Deflationism and Inferentialism

Jaroslav Peregrin[*]

The use theory of meaning

After putting forward his celebrated deflationary theory of truth (Horwich, 1998a), Paul Horwich added a compatible theory of meaning (Horwich, 1998b). I am calling also this latter theory *deflationism* (although it may be a slightly misleading name in that, as Paul himself notes, his theory of meaning is deflationary more in the sense of being forced by the deflationary theory of truth than of being particularly deflationary in itself). In contrast, what I call *inferentialism* is the theory of meaning which I am going to advocate here – the view, in a nutshell, that meaning is a matter of inferential role. Various versions of this theory have been defended by Wilfried Sellars, Robert Brandom and a couple of other philosophers including myself. And the thesis I wish to present in this paper – to put it as a provocation right off – is that Paul is an inferentialist led astray.

Both deflationism and inferentialism can be seen as elaborations of what can be called the *use theory of meaning*; for both seem to agree that:

(U)   *the meaning of an expression consists in the way in which the expression is used by the speakers of the relevant community.*

To elucidate the consequences of this view and to indicate why I think it naturally leads us to inferentialism, let me summarize the analysis I have given in greater detail elsewhere (see Peregrin, 2007).

First, let me stress that (U) is significantly more substantial than the *prima facie* similar claim:

(U*)   *any meaning an expression (i.e. a sound- or inscription-type) has, it has in force of the fact that it is treated in a certain way by the speakers of the relevant community.*

Few people would wish to oppose (U*): it is generally accepted that sounds or inscriptions do not mean anything *by themselves*, but only *due to us*.

One crucial difference between (U) and (U*) is that the term *use*, as employed in (U), is taken as something more specific than the term *treatment* in (U*); it amounts to treatment, as it were, *in the outer world* (as contrasted with the treatment *in the inner world of one's mind*, which we will call *conception*). Another difference is that while (U*) states that an expression has meaning *in force* of our treatment, (U) states that the meaning directly *is* the treatment

(usage). Hence, to get from the generally acceptable (U*) to the more controversial (U), we must make two substantial steps:

    (i)   identify any meaning-conferring kind of *treatment* with *use*;

and

    (ii)   identify meaning directly with the use.

Why should we (and should we at all) make these steps?

Use versus Conception

Let me consider (i) first. Why should we see meaning as a matter of usage rather than of conception? Why should we not say that an expression means thus and so iff the speakers conceive of it in a certain way, perhaps take it as a sign of something? The basic trouble is that as conception is a private, subjective matter (at least until it becomes manifested by behavior), it cannot serve as the foundation of the essentially intersubjective institution of meaning. As Davidson (1990, p. 314) aptly stressed, "that meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language".

To be sure, if an expression has a meaning within a linguistic community, then the speakers of the community *will* conceive of it in certain specific ways. However, this is not enough to establish the fact that it means what it does. An essentially private act of conception is not capable of grounding the essentially public institution of language. That people of some community mentally associate the word 'spider' with a certain kind of animal is a fact of their individual psychologies not capable of establishing the fact that 'spider' expresses, within their language, the concept of *spider* – for in order for it to express the concept, it is not enough that each person individually makes the association, he/she must also know that the others do the same, that he/she can use the word to refer intelligibly to spiders in various public circumstances etc. Hence what is needed alongside any private associations are some public practices that make the link between the word and a concept public and shared.

Moreover, once the practices are in place, the private associations become redundant – from the viewpoint of the institution of language (though, of course, not from the viewpoint of the psychology of communication) they become the idle wheel whose presence or absence makes no noticeable difference. This is the point of the famous case of 'Wittgenstein's beetle' (see Wittgenstein, 1953, §293). Wittgenstein invites us to imagine a game involving its players having boxes with 'beetles'; but none of the players can look into anyone else's box to see his or her 'beetle' (nor to see whether there is anything in the box at all). Wittgenstein then points out that, given this, the content of the boxes is wholly irrelevant from the viewpoint of the game – for the game can be played in the very same way independently of what really is

in the players' boxes. And, Wittgenstein concludes, in so far as our minds are such 'inopenable boxes', their contents are irrelevant from the viewpoint of our language games.

Another way of expressing the same point is to say, as Davidson does, that the very *point* of meaning is that it can be shared by many: that new people can always enter the realm of a language, learning the meanings of its words and then participating in the language games staged by its means. As Quine (1969, p. 28) stressed, "each of us, as he learns his language, is a student of his neighbor's behavior" and "the learner has no data to work with, but the overt behavior of other speakers". In so far as language and meaning is something essentially intersubjective, the contents of the minds of the speakers cannot be its *components*. Thus Quine (*ibid.*, 29) concludes: "There are no meanings, nor likenesses or distinctions in meaning beyond what are implicit in people's dispositions to overt behavior".

However, here we come to the neuralgic point of the usual formulations of the use theories of meaning. We want to express that meanings are a matter, not of the peculiarities of the ways individual speakers put a term to use, but rather of some 'principle behind this'. We want to state that there is a relevant interconnection between the occurrences of spiders and the utterances of 'lo, a spider!'; despite the fact that only some of the speakers, and only sometimes, would actually react to a spider with this very utterance. And the usual technique, adopted also by Quine, makes use of the concept of *disposition*. The sentence 'lo, a spider!' means what it does because the speakers have *the disposition* to utter it when confronted with a spider.

However, what is a disposition? A disposition is a property whose nature is unclear and which consequently is characterized in terms of the potential behavior of the entity in question in some special situations (thus, e.g., to say that sugar is soluble in water is to say that in the circumstance of being put into water, we should expect it to dissolve). But how should we characterize a disposition of the above kind, i.e. a disposition like to utter 'lo, a spider' when confronted with a spider? We should be able to say something to the effect that to be disposed to emit 'lo, a spider' in the presence of a spider is to emit it whenever there is a spider around and some further conditions are fulfilled – but which conditions? That the person in question has no reason to stay silent? That she wants to let others know? That she is not dumb, nor too lazy, nor afraid to talk, etc. etc.? Obviously none of this approaches an accurate characterization of the relevant circumstances.

These obstacles lead us into a true vicious circle: we claim that the meaning of a sentence is a matter of a disposition to utter the sentence; we reduce dispositions to specific behavior in specific situations; but in this case we are unable to specify the relevant circumstances other than as those circumstances in which the relevant sentence is really uttered; hence we say, in effect, that the meaning of a sentence is a matter of uttering the sentence in those situations in which it is really uttered. Of course, proponents of the dispositional analysis will claim that there *is* a possibility of characterizing the relevant circumstances explicitly (and that, moreover, the disposition is ultimately a matter of as yet unknown physical properties of the brain), but the fact that nobody has been able to progress very far in this direction seems to justify strong skepticism here.

I think that these obstacles are fatal; and hence that the dispositional elaboration of the use theory of meaning leads us up a blind alley. I think that the relevant relationship between 'lo, a spider' and spiders, the one which is responsible for the former to mean what it does, is of a different kind. To see of what kind, let us now turn our attention to the above mentioned step (ii).

Meaning versus Representing

Even if we do accept that meaning is determined by our usage rather than by our conception, why go on to say that meaning directly *is* the usage, instead of seeking a thing or an abstract entity which gets associated with an expression in terms of its use? Why give up the *prima facie* plausible picture of language according to which meaning is what is represented and embrace the much less intuitive notion of meaning as the 'way of usage'?

The core trouble is that it is hard to see the concept of *representing* (or *standing for*, or *naming* or *codifying* or ...) as an unexplained explainer. For what does it take for an expression to represent, for a community of speakers, something? That each of the speakers conceives of it in a certain way? We have already seen that the essentially private act of conception is not capable of grounding the essentially public institution of language. So it seems that the *representing* relevant here must be a matter of some communal *practices*. This indicates that an explanation of language which rests on the relation of *representing* cannot be considered as a satisfactory ultimate explanation, but only, at most, as an intermediary step, inviting a subsequent step: explaining the very relation of *representing*. And since resting this last explanation on the facts of conception would not do, we have to proceed to the level of social practices.

It was for this reason that Wittgenstein sought the ultimate explanation of semantics on the level of language games; and it has been, more generally, for this reason that various kind of use theories of meaning have started to flourish. However, we have seen that these theories are plagued with untenable dispositionalism. Can we improve on them?

Wittgenstein urged that, multifarious as our language games might be, they are typically g*overned by rules*, and moreover, by rules which appear to be somehow *implicit*. And Wittgenstein also urged that it is precisely in this way that an expression can acquire meaning other than by being made to stand for a thing. Meanings may be identified with the *roles* which the expressions play *vis-à-vis* the rules – roles of the kinds of those which make pieces of wood used in chess games into pawns, rooks, or kings[1].

Why *implicit* rules? Because, as Wittgenstein realized, the rules of language cannot be all explicit – in pain of a vicious circle. We do have explicit rules of chess – we can take a book and read them there. However, to do this, we must know how to interpret the signs in the book – we must know the rules of their interpretation. Perhaps also these rules are somewhere written, but it is clear that the regress must come to an end and at some point we must be able

---

[1] For a detailed discussion of the notion of meaning as a role see Peregrin (2006a).

to follow the rules of interpretation without their being explicit. Elucidation of the nature of the implicit rule-following practices was one of Wittgenstein's principal aims in *Philosophical Investigations* and subsequently became the topic of one of the most heated philosophical debates of the second half of the XX. century.

What is important is that the realization of the key role of rules enables us to dispose of the troublesome concept of disposition. The point is that, as we can now see, the correct description of the link between a sentence meaning that there is a tiger around and the fact that there is a tiger around is not that the speakers are *disposed* to utter the former in a case of the latter, but rather that it would be, for them, *correct* (conforming to certain rules of language) to do so. And, whereas saying that one is *disposed* to do something amounts to predicting that given suitable conditions one will inevitably do it, to say that one would be correct in doing so does not involve any prediction of this kind.

However, does it not follow that the correctness claim is merely chimerical in that it cannot be confirmed nor disconfirmed by anything the speakers of the relevant language actually do? Not really. The acceptance of rules, albeit implicit, *must* be manifested by what they do; but it is manifested "on the metalevel" – namely by what Brandom (1994) calls the competent speakers' *normative attitudes*. We *take* some utterances for correct and we *take* others for incorrect – which may be manifested in various ways, from praising or rebuking our children for the way they talk to granting our fellow speakers various kinds of statuses, from "respected" or "reliable" to "devious" and "untrustworthy".

Of course, our linguistic utterances can be classified as correct or incorrect in various senses; and consequently we have, if not entirely a motley of rules, then at least a multiplicity of their layers. An utterance may be correct in that it accords with the grammar of the language in question; it may be correct in that it says that things are in the way they really are; or, it may be correct, say, in that it is not offensive to an audience. The rules directly relevant for semantics are supposed to form one of these layers: namely the one which has to do with, as Aristotle put it, "saying of what is that it is" and which is normally associated with the concept of *truth*. Hence we may say that the relevant sense of "correct" is the one in which we can say that truth amounts to correct assertability.

To avoid misunderstanding: this does not pave the way to the straightforward naturalization of the concept of *truth* and *meaning*. The problem consists in singling out the kind of correctness which amounts to truth without relying on the very concept of truth. Understanding this peculiar kind of correctness is apparently a matter of acquiring a know-how which is explicitly manifested by our usage of the very concept of *truth* – with the result that the specification of the relevant kind of correctness has to rest on the concept of truth rather than vice versa[2]. Anyway, we may say, the recognition of the normative dimension of language, which is brought about by the realization of the key role of (various kinds of) rules within our language games makes us replace the concept of *disposition* engaged by the non-normative use-theories of meaning by the concept of *propriety*.

---

[2] See Peregrin (2006b) for more details.

Normative mode of speech

However, the deflationist might wonder, is the inferentialist not moving in a circle? I have rejected the possibility of explaining meaning in terms of dispositions to produce certain utterances and in its stead suggested explaining it in terms of taking these utterances for either correct or incorrect. This, I further suggested, is manifested by the speakers' perceptible attitudes to other speakers' (and also their own) utterances. But surely it is not the case that always when somebody does something incorrect, he comes to be beaten with sticks! So if we want to avoid subjectivism of the Cartesian kind, we must, it might seem, accept that the normative attitudes are a matter of mere dispositions to overtly treat others in certain ways. So are dispositions, which we have thrown out of the door, back via the window?

No. To say that it is correct/incorrect to utter a certain expression in a certain way in certain circumstances is *not* to say that the speakers (have a disposition to) reward/penalize this kind of usage. True, the events of such rewarding/penalizing constitute excellent *evidence* for the claims of correctness/incorrectness, but the normative claims are not translatable into the non-normative ones; they are not translatable into *any* non-normative claims. To say that something is correct is *not* to say that whoever does this will be rewarded, nor that it complies with a rule accepted by the relevant community, nor anything else of this kind.

I suspect that at this point the deflationist may gain the impression that the inferentialist is cheating. He declared he was going to rid us of the troublesome concept of disposition, but it has turned out that the price we must pay for it is the acceptance of the irreducibility of the normative idiom. This is correct: *If you want to buy inferentialism, you must buy the irreducibility of normative talk.* Normative claims neither correspond to, nor are verified by, the kind of facts which are spelled out by declarative sentences. However, it is no cheating but rather the appreciation of the true role of the descriptive idiom" namely that to rely on this idiom as if it were the only 'really' (or 'independently') meaningful one, and to think that any other kind of talk can become so only if it can be reduced to it, are mere superstitions.

Why should we require the reducibility of normative claims to descriptive ones? 'Either they state some facts, and then they are reducible to descriptions of these facts, or they do not and then they are meaningless!' However, why should we equate meaningfulness with stating facts? After all, questions are clearly meaningful and they certainly do *not* state facts. 'However, questions *do* have their peculiar role within the fact-stating business, they *prompt* stating facts!' Well, likewise, normatives play an important role within our linguistic activities, and are also interconnected with utterances which can be described as 'stating facts'.[3]

'However, what do normative claims *mean* if they are untranslatable into non-normative talk?' Well, we *know* what they mean: we *know* what it means to say that killing is wrong or that one ought to say that snow is white if snow is white. How do we know? Just like we

---

[3] And we can offer plausible stories about why and how we have developed them (such a story can be found in Brandom, 1994, Part One).

know what it means to say that snow is white or to ask whether it is white: we have learned our language which contains normative statements as an important part.

Take football. Nobody would dream of explaining what it is without referring to its rules and without pointing out that a player *ought not* to touch the ball with his hands, etc. Does this make this explanation meaningless or somehow illusory? Clearly not. Is the explanation translatable into a non-normative language? Can we, instead of saying that the player ought not to touch the ball with his hands, say that if he does so, the referee will punish him? Not really: the referee may fail to notice his foul. Can we say that the referee will have the disposition to punish him? Well, due to the problematic concept of disposition this would amount to a mere illusion of an explanation. Can we say that that there is a rule which prohibits touching the ball? Surely: but this does not move us towards the reduction of the "ought", for to explain what a *rule* and a *prohibition* are gives rise to the very same kinds of problems.

The point of the rules of football consists in that, expressed metaphorically, they open up a new space - an inexhaustible space of football games. Were it not for the rules, this space would not exist (as is well known to anybody who has ever tried to play football with people unwilling to follow the rules). And the inferentialist wants to see language precisely on the model of football (as Lance, 1998, duly pointed out, football may make a better case than the more popular chess): just as the rules of football open up the space of football games (populated by myriads of events from occasional matches to the World Cup), the rules of language open up the space of communication, and more generally of meaningfulness of the distinctively human kind (in which we virtually immerse ourselves during the process of our literarization).

Horwich's concept of acceptance

The inferentialist claims that the inferential role of an expression is a matter of an *inferential pattern* – a basic set of inferential rules which determine the overall inferential behavior of the expression. Thus, the meaning of *and* is determined by the pattern[4]

    A and B |-- A
    A and B |-- B
    A, B |-- A and B

What Horwich says, *prima facie*, is not strikingly different: the meaning of a term is a matter of its *acceptance property*; in particular the word "and" means what it does in virtue of the fact that we tend to accept "*p* and *q*" if and only if we accept "*p*" and "*q*". The difference

---

[4] With an oversimplification, of course. The English 'and' functions also in other ways; e.g. as indicating a temporal sequence.

seems to be Horwich's engagement of the term "tend"; but, as I have argued above, this difference is far from insubstantial.

We have seen that the dispositional words of the kind of "tend" do not allow for a reasonable explication. One possibility of their interpretation is psychological – explicating "tending to accept a proposition" as "being in a certain psychological relationship to the proposition". Disregarding my skepticism about the explainability of such a psychological relationship, this leads us back into the muddy waters of the subjectivism we were so happy to have extricated ourselves from. Another possibility is statistical – explicating "tending to accept a proposition" as "overtly assenting to the utterances expressing the proposition in more than x % cases". But this is clearly even more hopeless.

In fact, Horwich does give a rudimentary theory of accepting, which should help us understand what he means by the term.

(1)   For each observable fact O there is a sentence type "o" such that:

        O is instantiated in view of $S \leftrightarrow S$ accepts "o"

(2)   For each basic action type A there is a sentence type "a" such that:

        S does $A \leftrightarrow S$ wants "a".

(3)   The set of things S accepts conforms to principles of consistency, simplicity, and conservatism.

(4) S accepts "$p \rightarrow q$" iff S is disposed to accept "q" should hi come to accept "p".

(5)   (S wants "q" and S accepts "p $\rightarrow$ q") $\rightarrow S$ wants "p".

That this is not in itself adequate is clear; Horwich admits that it is merely a "crude first approximation". However, I am afraid that it is not easy to see how further, better approximations should go. In many cases a speaker would *not* accept "o" despite the fact that O is instantiated in her view; and we must adjust the definition to make room for this. And I can see only our two good old ways: *either* we can replace *S accepts "o"* by *S tends* (*has the disposition*) *to accept "o"*, *or* we can replace it by *S usually accepts "o"*. And, as explained above, I find neither viable.

Thus the inevitable step, in my view, is to *go normative*: to say *S ought to accept "o"* or *it is correct for S to accept "o"*. I can see no other reasonable way to round-off Paul's theory of acceptance; and hence I see Paul as striding the inferentialist path, only not as yet paying due attention to the signs of where it leads.

Conclusion

I think that semantics was long rather confused with respect to its subject matter: what are we studying when we study meanings? Where are the things which hang on our words and make them into the miraculously useful tools that so effectively help us become what we are –

rational, thinking, and communicating beings? Are they of the kind of the tangible things we encounter within the world, or denizens of some Platonist heaven, or perhaps some inhabitants of our minds? As each of these answers appeared to have some counter-intuitive consequences, none of them has been universally accepted as satisfactory. The choice of mentalism by the majority of semanticists was, I think, a natural outcome: the inside of the mind is itself so enigmatic and so multifaceted that it should be able to ingest the addition of another enigmatic kind: meanings. However, as explained above, my opinion is that mentalism is untenable.

I think that behaviorist, use-theoretic approaches to meaning popularized especially by Quine (and in a sense also by Wittgenstein, who was, to be sure, no behaviorist himself) served as a useful antidote to this: *nothing*, we can imagine Quine urging against the mentalists, just as his empiricist predecessors did against their rationalist opponents, *is in meaning that was not in behavior before*. According to the use-theories, semantics is basically not a theory of mental contents, but rather of certain behavioral patterns.

Although this insight may be truly liberating, it can still be dangerously misleading. For what are those patterns, the study of which it assigns to semanticists? Are they directly patterns of what members of the relevant linguistic community do or what goes on within their brains? This surely cannot be the case: I do not think there really is an empirically recordable regular co-occurrence of uttering/assenting to/thinking the sentence/thought *it rains and it is getting dark* and uttering/assenting to/thinking the sentence/thought *it rains*. The two sentences/thoughts *are* intimately related, but not because of any statistically significant tie between events involving the former and those involving the latter. Rather, I think, it is because the latter is *correctly inferable* from the former.

Hence, though I am very much in sympathy with Paul Horwich's anti-mentalist, use-oriented theory of meaning (and I think he did very much to help us see that many of the ideas we formerly held about truth and meaning were mere prejudices!), I doubt that his notion of acceptance properties, on which his theory ultimately rests, can be made adequate without becoming more inferentialist, in particular without accommodating the crucial role of *rules*, and especially *inferential rules*, within semantics.

References

Brandom, R. (1994): *Making It Explicit*, Harvard UP, Cambridge (Mass.).
Davidson, D. (1990): 'The Structure and Contents of Truth', *Journal of Philosophy* 87, 279-328.
Horwich, P. (1998a): *Truth (second edition)*, Clarendon Press, Oxford.
Horwich, P. (1998b): *Meaning*, Clarendon Press, Oxford.
Lance, M. N. (1998): 'Some Reflections on the Sport of Language', *Philosophical Perspectives* 12, 219-240.
Peregrin, J. (2006a): 'Developing Sellars' semantic legacy: meaning as a role', M. Lance and P. Wolf (ed.): *The Self-Correcting Enterprise: Essays on Wilfrid Sellars*, Rodopi, Amsterdam, to appear.

Peregrin, J. (2006b): 'Brandom and Davidson: what do we need to account for thinking and agency?', *Philosophica*, to appear.

Peregrin, J. (2007): 'The use-theory of meaning and the rules of our language games', in K. Turner (ed.): *Making semantics pragmatic*, Elsevier, London, to appear.

Quine, W.V.O. (1969): *Ontological Relativity and Other Essays*, Columbia University Press, New York.