Radically interpreting. On Davidson's theory of meaning*

Interpretando radicalmente. Acerca de la teoría del significado de Davidson

Interpretar radicalmente. Sobre a teoria do significado Davidson

Fecha de entrega: 13 de febrero de 2016
Fecha de evaluación: 18 de junio de 2016
Fecha de aprobación: 30 de junio de 2016

Edgar Eslava**

Abstract

The aim of this paper is to show the central issues of Davidson’s radical interpretation theory by making use of it in a concrete example in order to see how the interpretation proposal responds to the task it was designed for. In doing this, I’ll try to show radical interpretation as an independent theory and not just as a response to other theories of meaning, as it is usually seen.

Keywords: Radical interpretation, theory of meaning, truth value.

* Artículo resultado de investigación, como parte del proyecto Educación, Filosofía e Identidad, liderado por el autor.DOI: http://dx.doi.org/10.15332/s0120-8462.2016.0115.08

Resumen

El objetivo de este texto es presentar los aspectos centrales de la teoría de la interpretación radical de Davidson, haciendo uso de ella en un ejemplo concreto a fin de ver cómo esta propuesta interpretativa responde a la tarea para la cual ha sido formulada. Intentaré mostrar a la interpretación radical como una teoría independiente, y no solo como una respuesta a otras teorías del significado, como usualmente se la presenta.

Palabras clave: interpretación radical, teoría del significado, valor de verdad.

Resumo

O objetivo deste artigo é mostrar as questões centrais da teoria interpretação radical de Davidson, fazendo uso do mesmo em um exemplo concreto, a fim de ver como a proposta interpretação responde à tarefa ele foi projetado para. Ao fazer isso, eu vou tentar mostrar a interpretação radical como uma teoria independente e não apenas como uma resposta a outras teorias de significado, como é geralmente visto.

Palavras-chave: Interpretação radical, teoria do significado, valor de verdade.

Introduction

“I did not like the movie”. That was what my wife said after the show. At first I agreed by saying, “Neither did I”. But after thinking about it for a moment I realized that I wasn’t sure our judgment was really the same. I knew exactly what I meant, but there was no reason to assume that that was exactly what she meant. Was she talking about the performance of the protagonists, or the way the story developed, or the emotions the movie evoked in her? More than that, how could I be sure that I had gotten the exact meaning of her words? It is precisely to answer questions
like the latter one that the theory of radical interpretation has been advanced. The present text uses Davidson’s radical interpretation theory to undertake the task it was designed for, that is, making sense of other’s communicative efforts. Contrary to tradition, I will address radical interpretation not as an alternative to theories of meaning such as Quine’s radical translation theory, as for example the approaches found in Searle (1987), Malpas (1992), and Quine (1991), but as an independent and solid theory in itself.

**Making sense of a speaker**

From Davidson’s point of view, the central problem for a theory of meaning is to provide us with the tools for being able to understand the linguistic utterances of any speaker, and the aim of his philosophy of language is to the development of such a theory. This compromise requires, from his perspective, that there must be established both the formal elements and the set of facts that count as evidence for the theory being used. The way in which Davidson performs this double task is by supporting the system of inferences of his interpretative model on the formalism of a Tarski-like theory of truth, and using as evidence his own defined elements of linguistic behavior. But to see how those bases are built it is necessary to go into some detail.

“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” (Davidson, 1974, p. 142). This is the way Davidson expresses the idea that utterances, the starting point of communication, are charged with meanings and beliefs, and that if we are trying to understand somebody’s utterances, the relationship between meanings and beliefs needs to be clarified.

But, from Davidson’s perspective there’s a huge problem with such an attempt of clarification, because it is impossible to infer the beliefs of the speaker only by his utterances without knowing the meaning of his words. And at the same time it is also impossible to get the exact meaning of the words based exclusively on the sentences themselves, without the knowledge of his beliefs. Here, Davidson sees a parallel with Decision Theory; he sees this as a situation of “indifferent choice”. In such a case, a subject facing two alternative options shows no preference for either of them but forced by his needs finally chooses one without making explicit his reasons. Assuming that the choice is determined by the values and beliefs of the subject, it is necessary
to know at least one of them if we are trying to understand the way the choice was made. Davidson (1974) summarizes the situation this way:

Choices between gambles are the result of two psychological factors, the relative values the chooser places on the outcomes, and the probability he assigns to those outcomes, conditional on his choice. Given the agent’s beliefs (his subjective probabilities) it’s easy to compute his relative values from his choices; given his values, we can infer his beliefs. But given only his choices, how can we work out both his beliefs and his values? (p. 145).

The exit to the dilemma, in Decision Theory, is to use the subject’s previous choices as the base for the explanations of his preferences (choices). Then, using the new data, attribute subjective beliefs and values in order to get the “structure” of his behavior, structure that is going to be used as a starting point for the interpretation of the choices, past and future.

The result of the parallel with Decision Theory is that the theory of understanding utterances which we are looking for must be based on the utterances (sentences) themselves. And the attribution of beliefs and meanings, both at the same time, is part of the process of constructing the theory. “I conclude,” says Davidson, “that in interpreting utterances –in radical interpretation– we must somehow deliver simultaneously a theory of belief and a theory of meaning” (Davidson, 1974, p. 144). But beyond the parallel, and due to the fact that communication is composed of agents, the speaker and the listener, the attribution of beliefs and meanings is bidirectional. The speaker ascribes them to the listener when trying to be understood, in the same way the listener does to the speaker while attempting to understand the meaning of his speech. The process of attribution of meaning (and beliefs, as we just saw) on each occasion of use is an interpretative one, and the effectiveness of the communication between agents, rests on the ability of the speaker and the listener, as interpreters, to do their job “as good as it gets”.

The interpretation of utterances must have one special characteristic. It must offer a way to understand an infinite number of sentences, all the possible sentences to be uttered at any moment, based only on a finite number of them: those the agents know at a particular moment. In other words, the interpretative system to be built has to take single elements of the system itself and let the interpreter be able to use them
to make inferences over an infinitude of new ones. Formally speaking, the system (theory) must be recursive.

**Interpreting utterances**

Knowing the recursive character of Tarski’s theory of truth, Davidson finds this theory a perfect candidate to be the ground for the development of his own theory of interpretation. Let’s then briefly see Tarski’s approach. Tarski’s main interest is to develop a theory of truth that fits our intuition about what truth is, this is to say the Aristotelian conception of truth, which states that: “To say of what is that it is not, or of what is not that it is, is false while to say of what is that it is, or of what is not that is not, is true”. In doing this, Tarski suggests that the general formula for the truth of a sentence should be “X is true, if and only if p”, where p is a sentence of the language and X is the name (naming) of p. Sentences of that sort are so called T-sentences. Although T-sentences can be used, and are used, as reference for the truth value of other sentences of the language, it should be noted that

> Neither the expression (T) [the T-sentence] itself (which is not a sentence, but only the schema of a sentence) nor any particular instance of the form (T) can be regarded as a definition of truth. We can only say that every equivalence of the form (T) obtained by replacing ‘p’ by a particular sentence, and ‘X’ by a name of this sentence, may be considered a partial definition of truth, which explains wherein the truth of this individual sentence consists. The general definition has to be, in a certain sense, a logical conjunction of all this partial definitions (Tarski, 1944, p. 344).

This means that all that has been defined is the class of equivalence of the T-sentences of the elements of the theory; the characterization of the material adequacy of the sentences of the language.

Tarski proposed to name his conception of truth “semantical” because of his interest in bringing out a theory that deals with the relations of expressions of a language and the objects referred to by those expressions, making the theory dependent on the specification of the structure of the language over which truth is going to be defined.
As is easy to see, now there are two languages drawn into the definition of truth: One which is “talked about” and over which the whole discussion is made, the object language, and one in which we “talk about” and use to develop the definition of truth for the object language, the metalanguage. One of the features of the metalanguage that is relevant for the construction of a theory of meaning like the one Davidson is working on is that it contains the object language as a part, which allows us to name every sentence of the object language. In terms of speakers and listeners this means that the only thing we need to build a theory of interpretation of utterances is the original set of utterances that will be subsumed by the metalanguage of the interpreter. A final condition of Tarski’s theory is that the truth of a particular sentence is defined recursively from the former assignation of truth-values to prior sentences. Just what Davidson was looking for, since for him, “a theory of meaning for a language L shows “how the meaning of sentences depends upon the meanings of words’ if it contains a (recursive) definition of truth-in-L” (Davidson, 1967, p. 310). Further:

To give a recursive theory of truth for a language is to show that the syntax of the language is formalizable in at least the sense that every true expression may be analyzed as formed from elements (the ‘vocabulary’), a finite supply of which suffice for the language by the application of rules, a finite number of which suffice for the language (Davidson, 1970, p. 57).

But there’s a dramatic difference between Davidson’s theory of meaning and Tarski’s theory of truth. To obtain a definition of truth, Tarski needs to accept that the meanings of the sentences are shown clearly by the very utterance act, without further considerations. On the other hand, Davidson is trying to find out meanings and, obviously, he can’t have them as axioms of his theory.

In Tarski’s work, T-sentences are taken to be true because the right branch of the bi-conditional 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 recognized without pre-empting the point of radical interpretation: in empirical applications, we must abandon the assumption. What I propose is to reverse the direction of the 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 (Davidson, 1973, p. 134).
So, Davidson uses Tarski’s theory as a formal ground of the theory of interpretation accepting the help provided by its recursive character, but without the syntactical test of the T-sentences used by Tarski and without the search for an explicit definition of truth. Those exceptions are made because Davidson assumes that the task of understanding meanings of utterances can only be completed if sustained by a solid notion of truth.

But how is it that Davidson is so sure that truth can be established indisputably, “beyond any reasonable doubt?” His answer is clear and direct. He assumes the truth of the utterances of the speaker as a non-questionable fact “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” (Davidson, 1972, p. 152).

This Charitable Principle, or principle of charity, is the very central axiom of Davidson’s system and encloses important elements of the interpretation attempt. It makes the truth of an utterance relative to “observed circumstances”. Those circumstances are in turn dependent on the moment and context of the performance of the utterance. Therefore, the truth or falsity value of a sentence is something to be judged only by somebody who knows the actual situation under which the utterance was done. But because the interpretation is a two-way process, it also implies that the speaker must know that the conditions under which his utterances are made are such that the listener has a good chance to make sense of them. In other words, the principle of charity defines the conditions under which the linguistic behavior of the agents is necessarily rational. And with this we have completed the exposition of the formal elements of Davidson’s theory of meaning.

**Formalizing interpretation**

Now is the moment to ask how the formalism just introduced helps us in concrete cases of understanding meanings of utterances. Let’s go back to the movies. During the show all the information offered by the movie is incorporated into the interpretative structure of my wife as it is into every other audience member’s structure as well as mine. Becoming part of such a structure means that the feelings the director wants to transmit and the images brought out during the film will mix with all our personal feelings, memories, beliefs and expectations, making our interpretation system more
complex than it was originally before the show. At the end of the show, when the utterance “I did not like the movie” was made public, my wife communicated the final result of the application of her new interpretative system on the movie, and she judged the show based on both her previous and actual beliefs. What she means by her utterance is intended to be understood by me, the agent whose rationality is taken for granted in the very moment she directed her comment to me. The fact that she takes me as an acceptable listener is based on the application of the charity principle to one specific hearer at a particular moment.

When I try to understand the meaning of her sentence my starting point is, again, the charity principle, but knowingly applied to the speaker. Surely, “I did not like the movie” is only true if and only if she did not like the movie, so I accept the truth of her utterance as a fact. Beyond the fact of mere acceptances, this new information, the utterance, is now part of my system of interpretation, the one that I use to understand the utterance itself. So, one element that wasn’t part of the system until offered as a question to be solved (intended to be understood), becomes part of the set of tools used to solve the question it raises.

But, so far, the problem hasn’t been resolved. I still don’t know the meaning of her sentence. More than that, I don’t know if I have all the required information to do so. As I said before, formalism is just one of the two elements required for the theory. The definition of the acceptable evidence is the other one.

Based on the characteristics of the system used by Davidson, it seems natural to think that the empirical evidence for the theory is the expression uttered by the speaker. In fact, it is over those expressions that the formal analysis has been proposed. But if the expressions (sentences) are taken as evidence for the theory there’s at least one problem that is revealed immediately: What are the sentences evidence of? Of course the expressions are not evidence of any notion of truth because the assignation of truth is something that belongs to the formal kernel of the theory and not something to be determined by reference to any specific fact. Neither are sentences evidence of meaning or beliefs, in which case, the whole project would become superfluous and the final theory simply redundant. It seems that they are only evidence of the existence of a linguistic agent. But such a hollow conquest isn’t really helpful in order to complete the task assigned to the empirical evidence of the theory, which is used as an instance of validation of meaning. In Davidson’s (1972) terms:
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 (p. 128).

The question then is open. And, again, Davidson’s answer is sharp and direct. “Instead of utterances of expressions, I want to consider a certain attitude toward expressions... The attitude of holding true, relativized to time” (Davidson, 1974, p. 144).

This attitude is nothing but the principle of charity applied to the speaker’s intentions on temporarily relative (i.e. contextual) situations.

There are two basic conditions we need to fill to guarantee that the relevant evidence is complete. First, we have to be sure to know “all that can be known” about the history of the attitudes. This means that we have to be positive about the moments in which they had been held, and be able to predict when they are going to be held in the future. Second, we must be sure we can describe the circumstances under the speaker holds the attitude and those under which he does not. Those conditions are far from easy to determine, since they require an in-depth knowledge of the person who is uttering and having an interpretation system proved to be successful.

In effect, then, a theory of interpretation must be a theory for the understanding of persons, that is, of creatures who are conceived of as embodying a holistic system of belief and desires. As a theory of persons, such a theory must also be explanatory of their behavior as a whole, including [...] both linguistic and extra-linguistic behaviors. [...] Our theories of interpretation must be tested against the totality of behavioral and other evidence, for there is no way in which we could even begin to sort through the evidence independently of some theory of interpretation. The evidence is itself constituted by the theory of interpretation we employ. It becomes a matter of testing ‘total’ theories against ‘total’ evidence (Malpas, 1992, pp. 41-42).

Thus, the choosing of truth attitudes as empirical evidence for the theory of meaning makes it self-contained, i.e. makes the theory part of its own evidence. This is due, mainly, to the impossibility (in Davidson’s approach) of isolating a theory of behavior from a theory of belief and meaning. And the result of such theory is the necessity of
a holistic analysis of the linguistic behavior (which actually stops being only linguistic and becomes ‘total’). In Davidson’s (1979) own words:

It would, I hope it is clear, be a mistake to suppose that we somehow could first determine what a person believes, wants, hopes for, intends, and fears, and then go to a definite answer to the question what his words refer to. For the evidence on which all these matters depend gives us no way of separating out the contributions of thought, action, desire, and meaning one by one (Davidson, 1979, pp. 240-241).

Now that we have the evidential element of the theory let’s try to complete the interpretation of utterances in the cinema case.

When my wife made her statement about the movie public, her words were chosen to express her beliefs and in doing so she was exposing her whole interpretative system at once. It is the combination of her past and present experiences with the cluster of her desires, emotions and feelings that she offers to me. In return, as a rational agent and as a person, I base my understanding of her utterances on my own experiences, feelings and so on, in a process that makes me offer all that I am when trying to interpret her sentence. So, here we are, two interpreters that, based on the knowledge of one another, are capable of judging the truth value of the other’s pronouncements and that based on the same knowledge, are able to figure out the meaning of the other’s utterances. As soon as her sentence escapes her mouth it belongs to us. I’ll use it as an oracle. Based on the charity principle combined with all the information I have about the linguistic and extra-linguistic behavior of my interpretation partner in previous situations, I’ll try to make a correct interpretation of her sentence. And when the result of that analysis becomes public, is going to be the beginning of a new chain of interpretations, judgments and utterances.

**Interpretation in action**

We are now prepared for the final step, the discussion of the content of the sentences. The first thing that has to be noted is that we are not interested in designing a theory about the meaning of the words that compose a sentence, that can be found in a good dictionary. Besides, the knowledge of the meaning of the words doesn’t imply the correct understanding of the sentences such words belong to. That fact is evident in
those circumstances in which we had misunderstood meanings of sentences in the past (and that are now part of our interpretative system’s evidence, as seen before), and in the cases in which the speaker’s intention has to be read between the lines. It is in cases like the last one that Davidson makes his system work.

In his article on the extension of the radical interpretation to malapropisms (Davidson, 1986), Davidson focuses on the idea that in current speaking there are many ways to express one’s beliefs, even when using the same words. And, of course, very different things can be said using the same set of words in different ways. Davidson’s examples of these facts include Diogenes’s statement to Alexander: “I would have you stand from between me and the sun”, where the interpretation of the sentence is to be made by Alexander based on his knowledge of Diogenes’s beliefs. Also including expressions like “they got married and had a child” and “they had a child and got married”, where the order of the words (and not their meanings) are important to understand the content of the statements.

This dynamism of the meaning of the words and sentences, says Davidson, has been forgotten by both philosophers of language and linguistics when they analyze the content of the expressions of a linguistic agent, and this mistake can be corrected by the use of the radical interpretation system.

As a first step toward the clarification of linguistic utterances like malapropisms Davidson refers to Paul Grice’s theory of meaning. Grice makes explicit the difference between what he calls the natural or literal meaning and non-natural or implied meaning of sentences. On one hand they are meanings that can be obtained just by reference to the usual meaning of the words they are made of, those are the literally meaningful sentences. On the other hand, they are expressions whose meaning surpasses the current idiomatic definitions and makes use of an adequate interpretation of the context of utterance, something that must be received as part of the immersion in a language.

It must be noted that Davidson’s use of Grice’s definitions is not something that rests on the fact that they share the view of how a theory of meaning should look. That would be a very weak reason. The interesting fact is that the deep intentionalist base of Grice, which is the reason why he considers important the classification of

---
1 See Grice (1957) and Grice (1975).
meanings for the first time, makes it fit with Davidson’s system of radical interpretation. (Analysis of Grice’s intentional content can be found in Searle (1970) and Strawson (1971)). Davidson’s intent of using radical interpretation theory extends Grice’s differentiation of meanings of sentences to the whole process of understanding utterances. Such extension makes use of the holistic nature and the bi-directionality of the interpretation process. From his perspective, there are two interpretative systems or theories, the prior and the passing theories, that play their role in the understanding of sentences, each of which perform specific tasks when used by the speaker or by the listener.

For the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he does interpret the utterance. For the speaker, the prior theory is what he believes the interpreter’s prior theory to be, while his passing theory is the theory he intends the interpreter to use (Davidson, 1986, p. 472).

Let me exemplify this difference using my original problem. When my wife decided to make her comment, all the beliefs she possesses about my interpretative system are used to find the way to express the information she considers important to offer to me. This means that the way I usually respond to her words, her knowledge of my taste and sensitivity (or lack of taste and sensitivity), summed up with all the facts she knows about me and the context sentence to be uttered make her choose a particular set of words, intonation flux, facial expressions, that will be useful for the transmission of her beliefs. Using the words just defined her prior theory makes her choose the adequate communicative tools. The parts and relations of my interpretative system she is trying to make me use in order to fulfill the interpretation task represent her passing theory. In turn, as a hearer, all the information I have about the context of the utterance and the speaker make me anticipate the way I’m supposed to interpret the sentences to be offered to me. The way I finally decide to interpret the utterances may or may not respond effectively to that knowledge. It could be said that my passing theory doesn’t necessarily fit my prior theory.

Back to Davidson, what is essential for communication to be effective in transmitting meaning is that the interpreters share the passing theory. When involved in the task of transmission of beliefs, which is always the case in human communication, the conventional system used for that purpose is not as important as is the way in which
the utterances are actually interpreted by the speaker and the hearer. This could sound as if there were no place for conventions during the interpretation process. This is not totally true. But it must be pointed out that, as one Davidson’s commentator has said, “convention has a pragmatic role, but is not essential to interpretation and communication” (Evnine, 1991, p. 108). In fact, Davidson’s passing theory must fill two formal requirements. On one hand it must be systematic; the semantic structure of the elements of the utterances has to be solid enough to allow interpreters to understand the meaning of the sentences. As we saw before, the recursive character of the interpretative system that is part of the passing theory fulfills the requirement. On the other hand, speakers and listeners must share the passing theory. There’s no chance to communicate beliefs if the intended interpreter has no clues on how to deal with one’s utterances. The holistic character of the interpretative system guarantees that the passing theory is shared.

But there’s something astonishing about the idea of sharing passing theory as the only requirement for transmission of meanings. Because the passing theory cannot be understood as the interpreter’s linguistics competence, given the beyond-linguistic behavior characteristics it has been filled with, the final result is that what is generally taken for being a language: A “conventional assignment of interpretations to sentences” (Evnine, 1991) is not required for the communication process. Making this result extreme, Davidson’s (1986) advice is that:

There is no such thing as a language, not if language is anything like what many philosophers and linguistics have supposed. There is therefore no such thing to be learned, mastered, or born with. We must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases (p. 475).

Language’s death is the result of Davidson’s confidence on the recursive character of the passing theory adding with the holistic condition of the interpreter’s understanding system. It was a pretty natural death indeed². Now finally, we have all the elements required for the application of the radical interpretation theory on the resolution of the cinema-problem.

² This dramatic conclusion is in turn used as a base for the discussion of the existence of conceptual schemes, but that issue is completely beyond the scope of this paper.
Closing the circle

A brief review of what we have done so far. As an attempt to show her beliefs my wife has uttered a sentence. That sentence contains the information she considers necessary for me to make sense of her pronouncement; something that she takes for granted based on her knowledge of the context of the utterance and of me as an interpreter. Besides, her intention of being understood precisely makes her use a specific order of words and a particular intonation as part of the utterance. On my side, the knowledge I have of my communication partner plus the idea that she is actually trying to express her beliefs, make me hear her declaration and be confident she is telling me what she considers to be the truthful content of those beliefs. When I try to interpret her sentence I’m not only guessing what the words and intonation represent; I’m in fact choosing the interpretation content based on the expertise gained as an interpreter throughout my whole life and in the former cases when I performed this task with my actual companion.

But how could I be sure I did well on interpreting her utterance? I’ll have to follow the course of actions and pronouncements to do so. If after the “I did not like the movie” statement she suggests to go back and see the show again, I’ll have to revise my idea that she hadn’t enjoyed the movie. Or maybe she is trying to give the movie a second chance to be understood. Or maybe in one of the two situations, the utterance or the request, she is trying to trick me. The only way I have to recognize which is the case is to make a new sentence and offer it as a test; perhaps “sure, let’s see it again” would work. Depending on the answer to that last declaration, depending on whether she decides to start laughing at me or return to the cinema, I’ll have to choose which of the possible interpretations was the right one and act accordingly. Even in the case in which we don’t share the language in which the sentences were spoken the method would work the same, although perhaps it would take a little more time and experience to be sure of the success of the interpretation.

Therefore, it’s only during the very process of interpretation when interpretations are constructed and tested. The sentences, gestures and actions used as elements of the system are at the same time part of the system employed to make the interpretation and part of the behavior to be interpreted, joined with the information obtained looking at the temporal context of the utterance. And the conventional rules learned as a part of the language we share, by chance or choice but not as a necessity, help in
the process as facilitators of the interpretative task but are far from being condition *sine qua non* for the understanding of the meaning of the sentences.

So, the best thing I can do to be sure that my interpretative skills are working smoothly and appropriately is to invite my wife again to the movies. And maybe then she would like the show.

References


