

How radical interpretation works

Davidson's theory of radical interpretation is, in an important sense, not a theory of interpretation: it is not, that is, a guide as to how we should interpret people on a day to day basis. What is it, then? It is a foundational theory of meaning and belief: an account of what it is for a word to have a certain meaning.

You might think that it is weird to explain what someone means in terms of how someone or other would interpret them; you'd be right. Meaning does not seem to be hostage to (later) interpretation in this way. Davidson's idea is not that meaning is fixed by how someone or other interprets you, but rather by how an ideal interpreter, following certain rules of interpretation, would interpret you.

From this you can see that - although Davidson might not agree - the figure of the interpreter is only a heuristic. The theory can be expressed without talking about interpretation at all, but only about the rules that Davidson says that the interpreter should follow.

So what are these rules? We can see them as operating in stages.

Stage 1. A certain language-using agent holds true a series of sentences $S_1 \dots S_n$. We take note not only of the sentences held true, but also of the times at which they are held true.

Stage 2. We assume (for now) that the speaker has only true beliefs. Since, as Davidson claims, we can also assume that if the speaker holds true a sentence, then whatever that sentence means is something that the speaker believes, we can assume that the meanings of the sentences $S_1 \dots S_n$ must be truths. (More precisely: they must be true at the time at which the speaker held them to be true.)

Stage 3. So now we have a list of sentences, and a list of the true propositions which might be their meanings. The next task is to give a theory of meaning for the sentences by matching up the truths with the sentences. How do we do this? (i) The use of tensed statements. (ii) The requirement that words be interpreted constantly throughout $S_1 \dots S_n$.

Stage 4. Most likely, we will reach an impasse at Stage 3; there will be a conflict between the claim that the speaker believes only truths and the requirement that the theory of meaning interpret words in the same way throughout (and be compositional). Imagine, for example, that included in $S_1 \dots S_n$ are the sentences 'Grass is green,' 'Grass is in the yard', 'Cows eat grass', and 'Grass is a kind of fern.' How would a theory of interpretation work for such sentences? Here, as Davidson says, it is a matter of the best overall fit.