts a corner but begs no question: speakers belong to the same speech community if the same theories of interpretation work for them.

The obvious objection is that Kurt, or anyone else, may be wrong about whether it is raining near him. And this is of course a reason for not taking (E) as conclusive evidence for (GE) or for (T) ; and a reason not to expect generalisations like (GE) to be more than generally true. The method is rather one of getting a best fit. We want a theory that satisfies the formal constraints on a theory of truth, and that maximizes agreement, in the sense of making Kurt (and others) right, as far as we can tell, as often as possible. The concept of maximization cannot be taken literally here, since sentences are infinite in number, and anyway once the theory begins to take shape it makes sense to accept intelligible error and to make allowance for the relative likelihood of various kinds of mistake.

The process of devising a theory of truth for an unknown native tongue might in crude outline go as follows. First we look for the best way to fit our logic, to the extent required to get a theory satisfying Convention T, onto the new language; this means reading the logical vocabulary of first order quantification theory (plus identity) into the language, not taking the logical constants one by one, but treating this much of logic as a grid to be fitted onto the language in one fell swoop. The evidence here is classes of sentences always held true or always held false by almost everyone almost all of the time (potential logical truths) and patterns of inference. The first step identifies predicates, singular terms, quantifiers, connectives, and identity; in theory, it settles matters of logical form. The second step concentrates on sentences with indexicals: those sentences sometimes held true and

sometimes false according to discoverable changes in the world. This step in conjunction with the first limits the possibilities for interpreting individual predicates (for intuitively that is what we hope from T-sentences). The last step deals with the remaining sentences, those on which there is not uniform agreement, or whose held truth value does not depend systematically on changes in the environment. [14]

This method is intended to solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning. This is accomplished by assigning truth conditions to alien sentences that make native speakers right as often as plausibly possible, according, of course, to our own view of what is right. What justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement. Applied to language, this principle reads: the more sentences we conspire to accept or reject (whether or not through a medium of interpretation), the better we understand the rest, whether or not we agree about them.

The methodological advice to interpret in a way that optimizes agreement should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything.

Here I would like to insert a remark about the methodology of my proposal. In philosophy we are used to definitions, analyses, reductions. Typically these are intended to carry us from concepts better understood, or clearer, or more basic epistemologically or ontologically, to others we want to understand. The method I have suggested fits none of these categories. I have proposed a looser relation between concepts to be illuminated and the relatively more basic. At the center stands a formal theory, a theory of truth, which imposes a complex structure on sentences containing the primitive notions of truth and satisfaction. These notions are given application by the form of the theory and the nature of the evidence. The result is a partially interpreted theory. The advantage of the method lies not in its free-style appeal to the notion of evidential support but in the idea of a powerful theory interpreted at the most advantageous point. This allows us to reconcile the need for a semantically articulated structure with a theory

testable only at the sentential level. The more subtle gain is that very thin evidence in support of each of a potential infinity of points can yield rich results, even with respect to the points. By knowing only the conditions under which speakers hold sentences true, we can come out, given a satisfactory theory, with an interpretation of each sentence. It remains to make good on this last claim. The theory itself at best gives truth conditions. What we need to show is that if such a theory satisfies the constraints we have specified, it may be used to yield interpretations.

3. *If we know that a theory of truth satisfies the formal and empirical criteria described, can we interpret utterances of the language for which it is a theory?*

A theory of truth entails a T-sentence for each sentence of the object language, and a T-sentence gives truth conditions. It is tempting, therefore, simply to say that a T-sentence "gives the meaning" of a sentence. Not, of course, by naming or describing an entity that is a meaning, but simply by saying under what conditions an utterance of the sentence is true. [15]

But on reflection it is clear that a T-sentence does not give the meaning of the sentence it concerns: the T-sentence does fix the truth value relative to certain conditions, but it does not say the object language sentence is true *because* the conditions hold. Yet if truth value were all that mattered, the T-sentence for "Snow is white" could as well say that it is true if and only if grass is green or $2+2=4$ as say that it is true if and only if snow is white. We may be confident, perhaps, that no satisfactory theory of truth will produce such anomalous T-sentences, but this confidence does not license us to make more of T-sentences.

A move that might seem helpful is to claim that it is not the T-sentence alone, but the canonical proof of a T-sentence, that permits us to interpret the alien sentence. A canonical proof, given a theory of truth, is easy to construct, moving as it does through a string of biconditionals, and requiring for uniqueness only occasional decisions to govern left and right precedence. The proof does reflect the logical form the theory assigns to the sentence, and so might be thought to reveal something about meaning. But in fact we would know no more

than before about how to interpret if all we knew was that a certain sequence of sentences was the proof, from some true theory, of a particular T-sentence.

A final suggestion along these lines would be to say that we can interpret a particular sentence provided we know a correct theory of truth that deals with the language of that sentence. For then we know not only the T-sentence for the sentence to be interpreted, but we also know the T-sentences for all other sentences; and of course, all the proofs. Then we would see the place of the sentence in the language as a whole, we would know the role of each significant part of the sentence, and we would know a great deal about the logical connections between this sentence and others.

The suggestion fails. For how can it help in interpreting a single sentence to know the truth conditions of others? Of course, if we learn that a speaker also holds other sentences to be true or false, that may be a help. Indeed, enough more such information, and interpretation certainly will be possible. But enough more such information, and the theory isn't needed, for all that went into the theory was information about sentences held true under various circumstances. The point of the theory is to digest this fund of evidence and to deliver it in a form useful for the interpretation of isolated utterances. We must conclude, I think, that relativizing a T-sentence to a proof or theory is no help: if the theory does what it is designed to do, T-sentences taken alone must provide all we need for interpretation.

If we knew that a T-sentence satisfied Tarski's Convention T, we would know that it was true, and we could use it to interpret a sentence because we would know that the right branch of the biconditional translated the sentence to be interpreted. Our present trouble springs from the fact that in radical interpretation we cannot *assume* that a T-sentence satisfies the translation criterion. What we have been overlooking, however, is that we have supplied an alternative criterion: this criterion is that the totality of T-sentences should (in the sense described above) optimally fit evidence about sentences held true by native speakers. The present idea is that what Tarski assumed outright for each T-sentence can be indirectly elicited by a holistic constraint. If that constraint is adequate, each T-sentence will in fact yield interpretations.

A T-sentence of an empirical theory of truth can be used to interpret a sentence, then, provided we also know that the T-sentence is

entailed by some true theory that meets the formal and empirical criteria. It is not necessary to know what the theory is in a particular case, only that it is such a theory. For if the constraints are adequate, the range of acceptable theories will be such that any of them yields some correct interpretation for each potential utterance. To see how it might work, accept for a moment the absurd hypothesis that the constraints narrow down the possible theories to one, and this one implies the T-sentence (T) discussed previously. Then we are justified in using this T-sentence to interpret Kurt's utterance of "Es regnet" as his saying that it is raining. It is not likely, given the flexible nature of the constraints, that all acceptable theories will be identical. When all the evidence is in, there will remain, as Quine has emphasized, the trade-offs between the beliefs we attribute to a speaker and the interpretations we give his words. But the resulting indeterminacy cannot be so great but that any theory that passes the tests will serve to yield interpretations.

FOOTNOTES

1. Here and throughout this paper my debt to the work of W. V. O. Quine will be obvious. The term "radical interpretation" is meant to suggest a strong kinship with Quine's "radical translation" (*Word and Object*, Cambridge, Mass. 1960). Kinship is not identity, however, and "interpretation" in place of "translation" marks one of the differences: a greater emphasis on the explicitly semantical.
2. At one time I was convinced that unless such a finitely characterized theory could be provided for a language, the language could not be learned by a creature with finite powers. (See Donald Davidson, "Theories of Meaning and Learnable Languages", in *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*, Amsterdam 1966, pp. 383-394.) This still seems to me likely to be right, but Georg Kreisel has made me realize that the point is not obvious.
3. The idea of a translation manual with appropriate empirical constraints as a device for studying problems in the philosophy of language is, of course, Quine's. This idea inspired much of my thinking on the present subject, and my proposal is in important respects very close to Quine's. Since Quine may not have intended to answer the questions I have set, the claim that the method of translation is not adequate as a solution to the problem of radical interpretation may not be a criticism of any doctrine of Quine's.
4. Alfred Tarski, "The Concept of Truth in Formalized Languages", in *Logic, Semantics, Metamathematics*, Oxford 1956.
5. For a discussion of how a theory of truth can handle demonstratives, and how Convention T must be modified, see Scott Weinstein, "Truth and Demonstratives", *Noûs* (forthcoming, 1974).

6. See John Wallace, "On the Frame of Reference", *Synthese*, Vol. 22 (1970). pp. 61-94.
7. Tyler Burge, "Reference and Proper Names", *Journal of Philosophy*, Vol. 70 (1973), pp. 425-439.
8. Gilbert Harman, "Moral Relativism Defended", forthcoming.
9. John Wallace, "Positive, Comparative, Superlative", *Journal of Philosophy*, Vol. 69 (1972), pp. 773-782.
10. Donald Davidson, "On Saying That", *Synthese*, vol. 19 (1968), pp. 130-146.
11. Donald Davidson, "Causal Relations", *Journal of Philosophy*, vol. 64 (1967), pp. 691-703.
12. Donald Davidson, "Quotation", unpublished. Forthcoming as a chapter of *The Structure of Truth*, Oxford.
13. Michael Dummett, *Frege*, London 1973.
14. Readers who appreciate the extent to which this account parallels Quine's account of radical translation in Chapter 2 of *Word and Object* will also notice the differences: the semantic constraint in my method forces quantificational structure on the language to be interpreted, which probably does not leave room for indeterminacy of logical form; the notion of stimulus meaning plays no role in my method, but its place is taken by reference to the objective features of the world which alter in conjunction with changes in attitude towards the truth of sentences; the principle of charity, which Quine emphasizes only in connection with the identification of the (pure) sentential connectives, I apply across the board.
15. This idea, and others rejected here, will be found in various articles of mine: see "Truth and Meaning", *Synthese*, vol. 17 (1967), pp. 304-323, "Semantics for Natural Languages", in *Linguaggi nella Società e nella Tecnica*, Milan 1970, pp. 177-188, and "True to the Facts", *Journal of Philosophy*, vol. 66 (1969), pp. 748-764.

Donald Davidson
The Rockefeller University
and Princeton University
New York 10021