FODOR, Jerry. Semantics: an interview with Jerry Fodor. *Revista Virtual de Estudos da Linguagem - ReVEL*. Vol. 5, n. 8, março de 2007. ISSN 1678-8931 [www.revel.inf.br].

SEMANTICS – AN INTERVIEW WITH JERRY FODOR

Jerry Fodor
Rutgers University

ReVEL – There are certainly many different answers to a question such "What is Semantics?" But we will dare to ask you anyway: in your opinion, what is Semantics and what does it study?

Fodor - I suppose a semantics theory of a language, natural or artificial, is part of a grammar of that language. In particular, it's the part of a grammar that is concerned with the relations between symbols in the language and the things in the world that they refer to or are true of. The analogy is to a syntactic theory of a language. As that notion is understood by 'generative' grammarians, a syntax is about which expressions are 'well formed' in the language that it describes; in particular, it distinguishes the expressions that belong to the language from the ones that don't, and it represents certain structural properties of complex symbols including, importantly, their constituent structure. The intuition is that syntax is about how the expressions in a language are put together, and semantics is about how they relate to their referents in the nonlinguistic world.

This is to take a more or less 'Tarskian' view of semantics. So, as Tarski says, a correct semantics of language L would, at a minimum, determine the conditions under which the (declarative) sentences of L are true. A semantics for English would thus include, among its infinity of entailments, the theorem that the English sentence 'snow is white' is true if and only if the world is such that 'snow is white;' the English sentence 'Kant was a philosopher' is true iff 'Kant was a philosopher,' and so forth. Appearances to the contrary notwithstanding, such 'Tarski sentences' are by no means trivial or empty. You can see this if you

assume that the semantic properties of English sentences are described in some language *other than* English. That 'the cat said 'meow" is true if and only if le chat a dit 'meow' is patently *not* trivial. To the contrary, it's just the sort of fact that a French speaker who is trying to learn English (or an English speaker who is trying to learn French) would need to know.

Several caveats:

- I assume that the system of *mental* representations constitutes a language ('Mentalese'); accordingly, the proposed understanding of what semantics is about is intended to hold, inter alia, for mental representations (in particular, for concepts and the constructions that they enter into, both of which are thought of, for these purposes, as formulas of Mentalese). I think it is very likely that only mental representations have semantic properties (truth and reference) 'in the first instance'. Formulas in natural languages inherit their semantic properties from those of the mental representations that they are used to express. To a first approximation, 'snow is white' means that snow is white in English because it's the form of words that English speakers use to express the belief that snow is white.

- This proposal is, in several respects, a relatively exiguous understanding of what semantics is about. In the Empiricist tradition especially it has generally been supposed that semantics should specify which of the formulas in a language are 'analytic' or 'true in virtue of their meanings alone'. So, for example, a correct theory of the semantics of English would entail that 'x is a bachelor' means something like *x is an unmarried man,* and so is true of a person x if and only if x is male and unmarried. Accordingly, the sentence 'if x is a bachelor then x is unmarried' is true in virtue of the meanings of its constituent terms; it expresses an analytic truth. This view of semantics as concerned primarily with linguistic (/conceptual) truth is still widely prevalent among linguists; and many 'analytic' philosophers hold it in some version or other. That they do is not surprising. This kind of theory purports to explicate notions like 'conceptual truth', truth of 'informal logic,' 'truth of depth grammar' 'criterion' etc. These various notions differ from one another in various ways; but they share the idea that some truths are necessary and a

priori *because* they hold just in virtue of the meanings of the symbols that express them. If that is so, this kind of semantic theory would rationalize undertakings like the analysis of word meanings or of conceptual content; and, according to many philosophers, such conceptual/linguistic analyses are the typical products of philosophical inquiry. However, for (inter alia) reasons that are familiar from the work of Quine and his followers, I very much doubt that this conception of semantics can be sustained.

- I assume that, in all the languages of interest (including English and Mentalese), there are infinitely many expressions that can be evaluated for truth and reference: 'this is the cat', 'this is the dog that chased the cat', 'this is the dog that chased the cat that ate the rat'... and so forth indefinitely. Most of the work of a semantic theory is to explain how the semantics of these (infinitely) many formulas are determined by their syntax together with the semantics of their (finitely) many primitive constituents. So, very roughly, 'the cat ate the rat' is true if and only if the cat in question ate the rat in question; and that, in turn, is true because 'the cat' refers to the cat and 'the rat' refers to the rat and the world is such that the former ate the latter. This is the sort of thing semanticists have in mind when they say that the semantics of natural languages (and of 'Mentalese') must be "compositional". As it turns out, compositionality is a very strong constraint on semantic theories; one which, quite possibly, can be met by only theories that identify the fundamental semantic properties of symbols as truth and reference.

ReVEL – In your opinion, why does "meaning" seem to be central to everything human, as Ray Jackendoff puts it in his *Foundations of Language*?

Fodor – Actually, I think it isn't. If meaning seems to be ubiquitous that's because it's used as a cover term for all sorts of things that are, in fact, quite different from each other. 'Meaning' is, in short, radically ambiguous; not noticing that it muddies the water in all sorts of ways.

For example, the meaning that's referred to in notions like 'the meaning of a word' is quite different from the meaning that is referred to in 'smoke means fire', which is again quite different from the meaning that is referred to in 'I can't tell you how little existential phenomenology means to me.' It's quite easy to show that this is so. Consider the following argument: "'Smoke' means smoke; smoke means fire; therefore 'smoke' means fire'. Clearly the argument trades on the ambiguity of 'means': In the first premise, 'means' means something like REFERS TO; in the second premise it means something like INDICATES. If, however, you just take for granted that 'means' means the same thing in both premises, you won't be able to explain why the argument isn't valid.

Pace Jackendoff, I think that for purposes of theory construction, we would be well advised to forget about the everyday notion of meaning, the one according to which it seems to saturate our lives. Even in what purports to be scientific discourse, there are all sorts of things that psychologists and philosophers have meant by meaning, and attempts to bring all of them into the same framework of theory have quite generally been unsuccessful. Consider psychological associationism as a pertinent example. Heaven only knows how many books and articles have been written, over the last 150 years or so, which claim that the meaning of a word or the content of a concept is the set of associations that it has a high probability of evoking. That can't be true, of course; 'Dog' is a high associate of 'cat', but 'cat' doesn't mean 'dog'. Nonetheless, the confusion of meaning with association persists in psychology and continues in cognitive science at large. Such currently fashionable views as that conceptual contents are stereotypes or that concepts are arranged in a 'neural network' are current version of traditional associationism and they succumb to the traditional anti-associationist arguments.

Still, I do think that there is a semantic notion that is of central theoretical interest in psychology and linguistics and that does some of the work that meaning has traditionally been supposed to do; namely, *REPRESENTATION*. Perhaps the most important thing to understand about the cognitive mind is that it is somehow able to represent the world. What makes that so important is

that, all else equal, how one acts is determined by how one represents the world (rather than by how the world actually is.) Of course, when all goes well and one's belief about the world is true, the way that one represents the world as being *is* the way that the world actually is. It's in such cases that actions based on one's beliefs are most likely to be successful. By contrast, if actions based on false beliefs ---that is, on misrepresentations of the world--- succeed, it's an accident that they do so. Likewise, it's part of what English-speakers know about English that if someone utters 'John is hungry' (and a variety of what John Austin called 'felicity conditions' on speech acts are satisfied) he is telling you that John is hungry. It's in such cases that actions based on what one has been told are most likely to be successful.

That's fine as far as it goes; it's entirely plausible that, from the semantic point of view, the essence of language and mind is representation. But this claim lacks a metaphysics; it doesn't tell us what representation *is* except that it's typified by symbol-world relations like truth and reference. How to understand the metaphysics of representation, is among the deepest and most hotly debated of current philosophical issues; all the more so if one accepts on the assumption that the metaphysics of representation must be 'naturalistic'. That is, a representational psychology (/linguistics) must be compatible with other empirical theories that we independently have reasons to believe are true; for example, with 'brain science', to say nothing of nonbiological theories in chemistry, physics and so forth. Lots of philosophers who do assume some sort of naturalism think that a proper theory of representation would construe semantic properties as in some way constituted by *causal* relations between the mind and the world. That strikes me as plausible prima facie since, in many kinds of cases, it would seem to be our causal encounters with the world that make our thoughts have the contents that they do. Very, very roughly, the paradigm might be that the concept DOG represents dogs because interactions with dogs cause us to think *dog*. The operative phrase here is 'very, very roughly'. Nobody knows, in any detail, how a causal theory of representation might actually work; but it had better work somehow or other if the line of thought I've been pursuing is even close to being right.

ReVEL – The problem of meaning and reference is classical in the study of Semantics and Philosophy of Language. And you have made some interesting contributions to it. How can you compare your approach to the problem to the approach given by Noam Chomsky, for example?

Fodor - I'm not at all sure that I understand how Chomsky views semantics. But, as far as I can tell, he thinks that it's not about *relations between ideas and the world*, but rather about *relations among the ideas themselves*. The typical kind of semantic relations among ideas, on that sort of view, are the ones that engender analyticities (such as that whatever falls under the concept BACHELOR ipso facto falls under the concept UNMARRIED MALE; see above.) There is, in philosophy a very long history of holding such views (Hume, for example, appears to have done so; maybe Kant did too.) Often the reasons for holding it have been epistemological. The line of thought is something like this: Since knowledge involves representation, one can't know what the world is like 'in itself' viz how it is independent of the ways that we represent it. So, if representation is itself a kind of mind-world relation, we can't know whether we ever do succeed in thinking about the world. (/about what our words mean, etc.) Suppose, however, that representation is constituted by relations *among* thoughts. Since we *can* know about such relations (by introspection for example) we likewise can know for sure such putatively analytic truths as that bachelors are unmarried, that cats are animals, and so forth. In effect, the proposal is to avoid skepticism about knowledge by adopting a sort of Idealism about meaning: all our ideas are ideas about ideas.

As I say, I'm not at all sure that this is Chomsky's view. I hope it isn't because, whereas my confidence in many other of Chomsky's views is practically unbounded, succumbing to representational Idealism strikes me as a strategy that is to be avoided at all costs. Here are some reasons, all of which I take to be more or less truistic:

- It is wildly implausible that we don't, at least some of the time, think about the world. Semantic Idealism seems to deny this and hence to be false on the face of it.

- The Idealist sort of semantics requires that there are lots of analytic propositions (enough to fix the content of each of our concepts). But the evidence is that there are, at most, not very many. (Quite possibly there are none that are untendentious including 'unmarried men are bachelors'. Is the Pope a bachelor?)

- The view of meaning that I'm supposing Chomsky endorses avoids skepticism about whether bachelors are unmarried; we really can know that they are; in fact, anyone who has the concept BACHELOR *must* know that they are. Likewise for the knowledge that if John killed Mary, then Mary is dead, etc. But it's very unclear how this is supposed to work for knowledge of 'contingent' propositions (for example the case of one's perceptually grounded true belief that the cat is on the mat.) In such cases, our knowledges simply *can't* come from our grasp of relations among ideas: It's not part of the idea CAT that this one (the one I'm, just now looking at) is on a mat; and it's not part of the idea MAT that this one now has a cat that's on it. It's plausible on the face of it, that empirical knowledge is a mind-world relation. That being so, it would seem that Semantic Idealism avoids skepticism about 'conceptual truths' only at the cost of making a total mystery of empirical truth. (It's notable that the currently fashionable post-Modern relativism about truth, knowledge and the like (which, by the way, I entirely abhor, and so should you) invariably starts by assuming that there is 'nothing beyond the text'; viz that our concepts are constrained by the their relations to one another but not by their relations to the world.

- For essentially similar reasons, semantic Idealism can't account for the fact that, at least some times, we are able to make rational choices among conflicting beliefs; in particular, among conflicting scientific theories. Rather, according to semantic idealists, theories can't be rationally compared because what their terms in a theory means is determine internal to the theory. If I think dogs have tails and you think they don't, then we must 'mean something different' by 'dog'

so there's no way of settling what appears to be the disagreement between us. In fact, strictly speaking, there *isn't* a difference between us. For a really flagrant example of this dialectic working itself out, see Thomas Kuhn's *The Structure of Scientific Revolutions,* according to which scientists whose theories differ radically "live in different worlds". There must be something wrong with a semantics that entails this sort of thing since, patently, we all live in the very same world; viz *this* one.

The long and short is that Idealistic semantics rejects the notion of mind-world correspondence, renders the content of our beliefs intractably holistic, and makes it unintelligible that any of our beliefs are rational. I doubt that Idealist semantics is worth having at that price.

- If semantic relations hold only among ideas, then everything we can think about is mind-dependent. But it's simply untrue that whatever we can think about is mind dependent. For example, we can think about The Grand Canyon, which surely was around before there were any minds and presumably will continue to be when all the minds are gone. The world (consider as the potential object of indefinitely many thoughts) is *prior* to the mind. A fortiori, the objects of thought can't themselves all be mental.

It's an infallible sign of bad semantics that it leads to bad metaphysics.

ReVEL – How does the idea of *Mentalese* relate to other semantic theories available on the market? We know that this idea has already been challenged by a variety of philosophers and linguists. What are the main arguments from the critics and how do you respond to them?

Fodor – The story about 'Mentalese' isn't, and has never purported to be, a version of semantics. To the contrary, if (as I suppose) Mentalese is a language (viz the language in which one thinks) then it itself requires a semantics, just like any other language. On the roughly Tarskian view of semantics that I've

been sketching, an acceptable semantics for Mentalese would have to entail, for example, that the concept DOG is satisfied by and only by dogs; that the thought that *that's a dog* is true if and only if, that's a dog; and so forth for the indefinitely concepts and thoughts that our psychology permits us to entertain.

In short, qua language, Mentalese needs a theory of the truth and reference of its formulas. Indeed (as remarked above) it may be that Mentalese is the *only* language that needs a semantics; for example, that words and sentences of English have the contents they do because they are used to express the content of the corresponding words and sentences of Mentalese. This is, to be sure, a 'psychologistic' theory of linguistic content. As far as I can tell, it's none the worse for being so.

So, the Mentalese story is about (not semantics but) the character of mental representations; it's the theory that the mental symbols that we use to represent the world in our thoughts are like sentences (and not much like, for example, pictures). The arguments for this view are, I think, pretty nearly overwhelming. On the one hand, it's needed to account for the productivity of thought; just as English grammar places no upper bound on the number of sentences that are available for us to utter, so the grammar of Mentalese places no upper bound on the number of thoughts that are available for us to have. Likewise, it's required in order to connect cognitive psychology with logic. It does so by explaining how the '1ogical form' of inferences can affect the course of our thinking in inferential processes and it's required to connect cognitive psychology with the theory of computation. It does so by explaining how mental *objects* like thoughts and concepts can provide domains for mental processes like *reasoning*; namely, by treating mental processes as species of computations which are, by definition, formal operations defined over the syntactic structure of representations.

This last point is no small matter. One of the main things wrong with the empiricist tradition in representational theories of mind was its commitment to an associationistic treatment of cognitive process which, in the event, proved to be utterly untenable. That the computational treatment of cognitive processes

offers a radical break with the associationist tradition may well be the most important idea that grounds our current cognitive science. And, to repeat, the theory that mental processes are computations depends on the theory that mental representations are sentence-like; in particular computations that mental representations have constituent structures.

It seems to me overwhelmingly plausible that, if one is going to endorse a representational theory of mind at all, one ought to opt for the Mentalese version. But the arguments for Mentalese, though they support the theory that we think in some sort of language, leave open which language it is. They don't, for example, rule out the possibility that 'we think in English'. So, although the canonical versions of the language of thought thesis hold that Mentalese is an unlearned representational system with many of the formal properties of a logic, it is possible to hold a much less dramatic version of the thesis consonant with its letter if not with its spirit. One might hold that we think in whatever language we speak in; e.g. that English speakers think in English, French speakers think in French, and so forth. This is, perhaps, the only version of a representational theory of mind that many philosophers are even remotely inclined to accept. Unaccidently, it has striking affinities with the sort of Watsonian behaviorism according to which thinking is a kind of talking to oneself. Though few philosophers explicitly endorse behaviorism these days, it's surprising how much of it one finds alive just below the surface.

I think, however, that the identification of Mentalese with English really isn't an option. The most persuasive consideration is the truism that English has to be *learned,* presumably by some sort of inductive (or abductive) inferences over what one hears in one's linguistic environment. But since drawing inferences is itself a kind of thinking, the theory that one thinks in a language that one has learned is doomed to circularity; the identification of Mentalese with English (or, mutatis mutandis, any other 'natural' language) really is out of the question.

If that's right, then there is decisive reason to dissociate two uses of languages that philosopher often run together, sometimes as a matter of principle. On the one hand, there's the role of language as the medium in which cognitive

processes are typically couched; and, on the other hand, there's the role of language in mediating communication between speakers and hearers. The line of thought we've been pursuing suggest that these functions must be performed by *different* languages: Mentalese is employed for the first, but not for the second; English is employed for the second but not for the first. This bears emphasis in light of claims by Wittgenstinians (and, by the way, of Whorfians) which would seem to deny that such a dissociation is possible; to hold, in effect, that only a 'public' language really *is*. To my knowledge, no serious argument for that view has ever been proposed; nor, as far as I know, to my knowledge, has any Wittgensteinian (or Whorfian) offered so much as a sketch of an account of how, if English is itself the vehicle of thought, how learning English is so much as possible (Wittgenstein says, not very helpfully, that it's a matter of 'training'.)

If, however, that line of argument doesn't convince you, there are others on offer. On even cursory examination, English would appear to be a bad choice as the representational format for thought. I mention just two of the many reasons. First, English is full of both structural and lexical ambiguities, and it's thoroughly unclear what it could be to think an ambiguity. Notoriously, 'everyone love someone' is ambiguous with respect to the scope of the quantifiers. But could one think the thought that everyone loves someone without choosing between the possible scopes? Is it possible to think that everyone loves someone and simply not know whether one is thinking that there is someone that everyone loves? What would it be like to be in such a state? And, importantly, how on earth could one get out of it if one had once gotten into it?

The point this illustrates (that the language of thought must be ambiguity free) is, in fact, special case of a quite general consideration: mental representations must be explicit as to their logical form. Philosophers have been pointing out, literally since Aristotle, that natural languages don't meet this condition; that is, they aren't explicit about the properties of thoughts that determine their roles in inference. The inevitable conclusion seems to be that one doesn't think in a natural language.

I repeat the point I made above: If one is going to have a representational theory of mind at all, the version of choice is clearly the one that makes Mentalese its format (where, by assumption, Mentalese isn't a natural language; it isn't ever used as a vehicle of communication.) To be sure, this leaves it open that one might refuse to endorse a representational theory of mind of any sort. But the only kind of alternative I've heard of is some sort behaviorism, and we've been down that road before. It leads nowhere.

ReVEL – As an experienced semanticist, philosopher and cognitive scientist, can you please suggest some basic, essential, outstanding or classical readings in the field of Semantics?

Fodor – This is a hard question to answer, because many of the classical books and papers that treat semantics in the way I've endorsed are relatively technical (some logic and/or some linguistics is required to read them) and they don't even purport to provide an overview of the theoretical options. And, as the previous discussion must have suggested, there is considerable and vehement disagreement not just about which semantic theory is best, but even about what problems such theories are supposed to address and what data they are supposed to accommodate. The best I can do is suggest a scattering of texts which provide a more or less untechnical treatment of some kinds of issues that I take to be central. The items I've starred are collections of papers, many of which are foundational.

*Antony and Hornstein (eds) CHOMSKY AND HIS CRITICS
Fred Dretske, KNOWLEDGE AND THE FLOW OF INFORMATION
Hartry Field, 'Tarski's theory of truth'
*Geirsson and Losonsky (eds) READINGS IN LANGUAGE AND MIND
Christopher Hughes KRIPKE; NAMES, NECESSITY AND IDENTITY
Saul Kripke, 'NAMING AND NECESSITY
Hilary Putnam, 'The meaning of 'meaning"
Jerry Fodor, THE LANGUAGE OF THOUGHT

Jerry Fodor and Ernie Lepore, THE COMPOSITIONALITY PAPERS
Jerry Fodor and Zenon Pylshyn ‘Connectionism and cognitive architecture’
*Margolis and Laurence (eds) CONCEPTS
*Stich and Warfield (eds.) MENTAL REPRESENTATION