Univ. of Peshawar (Pakistan)

Language and Translatability: Tarski versus Davidson

Mohammad Ashraf Adeel

THE CONFLICT of opinion between Tarski and Davidson regarding the semantics of natural languages is sharp indeed.[1] It is sharp because Tarski despairs of applying the methods of formal semantics to natural languages while Davidson bases his entire program in semantics on such a possibility. What is puzzling, however, about Davidson's attitude in this regard is that he visualizes a marriage between his program in semantics and his refutation of conceptual relativism. It is tempting enough to apply some of the fine-grained formal distinctions to the apparently hapless natural languages; but the Davidsonian package purports to go beyond that. It aims ambitiously at retrieving our common-sense realist intuitions by destroying conceptual relativism root and branch. A conceptual relativist is someone who argues that people can have radically different world views, world views that cannot so much as be stated in each other's terms. In other words, purported truths of a given world view or theory are not accessible to the believers of a radically different world view. The instrument of Davidson's onslaught on conceptual relativism is an argument which concludes with his criterion of languagehood. In what follows, I will sketch an argument to show that this criterion of languagehood cannot be accepted simultaneously with a rejection of Tarski's deep-seated pessimism regarding the applicability of formal methods to the semantics of natural languages. In a nutshell, I will argue that the Davidsonian package (Davidson's program in semantics plus his refutation of conceptual relativism) is afflicted by a contradiction. By way of introducing the problem let me state, very briefly, what I believe is contradictory in Davidson's package.

Tarski claims that natural languages are universal. He means two things by this universality thesis: a natural language contains its own metalanguage, and any other language can be translated into it. Formalized languages, according to Tarski, lack both these properties and, hence, they are not universal. Now if Davidson wants to apply the methods of formal semantics to natural languages he must deny Tarski's claim about their universality. In other words, he must divide a natural language into two languages, the object language and the metalanguage. But denying Tarski's universality thesis carries with it the implication that not everything can be said or translated into the language under study, i.e., the object

[1]All references to Donald Davidson's papers are from his *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984). References to Alfred Tarski's work are confined to "The Concept of Truth in Formalized Languages" from his *Logic, Semantics, and Metamathematics* (Oxford: Clarendon Press, 1956), hereafter "Tarski."

language. In other words, when one denies universality of a language, one thereby grants that there is going to be some failure of translation between the language in question and some other languages.

Now Davidson does declare Tarski's universality thesis as suspect in order to motivate his semantic program. In his refutation of conceptual relativism, however, he also declares translatability to be a criterion of languagehood. Davidson argues that there can be no failure of translation, either partial or complete, between languages. This is where the contradiction arises in Davidson's package. In order to motivate his semantic program he must admit failure of translation between languages. But in order to refute conceptual relativism he must deny such failure. The rest of my paper is an elaboration of this claim.

I

As noted above Tarski thinks that natural languages are universal while scientific, or formal ones, are not. He gives us the following necessary condition for the universality of a language in his "The Concept of Truth in Formalized Languages":

A) If L is a universal language then L must contain, in addition to its sentences and expressions, the names of these sentences and expressions, sentences containing these names, and semantic expressions such as "is a true sentence," "names," "denotes," etc. (Cf. Tarski, 164)

Tarski also provides us with a sufficient condition for the universality of a language:

B) If for every meaningful language X, X is translatable into L, then L is universal. (Ibid.)

Any L that satisfies the conditions (A) and (B) is universal and therefore, a natural language for Tarski.

Now Tarski denies that formalized languages are universal. So the minimal distinction between formalized and natural languages for him is that formalized ones do not satisfy conditions (A) and (B). That is to say, if K is a formalized language, then K does not contain any apparatus for referring to its own expressions and sentences, nor does it contain its own truth-predicate or satisfaction relation, etc. These resources belong to MK, the metalanguage for K. Further, not every meaningful language can be translated into K. This is not just a practical limitation of K arising from the meagerness of its resources. For resources of a language can, in principle, always be augmented either by introducing completely new terms into it or by appropriating terms from another language. K is, *in principle,* not capable of translating every meaningful language into it. Specifically MK or any other higher-order language cannot be translated into K on pain of paradox. In a nutshell, then, the distinction between natural and formalized languages is that natural languages contain their own metalanguage while formalized languages do not. The metalanguage for K, for example, is a separate language MK which, in homophonic cases, includes K but has some further

resources which K does not have. It is these peculiarly metalinguistic resources of MK that cannot be translated into K on pain of paradox.

The universality of natural languages makes them inconsistent. The reason is easy to see. Given the resources that L must have if it is universal, L cannot stay consistent. That is, its quantifiers have in their range not only its sentences and other expressions but also sentences containing the names of its sentences and expressions. Further, L contains semantical expressions such as "is a true sentence," etc. These resources in L give rise to semantic paradoxes such as the liar, Grelling's heterological paradox, and others. So Tarski's necessary condition for the universality of a language implies that such a language is inconsistent. In an inconsistent language, however, we can prove any given statement to be both true and false. This is one of the central reasons for Tarski's deep-seated pessimism as far as the applicability of formal methods to the semantics of natural language is concerned.[2]

II

Davidson does not share Tarski's pessimism in this regard. He believes that we can apply formal methods to the semantics of natural languages.

In particular, he suggests that we split a natural language, such as English, into two languages, English_0 which is the object language and English_1 which is the metalanguage for English_0. Moreover, English_1 contains English_0 (in the homophonic case) as well as devices for referring to the English_0 sentences and expressions, plus a satisfaction relation. The crux of the Davidsonian program in semantics for natural languages implies that we recursively characterize (rather than define explicitly) "true-in-English_0" in the metalanguage English_1 such that for every sentence of English_0 there follows, from our recursive theory of truth, a sentence of the form:

D) S is true-in-English_0 iff p,

where S is replaced by a structural description of the sentence in question and p by the sentence itself, since the metalanguage includes the object language. Now the substitution instances of (D) such as " 'snow is white' is true-in-English_0 iff snow is white" are recognized by us to be trivially true. But the totality of these sentences (Tarski's T-sentences) of English_1 uniquely determines the extension of the truth-predicate for English_0. So a recursive truth theory, which satisfies appropriate formal (and empirical) constraints, delivers for us the truth conditions of every sentence in English_0, which is the same as delivering meanings of these sentences.[3]

Our concern in this paper, however, is not the details of Davidson's theory of meaning. For our purposes, it suffices to note that Davidson splits English into two languages and the language interpreted is distinct from the language in which the interpretation is given. It is, however, exactly the formal procedures, Tarski

[2]The other reasons for Tarski's pessimism in this regard are that natural languages are vague and ambiguous, and in general, inductively non-specifiable. (Cf. Tarski, 267)

[3]Donald Davidson also requires the truth theory to be finitely axiomatizable as well as empirically verifiable in a holistic sense.

thought, that cannot be applied to natural languages if their universality is to be maintained. We are at the heart of the conflict between Tarski and Davidson now. Davidson has a program for interpreting natural languages and the program is based on the application of formal methods to these languages. Tarski rejects such an application on the grounds that natural languages are universal and hence inconsistent. Davidson's way out is to reject the claim that natural languages are universal. And that is exactly what he does in order to motivate his program. He says: ". . . [i]t seems to me that this claim, now that we know such universality leads to paradox, is suspect."[4]

To reject the universality claim is obviously to reject Tarski's (A) and (B) above. To repeat rejecting the universality claim for natural languages translates into the idea that not everything that can be said meaningfully is translatable into a natural language. When we split, for example, English into $English_0$ and $English_1$, the idea is to avoid semantical paradoxes by imposing restrictions on what can be said or translated in the object language $English_0$. In particular, $English_1$ cannot, all of it, be translated into $English_0$ without collapsing the distinction between the two languages. We must grant, therefore, that not every meaningful (piece of) language can be translated into $English_0$.[5]

III

Davidson, however, wants to refute conceptual relativism as well. For this purpose he argues for the following: The criterion of languagehood is translatability. The question is how does this criterion of languagehood refute conceptual relativism? We know that a conceptual relativist argues for a plurality of conceptual schemes. Davidson rightly contends that the best construal of such a position is to identify conceptual schemes with languages. On such a construal, sets of intertranslatable languages are taken to embody the same conceptual scheme, while failure of intertranslatability is taken to signal a difference in conceptual scheme. After having laid down this criterion of difference in conceptual schemes, Davidson embarks on the task of showing that failure of translation between languages, either partial or complete, is impossible. Such is his refutation of conceptual relativism. But let us look into the details of his argument for his criterion of languagehood.

Here is how Davidson characterizes complete or partial failure of translation between languages: "There would be complete failure if no significant range of sentences in one language could be translated into the other; there would be partial failure if some range could be translated and some range could not. . . ."[6] After having given this characterization, Davidson proceeds to raise objections against the idea that there could be any complete or partial failure of translation between languages.

Davidson's first general point in support of his criterion of languagehood is that "nothing . . . could count as evidence that some form of activity could not be

[4]Davidson, "Truth and Meaning," p. 29.

[5]Davidson seems to acknowledge this when he says: ". . . there may in the nature of the case always be something we grasp in understanding the language of another (the concept of truth) that we cannot communicate to him" ("Truth and Meaning," p. 29).

[6]"On the Very Idea of a Conceptual Scheme," p. 185.

interpreted in our language that was not at the same time evidence that that form of activity was not speech behaviour."[6a] In other words, if some form of activity is speech behavior, then it must be understandable in principle; we should be able to translate it into our own language. Otherwise there would not be any grounds for us to call that activity speech behavior. Furthermore, Davidson rejects the claim that "[b]y imagining a sequence of languages, each close enough to the one before to be acceptably translated into it, we can imagine a language so different from English as to resist totally translation into it."[7] This claim is unacceptable because we could never recognize any member of such a sequence as translatable into the one before if it was not translatable into our home language, English. That is, if language B is translatable into English and language C is translatable into B, then C must be translatable into English. To imagine even a partial failure of translation from C into English would cast doubt on our ability to recognize that C can be translated into B.

This general line of argument is supplemented in Davidson's paper by his attack on the distinction between conceptual scheme and empirical content, the so-called scheme-content dualism, which many philosophers employ in order to strengthen the idea that there are different conceptual schemes. This dualism is the first cousin of analytic-synthetic dualism which, as we know, divides up the sentences of our language into two groups, those true (or false) by virtue of their meaning alone and those true (or false) by virtue of both their meaning and empirical content. Giving up the analytic-synthetic dualism, however, does not mean that we give up the notion of empirical content. Davidson explains this in the following words:

> . . . [W]e can hold, if we want, that *all* sentences have empirical content. Empirical content is in turn explained by reference to the facts, the world, experience, sensation, the totality of sensory stimuli, or something similar. . . . [I]t is possible . . . to give up meanings and analyticity while retaining the idea of language as embodying a conceptual scheme. Thus in place of the dualism of analytic-synthetic we get the dualism of conceptual scheme and empirical content.[8]

Davidson calls this dualism the third dogma of empiricism and chases it through the writings of such authors as Whorf, Kuhn, Feyerabend, and Quine. His contention is that the scheme-content dualism is as dubious and untenable as the analytic-synthetic dualism. The scheme-content dualism presupposes "something neutral and common that lies outside all schemes." This neutral and common stuff is somehow *organized* by alternative schemes. The problem, however, is that "[t]his common something cannot . . . be the *subject matter* of contrasting languages, or translation would be possible."[9] So those philosophers who claim that the same world can be approached by different scientists (observers) with incommensurable concepts are mistaken in a very basic sense. If the "common something," the world, is the subject matter of contrasting languages (schemes), then translation between such languages cannot fail. But if we assume that there

[6a]Ibid.
[7]Ibid., p. 186.
[8]Ibid., p. 189.
[9]Ibid., p. 190.

is a failure of translatability between such languages, then we will have no grounds for claiming that these languages have the same world as their *subject matter*. Another idea is that if we assume our conceptual resources as fixed and also assume that we generate new schemes out of the old ones by some systematic redistribution of truth values over the sentences of our language, then we can end up talking about different possible universes. The problem with this idea is the same. How could we know that the new scheme is new if there were no translation between the new scheme and the old one?

Davidson goes on to point out that those committed to the scheme-content dualism work basically with two kinds of metaphors when it comes to describing the relation between the conceptual schemes and the world, reality, or sensory promptings, etc. According to the first metaphor, the conceptual schemes *organize*, *systematize*, or *divide up* the *uninterpreted content*, *the world*, *the stream of experience*, or *the surface irritations*. According to the second metaphor, a conceptual scheme *fits* our *sensory promptings*, or, *faces* the tribunal of experience, etc. The first metaphor focuses on the referential apparatus of language, its singular terms, quantifiers, predicates, etc. It is this apparatus which is thought to *organize* or *divide up* the uninterpreted content. The second metaphor focuses on whole sentences. We use sentences to predict things or cope with them. Also it is sentences that fit sensory promptings or face the tribunal of experience collectively.

Davidson argues that both kinds of metaphors fail to make an acceptable case for conceptual relativism. Regarding the first metaphor, he says that "[w]e cannot attach a clear meaning to the notion of organizing a single object (the world, nature, etc.) unless that object is understood to contain or consist in other objects."[10] If, however, we take the metaphor as suggesting that different languages organize the same items or entities in the world, then we must grant that such languages have a shared ontology. But, then, such languages cannot fail to be intertranslatable. So the philosophers who employ this kind of metaphor in support of conceptual relativism fail to show how languages can fail to be intertranslatable.

At this point in his argument Davidson talks about what he calls "local" cases of failure of translation between languages and declares them to be intelligible:

> A language may contain simple predicates whose extensions are matched by no simple predicates, or even by any predicate at all, in some other language. What enables us to make this point in particular cases is an ontology common to the two languages, with concepts that individuate the same objects. We can be clear about breakdowns in translation when they are local enough, for a background of generally successful translation provides what is needed to make the failures intelligible.[11]

Apparently the distinction between local failure of translation between languages and the partial failure, which Davidson thinks is impossible, is drawn in the following terms: If English contains some simple predicates which are not matched by any counterparts in Hindko then the failure of translation from English into Hindko is a case of local failure. Partial failure of translation,

[10]Ibid., p. 192.
[11]Ibid.

however, between Hindko and English occurs if some range of English sentences can be translated into Hindko while some range of English sentences cannot be so translated. This way of characterizing the distinction suggests that every case of a local failure of translation is a case of partial failure, though not vice versa necessarily. The reason is easy to see. If there are cases of local failure of translation between Hindko and English, then there are some predicates in English that do not have any counterparts in Hindko. For the sake of example, let us assume that Hindko does not have the predicate ". . . is grue." Now the problem is that there are countless sentences of English in which the predicate ". . . is grue" can occur potentially. If Hindko cannot translate ". . . is grue," then, presumably, it cannot translate a whole range of English sentences in which the said predicate occurs. Therefore, if there is a case of local failure of translation between Hindko and English, then there is partial failure of translation between them. So local failure of translation between languages turns out to be sufficient for a partial failure, although it may not be necessary. But Davidson denies that there can be a partial failure of translation between languages. Hence he has a basic tension even within his argument for the criterion of languagehood.

Davidson apparently introduces the notion of "local" failure of translation in order to save our ordinary workaday intuitions regarding languages. Ordinarily we do think that not all languages have all the predicates which, let us say, English has. But Davidson's notion of "local" failure turns out to be an ad hoc device, since it runs counter to his criterion of languagehood which disallows either partial or complete failure of translation between languages.

The second metaphor which, Davidson says, some philosophers employ in order to bolster relativism is that of a scheme (or language) *fitting* our sensory promptings or *facing* the tribunal of experience, etc. This just means that the conceptual scheme, a broad theory of the world, is borne out by evidence. Commenting on this Davidson says: "In the common course of affairs, a theory may be borne out by the available evidence and yet be false. But what is in view here is not just actually available evidence; it is the totality of possible sensory evidence past, present, and future."[12] A theory which is borne out by all possible evidence, however, is a theory simply true. To talk about a fit between the theory and evidence, in Davidson's words, "adds nothing intelligible to the simple concept of being true."[13] Therefore, for any two theories to fit the totality of all possible evidence just means that both these theories are true. If, however, the metaphor has to support relativism, then these theories must not be intertranslatable. But if we take them to be incommensurable, we commit ourselves to the claim that we can understand what it means for a theory to be true without being able to translate it into our own. This, however, is an untenable claim. But if we allow the theories to be intertranslatable, then the metaphor of a fit between theories (schemes) and evidence obviously fails to supply us with a criterion of languagehood that would entail failure of translatability.

There are other aspects to Davidson's argument in "On the Very Idea of a Conceptual Scheme," but we need not detail them here. Suffice it to say that he concludes with the claim that the criterion of languagehood is translatability and

[12]Ibid., p. 193.
[13]Ibid., pp. 193–94.

either complete or partial failure of translation between languages is impossible. Davidson, of course, concedes that there is such a thing as *local* failure of translation, but as I pointed out, this concession he cannot make consistently with his denial that there can be partial failure of translation between languages.

IV

As noted above, Davidson rejects the Tarskian thesis that natural languages are universal in order to motivate his semantic program. This means that he divides English, for example, into two languages—the object language English_0 and the metalanguage English_1. English_1 includes English_0 in it but exceeds English_0 in having the truth-predicate or satisfaction relation for English_0 and the apparatus for referring to the expressions and sentences of English_0. While the introduction of these levels helps us avoid semantical paradoxes, it does so by barring translation of certain parts of the metalanguage, such as the satisfaction relation, into the object language. As W. V. Quine puts it, though in a slightly different context than ours, the satisfaction relation becomes an "untranslatable foreign language" as far as the object language is concerned.[14]

Let me try to put the same point differently. Tarski's necessary condition for the universality of natural languages is that they contain their own metalanguage. This, of course, makes natural languages inconsistent. His sufficient condition for the universality of these languages is that every meaningful (piece of) language can be translated into them. Now anybody who wants to introduce a Tarski-style formal distinction between object language and metalanguage into a natural language must deny Tarski's necessary condition for the universality of that natural language. But this also means that not every meaningful (piece of) language can be translated into the resulting language(s). Therefore, anyone who is anxious to apply Tarskian formal methods to the semantics of a natural language must perforce grant that there is some failure of translation between the language he is studying and some other languages. It is, however, only a metalanguage, parts of which can be plausibly maintained to be untranslatable into its object language. Any other languages which would be plausible candidates in this regard are hard to find. But my basic point is that if Davidson wants to apply formal methods to the study of natural languages, he cannot avoid allowing for some failure of translatability.

This is exactly what Davidson's criterion of languagehood is meant to deny, however. According to this criterion there is no failure of translation, either partial or complete, between languages. Hence Davidson's criterion of languagehood clashes with the basic motivation (his denial of Tarski's universality thesis) of his semantic program, and he is involved in a contradiction.[15]

[14]Cf. W. V. Quine, *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970), p. 45.

[15]I am indebted to Professors Irving M. Copi and Larry Laudan as well as to Dr. George Rudebusch and Russell Alfonso for helpful conversations on the matters discussed in this paper.