Jaroslav Peregrin

DEVELOPING SELLARS'S SEMANTIC LEGACY: MEANING AS A ROLE*

ABSTRACT. Wilfrid Sellars's analysis of the concept of meaning led, in effect, to the conclusion that the meaning of an expression is its inferential role. This view is often challenged by the claim that inference is a matter of syntax and syntax can never yield us semantics. I argue that this challenge is based on the confusion of two senses of "syntax"; and I try to throw some new light on the concept of inferential role. My conclusion is that the Sellarsian view that something can have meaning only if it is subject to inferences is viable, and that *inferential role* is a plausible explication of *meaning*. However I also argue, *pace* Sellars, that the inferential nature of meaning does not prevent us from engaging in the enterprise of Carnapian formal semantics.

1. Sellars on Meaning and Abstract Entities

In his seminal article "Abstract entities," Wilfrid Sellars claims that the meaning of a linguistic term is best construed as sort of role played by the term within the drama of language (and that, consequently, abstract entities, which are normally seen as expressed by such terms are linguistic in nature). "Redness," writes Sellars:

> [A]s a first approximation, is the word •red• construed as a linguistic kind or sort which is capable of realization or embodiment in different linguistic materials, e.g. *red*, *rot*, and *rouge*, to become the English word 'red', the German word 'rot', and the French word 'rouge' . . . Thus •red• is a type which is shared by the English word 'red', the German word 'rot', and the French word *'rouge'*. (Sellars 1963, p. 267)

(Here the asterisks are supposed to form names referring to expressions as mere shapes, quotation marks form names referring to expressions qua

* Work on this paper was supported by research grant no. A9009204 of the Grant Agency of the Academy of Sciences of the Czech Republic.

In: M.P. Wolf and M.N. Lance (eds.), *The Self-Correcting Enterprise: Essays on Wilfrid Sellars (Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 92), pp. 257-274. Amsterdam/NewYork: Rodopi, 2006.

expressions of a particular language, and the names formed by the dots refer to the roles instantiable by different expressions in different languages.)

Crucial for this conception is the distinction between what can be called "generics" and "true universals." When I say "The lion is tawny" and use "the lion" as expressive of the generic, my claim is reasonable and true. On the other hand, should I construe it as standing for the universal ("Lionhood is tawny"), what I say would be false (if not utterly meaningless) – a universal surely is not tawny (see Sellars 1963, p. 627). Now what Sellars claims is that universals are in fact a kind of "second-order" generics – they are generics whose instances are linguistic expressions. Thus the universal *lion* is in fact the role played by the word 'lion' in English (and by other words in other languages).

Hence Sellars may be seen as distinguishing between *generalization*, which is manifested by the step from particular statements to a general one ("All lions are . . . " resp. "The lion is . . . ") and *abstraction*, which is a matter of a *further* step to a particular statement about a universal, which is the notorious "e pluribus unum" ("Lionhood is . . . "). And while philosophers had traditionally wanted to analyze the abstraction as a matter of a peculiar knack of human mind to put many things into a nutshell of one, Sellars proposes to analyze it as a capacity into which the capacity of generalization mutates when language sets in. For language itself provides us with a certain "one over many" (one word 'lion'[1] over many particular lions); and when we apply our capacity of generalization to it, we find the abstract universal (lionhood) as the one over the many words playing the role assumed by 'lion' in English within other – actual or possible – languages. Hence the peculiar ability of our minds to abstract appears as secondary to its ability to (generalize and) "language about languagings" (1974, p. 425). This leads Sellars to the conclusion that the senses of our expressions (or their intensions) are best seen as certain *roles* of the expressions.

What is a role? Sellars invokes probably the most frequented simile within modern philosophy of language and compares language to chess.[2] It is rather clear what the role of, say, pawn is: it is delimited by the rules saying how the pawn moves, how it interferes with other pieces, etc. Sellars points out that to be a pawn usually involves *two kinds* of criteria: descriptive criteria ("to look

---

[1] The word as such is, of course, itself an abstract item, one shape over many particular inscriptions or sounds. However, it seems that Sellars takes this kind of abstraction as somehow straightforward and unproblematic.

[2] In this way he queues up into the long line of linguists and philosophers already containing Husserl, Frege, de Saussure, Wittgenstein, Jakobson, etc.

thus and so") and prescriptive criteria ("to be correctly moved thus and so"). Now the role in his sense is what remains if we subtract the descriptive part.

Sellars (1963) invites us to imagine an alternative game, "Tess," which is played in Texas and which is just like chess, save for the fact that the various kinds of pieces are replaced by various types of cars and the fields of the chessboards are replaced by counties. The same role which is assumed within chess by our pawn is assumed in Tess by, say, their Volkswagen; and the role is what these two utterly dissimilar objects share.[3]

Now if we accept that language is something like a "rule-governed activity" (as urged not only by Sellars, but of course also by the later Wittgenstein and his followers), the idea that also meanings might be sort of roles comes quite naturally. But what kind of rules is constitutive of this kind of roles? There appears to be an abundance of linguistic rules: we have, for example, the rule that it is correct to change 'read' in the third person singular to 'reads', or the rule that claiming that somebody stole something commits the claimer also to the claim that the very somebody is a thief; or the rule that it is impolite to say "Ciao" to the Queen of England. Which of them, if any, represent the ones that are responsible for meaning? Sellars indicates (see esp. 1953; 1992) that the rules which are crucial in this respect are the *rules of inference* (Sellars 1992 calls them "consequential rules," see p. 114), i.e., the rules exemplified by the middle one of the three examples.

2. What Is an Inferential Rule?

Inferential rules are often taken as kind of prescriptions for passing from an antecedent to a consequent: perhaps from asserting the former to asserting the latter, or from believing the one to believing the other. But this is clearly ill-conceived: when we say that it is correct to infer *Fido is a mammal* from *Fido is a dog*, then we surely do not want to say that whoever asserts the former is obliged to assert the latter – if we were always to assert all consequences of everything we assert, we could never stop making assertions. And also we surely cannot want to say that one is obliged to believe all the consequences of what she believes – we do not control our own beliefs in such a way that this kind of obligation would make more sense than, say, the obligation to have luck.

---

[3] These themes are further elaborated especially in Sellars (1950; 1954; 1963; 1969; 1974; 1992; 1997). See also Harman (1970), Marras (1973) and Seibt (1990).

To understand the Sellarsian notion of inference, we consider his distinction between what he calls the "rules of doing" and the "rules of criticizing" (1999, p. 76). Inferential rules are basically the latter – they do not tell us *what to do*, but rather what to despise – and hence are really more the rules of what *not* to do. They delimit a space of what is approvable: if you assert that Fido is a dog, then you should not deny that Fido is a mammal; and if you do deny it, you are a legitimate target of criticism.

Hence the rules of inference – which are constitutive of conceptual content – are not a matter of regularities of linguistic conduct, of "passing over" from some utterances or beliefs to other ones (whatever the "passing over" might mean). They are a matter of what Brandom (1994) calls *normative attitudes*: attitudes to one's own and others' linguistic utterances, of "languaging about languagings." This is an important thought: conceptual content is not a matter of regularities of linguistic behavior, but rather of (regularities of) a "*meta*linguistic behavior" (see Sellars's (1950) criticism of "regulism").

This is closely related to the Sellarsian "dialectics" of *ought-to-do* and *ought-to-be*: whereas an *ought-to-do* in a sense directly *is* also an *ought-to-be* (and hence a rule of criticism directly piggybacks on every rule of doing) an *ought-to-be* bears an *ought-to-do* via a kind of a "practical syllogism" (see Sellars 1969): via the conviction that we ought to do what brings about that which ought to be. (But of course that if roles of inference are really rules of what ought *not* to be, they lead us to what *not* to do.)

3. Can We "Get Semantics out of Syntax"?

One of the main reasons why many people would reject the proposal to see semantics as a matter of inferential rules right off stems from the combination of two "received wisdoms":

(i) inferential rules are a matter of syntax; and
(ii) syntax is not enough for semantics.

I think that for each of the two claims there is a sense of 'syntax' in which it is true. The trouble is, however, that the two needed senses are *different*; and it is impossible to take the two claims to be true jointly (without taking the two occurrences of 'syntax' in them to be mere homonyms). And in particular, it is impossible to infer that inferential rules are not enough for semantics.

What does the word 'syntax' mean? In one sense, syntax is the theory of well-formedness: it specifies rules for composing more complex expressions out of simpler ones, thus delimiting the space of all expressions, and especially all sentences, of a language. It is clear that if we understand syntax in this way, (ii) will be true, but (i) not.

But syntax is sometimes understood in a wider sense, as amounting to every property of expressions (and every relation between them) specified without reference to anything else than the shapes of the expressions (this is the sense put into circulation by Carnap).[4] Taken in this way, (i) becomes true, but at the same time the falsity of (ii) ceases to be a matter of course. Let us see why.

Let us consider a list of names, such as a telephone directory. We can say that the names it contains have two kinds of properties: those which they have "by themselves" (and which thus can be "read off" them), and those which they have as a consequence of being names of particular persons (and which are thus not "visible"). Examples of the former properties are *to begin with 'N'* or *to consist of more than ten characters*; an example of the latter is *to be a name of a bald man*. Let us call the former "syntactic," and the latter "semantic."

Now suppose somebody finds out that, just by chance, all and only bald men have names beginning with '*N*'. With this knowledge, it becomes possible to read the property *to be a name of a bald man* off the name – it is enough to check whether it starts with '*N*'. Does this mean that we have managed to change the property in question from a "semantic" to a "syntactic" one? Surely not, we have only found its "syntactic symptom" and thereby a way to "represent" it by a "syntactic" property.

Notice that the same holds if we are not so improbably lucky as to find out that the names of all bald men share such a simple property. We achieve the same by simply making the list of all the relevant names: then it will be again possible to find out whether a name is that of a bald man by simply inspecting its shape – namely by comparing it with a given list of shapes. If our directory is finite, then we can "syntacticize" *any* property in this way; if it is infinite, then perhaps not any, but surely we still can do so with many.

Now it seems to be clear that *correct inferability* is basically a semantic relation – of course that we cannot see whether a sentence is correctly inferable from other sentences by merely inspecting the shapes of the sentences (i.e., without knowing what they mean). However, we can make a

---

[4] Elsewhere (Peregrin 1999) I argued that once we abandon the view that language is a collection of labels stuck to things and see it as a means of an activity, the Carnapian concept of syntax/semantics boundary is no longer available. For then the only perceptible boundary is that between the question which strings are employed (syntax) and the question how they are employed (semantics-cum pragmatics).

list of correct inferential patterns (such as the one assembled by Frege in his *Begriffsschrift* as well as by the host of his followers), and given we know this is a (complete) list of valid patterns, we can recognize inferability simply from the shapes.[5] (In fact, this is what appears to have been behind the famous Hilbert's program: once we are able to show that whenever *A* follows from *X*, it is derivable from it in terms of our rules and *vice versa*, we may reduce truth to provability and hence to a "mechanical manipulation" of symbols.)

This means that inference, though its rules are articulated in syntactic terms, is not syntactic in the other – more important – sense; and this removes the absurdity from taking *inferences* as constitutive of *meanings*.

4. A Digression: Keepers, Joiners, Crashers, and Co.

To throw more light on the nature of roles which we will claim underlie meanings, let us make a digression. Suppose we have an alphabet and consider strings which can be formed by its means. Suppose the space of all such potential strings is divided into two groups: the "good" and the "bad" ones; and we are interested in the ways of building the good strings.

When we toy with putting such strings together, we may come to notice that various letters and strings display specific kinds of "behavior." We may, for example, discover that some letters/strings have the property that when concatenated with any strings whatsoever they always produce bad strings; and we may conceive of calling such rogue strings *crashers*. Other letters/strings may do the same thing when they are concatenated with other strings *in a particular way*: we may, for example call a string an *l-crasher* iff it produces a bad string whenever it is prefixed to a string. Still other strings may have the property that they produce a bad string whenever they are prefixed to *good* strings – we may call such a string *l-killers*. Or we may come to call a string a *joiner* iff it produces a good string if and only if it is put in between two good strings.

Here are some examples of the basic categories of strings we may discover when studying buildability of good strings:

S is a *crasher*: any string containing *S* as a substring is bad
S is a *l-crasher* (*r-crasher*): any string starting (ending) with *S* is bad

[5] Of course that as the number of such patterns is unlimited, the list can be given only "potentially," via a finite list of some basic patterns (axioms) and some derivation rules. A valid proof is then the decomposition of a given inference into a chain of the axiomatic patterns.

S is a *l-killer* (*r-killer*): any string consisting of *S* followed (preceded) by a good string is bad

S is an *l-inverter* (*r-inverter*): a string consisting of *S* followed (preceded) by another string is good iff the other string is bad. (Every *inverter* is obviously an *killer*.)

S is a *l-keeper* (*r-keeper*): any string consisting of *S* followed (preceded) by a good string is good

S is a *l-superkeeper* (*r-superkeeper*): any string consisting of *S* followed (preceded) by a good string is good iff the string is good. (Every *superkeeper* is an *keeper*.)

S is a *joiner*: any string consisting of two good strings with *S* between them is good

S is a *superjoiner*: any string consisting of two strings with *S* between is good iff the two strings are good. (Every *superjoiner* is a *joiner*.)

Consider, for the sake of illustration, what we will call Language 1. The vocabulary is

a, **b**, **c**, **d**, **e**

and the delimitation of the good strings over it is as follows:

(i) **a**, **b** and **c** are good strings;
(ii) if *s* is a good strings, then $\mathbf{d}^\frown s$ (i.e., the symbol **d** followed by the string *s*) is a good string;
(iii) if s_1 and s_2 are good strings, then $s_1{}^\frown\mathbf{e}^\frown s_2$ is a good string;
(iv) no other string is good.

The words of this language can be clearly classified into three categories. First of them is constituted by the letters **a**, **b** and **c**; the second one by the letter **d**, which is an *l-superkeeper*; and the third by the letter **e**, which is a *superjoiner*.

Now suppose we distinguish a further subset of the set of good strings; call its elements "perfect" and call those good strings which are not perfect "imperfect." We can, of course, again start to classify the roles of letters and strings with respect to the new set. Returning to Language 1, imagine that **a** and **b** are perfect strings, while **c** is imperfect. Further imagine that the string $\mathbf{d}^\frown s$ is perfect if and only if *s* is imperfect; and that $s_1{}^\frown\mathbf{e}^\frown s_2$ is perfect iff both s_1 and s_2 are. It is easy to see that from the viewpoint of the delimitation of the perfect strings, then, **d** comes to be an *l-inverter*, whereas **e** is again a *superjoiner*.

Now consider a little more sophisticated kind of language, Language 2, based again on the letters **a**, **b**, **c**, **d**, and **e**. Suppose that a string over this alphabet is good if it is of the shape $s_1{}^\frown\mathbf{e}^\frown s_2$ where the strings which are usable in the position of s_1 and s_2 are defined recursively: **a** is usable, any usable string preceded by **b** is again usable, and any string which is of the shape $s_1{}^\frown\mathbf{c}{}^\frown s_2$ is $s_1{}^\frown\mathbf{d}{}^\frown s_2$ with s_1 and s_2 usable is again usable. Examples of good strings are

aea
bbaea
baebbbbba
acbaeba
acaebbbadbbbba

To characterize the roles of the strings of this language we do not make do with the simple concepts of joiners, inverters etc. We first have to characterize the strings which may surround **e** in a good string – call them **e**-*joinables*. **a** is an **e**-joinable. **b** is what could be called **e**-*joinable-l-keeper*, whereas **c** is what we could think of calling **e**-*joinable-superjoiner*.

Now imagine that the perfect strings of this language are defined in the following way:

aea is perfect;
if $s_1{}^\frown\mathbf{e}{}^\frown s_2$ is perfect, then so is $\mathbf{b}{}^\frown s_1{}^\frown\mathbf{eb}{}^\frown s_2$;
$s{}^\frown\mathbf{cae}{}^\frown s$ is perfect;
if $s_1{}^\frown\mathbf{c}{}^\frown s_2{}^\frown\mathbf{e}{}^\frown s_3$ is perfect, then so is $s_1{}^\frown\mathbf{cb}{}^\frown s_2{}^\frown\mathbf{eb}{}^\frown s_3$;
$s{}^\frown\mathbf{daea}$ is perfect;
if $s_1{}^\frown\mathbf{d}{}^\frown s_2{}^\frown\mathbf{e}{}^\frown s_3$ is perfect, then so is $s_1{}^\frown\mathbf{db}{}^\frown s_2{}^\frown\mathbf{e}{}^\frown s_3{}^\frown\mathbf{c}{}^\frown s_1$;
nothing else is a perfect string.

Again, these stipulations furnish the strings with roles, which are now rather intricate. Call two **e**-joinable strings s_1 and s_2 **e**-*comparable*, if $s_1{}^\frown\mathbf{e}{}^\frown s_2$ is a perfect string. We may say that a string *cancels s from the left in the end position* (in short, is a *s-l-e-canceller*) if a string ending with $\mathbf{c}{}^\frown s$ is **e**-*comparable* to a string without this ending. And we may say that it *shifts s from the left to the beginning* (is a *s-l-b-shifter*), if a string containing $\mathbf{c}{}^\frown s$ is **e**-comparable with a string starting with s and continuing by something **e**-comparable with the original string with s excerpted. Using this terminology, **c** is a **a**-*l-e-canceller* and a **b**-*l-b-shifter*.

5. Meanings as Roles

We have seen that any delimitation of a subset of the set of strings over an alphabet furnishes the strings with certain roles, which are a matter of their "behavior" with respect of the distinguished set and which may be quite complicated. Especially any language and any axiomatic system (which also delimits a set of sentences, namely the set of its theorems) does so. Now the meaning which an axiomatic system confers on an expression of its language can be seen as precisely this kind of role.

Consider formal languages of the kind that we can formulate within the framework of the propositional and the predicate calculus, with good strings being the well-formed formulas. It is, for example, easy to see the usual syntax of the propositional calculus is only a minor variation on our Language 1. We also have three categories of signs: the propositional letters, the *l-superkeeper* $\neg$ and the *superjoiners* $\wedge$, $\vee$ and $\rightarrow$.[6] More generally, taking the set of good strings as the set of well-formed formulas, the roles expressions thus acquire are usually called syntactic categories. (We would not say that, $\wedge$, $\vee$ and $\rightarrow$ are *superjoiners*, but rather that they are *propositional connectives*.)

Now, moreover, take the set of perfect strings to be the set of true sentences (under some interpretation). What, then, do the resulting roles correspond to? Consider conjunction: we have noticed that with respect to the perfect (i.e., true) sentences it acts as a superjoiner. This means that it produces a perfect output iff the two strings which are appended to it from the sides are both perfect, i.e., it maps *perfect* and *perfect* on *perfect*, whereas *perfect* and *imperfect* on *imperfect* etc. However, as *perfect* is *true* and *imperfect* is *false*, we have the function mapping T and T on T and any other combination of the truth-values on F. This is to say that the truth table usually thought of as the meaning of conjunction can be plausibly seen as the recapitulation of the role of a superjoiner. This indicates that to be the conjunction simply *is* to be the superjoiner (with respect to the class of true sentences), to be the negation *is* to be an l-*inverter*, etc. Thus our Language 1 amounts to *nothing else* than the classical propositional calculus.

Now look at language Language 2. If we replace the signs **a**, **b**, **c**, **d** and **e** with **0**, **Succ**, **+**, **×**, and **=**, respectively, we can see that Language 2 constitutes a kind of rudimentary arithmetic. In this case, the roles are much more complicated and they are mutually entangled (so that it is not possible to specify the role of, say, **+** independently of that of **0**), but nevertheless they are still roles in the very sense as before. Hence, we may say, the meaning arithmetic confers on the sign + is the role of a "**0**-*l*-*e*-canceller" and a "**Succ**-l-*b*-shifter." (Expressed in a more prosaic way, this means that to add zero is to add nothing and to add a successor of a number is to produce the successor of the addition of the number.)

---

[6] This is an oversimplification for if we consider the standard form of the syntax, we need brackets; and if we consider the Polish variant, the connectives will not be literally superjoiners but rather something which could be called, say, *l*-bikeepers.

However, as Lakatos (1976) reminds us, this "Euclidean" (may be we should better say "Hilbertian"[7]) way of viewing an axiomatic system (according to which any meanings the signs have are conferred on them by the axioms) is only one possible way; and in fact such systems are often understood in a quite different way, which Lakatos calls "quasi-empirical." From this point of view, such a system is seen as codifying some pre-existing, proto-theoretical knowledge, and hence is responsible to it: if it produces theorems whose interpretation is at odds with what we take for granted within the proto-theory, it is revised.

Let us take a formula of PA. The answer to the question whether it is true can be sought in two quite different ways: we may check whether it is deducible from the axioms; or we may check whether the "informal" mathematical sentence which is taken to be "captured" by it (in view of the fact that the basic vocabulary and grammatical rules of PA more or less copy the basic vocabulary and rules of our pre-formal arithmetical language) is true. If we see the system in the Euclidean way, then a possible discrepancy between the two results need not worry us – for the latter is completely irrelevant for us. However, if we see it in the quasi-empirical way, then the discrepancy might lead us to revise the axioms.[8]

Hence we should qualify the claim we made above: any axiomatic system furnishes the expressions of the underlying language with *Euclidean* meanings. But is it not so that such an Euclidean meaning is restricted to formal systems only, and that the meaning in normal languages is something quite different which is perhaps "incommensurable" with it?

Observe, first, that we can reconstruct, and in some special cases perhaps even replace, parts of natural language by means of formal languages constituted in the axiomatic way. The axiomatic Peano arithmetic is not only a commonly accepted codification of pre-formal arithmetic, but in the eyes of many mathematicians it simply replaces it. This means that Euclidean

---

[7] In a letter to Frege, Hilbert writes: "You write: 'From the truth of the axioms it follows that they do not contradict one another'. I was very interested to read this particular sentence of yours, because for my part, ever since I have been thinking, writing and lecturing about such matters, I have been accustomed to say just the reverse: if the arbitrarily posited axioms are not in mutual contradiction with the totality of their consequences, then they are true – the things defined by the axioms exist." Hence, according to Hilbert, each delimitation of a set of strings constitutes an Euclidean semantics and this is the only semantics we could possibly require; whereas according to Frege, Euclidean semantics of the systems we assemble is not the true semantics, for the true semantics is what expressions of our natural language have and what we cannot freely create.

[8] I have discussed the consequences of these two radically different aspects of logical systems elsewhere (see Peregrin 2001).

meanings can be not only seen as suitable *explications* of the corresponding "natural" meanings, but sometimes even as their *descendants*. What makes this possible?

The Sellarsian answer is that natural languages are themselves sort of Euclidean systems, only with axioms and rules not being explicit in the form of formulas or statements, but implicit within the praxis of using language. Our arithmetical praxis is governed by the rule that adding zero to x means leaving x unchanged just as the axiomatic PA is – though it need not be written anywhere explicitly. Children learn to add in this way by being corrected – the "ought-to-do" imposed on them by their teachers thus becoming their "ought-to-be" (see Sellars 1969). And just like within the Peano arithmetic there is nothing to semantics save the explicit rules (axioms and derivation rules), there is, in natural language, nothing to semantics save the implicit rules of our "language games." Thus, the claim is that every meaning is *a sort of* Euclidean meaning.

This is not to deny that there is an important difference between systems constituted by explicit and hence rigid rules and those governed by implicit and hence flexible ones, and so consequently between the respective "formal" and "natural" meanings. However, let us postpone further discussion of this difference to the concluding section, and let us now accentuate the common Euclidean character of both kinds of meaning: the former are explicitly Euclidean, the latter are implicitly Euclidean and can be explicated in terms of the explicitly Euclidean ones. Given this, we may want to consider the following two general questions:

(1) Is every "formal meaning" a meaning worth its name?
(2) Is every "natural meaning" reconstructible as a "formal" one?

The answer to the first question is clearly *no*. We may have a formal language with a single sentence; or a language without axioms, or such that any sentence is inferable from any other. No such system obviously yields the rich variety of roles characteristic of what we would be willing to call language, and hence furnishes the expressions with what we would be willing to call meanings. (The last example covers the case of Prior's 1960/61, **tonk**, which is a grand example of the fact that nontrivial axioms need not lead to a nontrivial semantics. This is important, but it does not show, *pace* Prior, that axioms or inference rules are generally short of conferring meaning. See Peregrin 2001, §8.5 for more details.)

The other question is more intricate; and *prima facie* there seem to be at least two counterexamples. First, meanings of empirical terms; and, second, meanings of some common non-empirical terms which do not seem to be conferrable by means of rules. Let us deal with the two topics in turn.

6. Empirical Terms

Of course that Sellars does not claim that the meaning of an empirical term is a matter of nothing else than inferences in the usual sense on the term. What he claims (1974, pp. 423-424) is that, in general, the ways in which expressions are used (and hence their meanings) are a matter of three kinds of regularities or patterns of linguistic activity[9]:

> (1) Language Entry Transitions: The speaker responds to objects in perceptual situations, and in certain states in himself, with appropriate linguistic activity.
> (2) Intra-linguistic moves: The speaker's linguistic conceptual episodes tend to occur in patterns of valid inference (theoretical and practical), and tend not to occur in patterns which violate logical principles.
> (3) Language Exit Transitions: The speaker responds to such linguistic conceptual episodes as "I will now raise my hand" with an upward motion of the hand, etc.

Hence if we want to restrict the term 'inference' to the language-language moves, we cannot claim that meaning is in general the *inferential* role, but rather a role with respect to more inclusive kinds of rules. (The term inference is sometime stretched to cover even Sellars's (1) and (3), which then allows for equating meaning and inferential role quite generally, but this appears to be at odds with how the term 'inference' is most often used.)

We must, however, notice that it is only the intra-linguistic moves which we can nontrivially capture when we are doing a theory of semantics. The trouble with the other two is that they consist of a language-part and a world-part, and when we want to talk about the latter, we have no choice but to put it in words, by which it becomes identical with the former. We can practically teach somebody to recognize the circumstances in which it is appropriate to assert "A dog has just run by," but when we want to capture them within a theory, we end up with the disquotational claim "It is appropriate to assert 'A dog has just run by' iff a dog has just run by." As Wittgenstein (1977, p. 27) remarked "The limit of language manifests itself in the impossibility of describing the fact that corresponds to (is the translation of) a sentence without simply repeating the sentence." True, we need not use the *very same* sentence to describe the circumstances, but if we do not do so, then the ensuing

---

[9] A more fully elaborated (if not overelaborated) version of this classification is presented by Sellars (1992, p. 114).

nontriviality can be seen as a matter of the intra-linguistic relationship between the two sentences. This means that insofar we take semantics to deal exclusively with *language*, we cannot expect that it completely elucidates the meanings of non-empirical terms: part of their meanings is a matter crucially depends on their relationship to the extralinguistic world. After all, the delimitation of the meaning of "dog," i.e., of what a dog is, should not be expected of a semanticist, but rather of a biologist.

The situation is, however, different, as far as non-empirical words are concerned. Their meaning *can* be usually seen as a matter of inferences alone. There is nothing more to the meaning of 'and' than the inferences which govern the term: the inferences from *X and Y* to both *X* and *Y*, from *X* and *Y* to *X and Y* and perhaps some more.[10] This means that while the meaning of an expression is, from the current viewpoint, the role of the expression within the relevant "language game," the theoretically articulable part of the meaning (which may in some cases, like that of 'and', wholly exhaust it) consists in its *inferential role.*

Hence, we may say, *syntax, in the form of inferential rules in some cases can alone be sufficient for semantics*. (The meaning of an non-empirical term, such as 'and' can be directly seen as consisting in its inferential role.) And there is an asymmetry between inferences (the language-language kind of moves) and the other two kinds of moves in this respect: while inferences alone are (sometimes) capable of conferring a meaning, neither the world-language moves, nor the language-world moves alone are. An expression cannot become truly meaningful by being subjected to merely language entry rules, i.e., by becoming a correct reaction to something extralinguistic: it must *also* be subject to inferences.

The reason is that language is a holistic, interconnected structure – an expression which would not be connected, by the rules of language, to other expressions would be the notorious "idle wheel." (See Wittgenstein 1953, §271/p. 95e.) Just like there cannot be a chess piece without a potential interaction with all other pieces, there cannot be an expression which would be connected only to something extralinguistic and not to other expressions. And it is only inferences which are capable of interlinking all its elements. Hence *syntax, in the form of inferential rules, is in all cases essential for semantics*: for something can have meaning only if it is subject to inferences, i.e., if it is in inferential relations to other expressions.

---

[10] The proviso is here because the meaning of the English 'and' does not utterly coincide with that of the classical conjunction operator, it, e.g. often behaves non-symmetrically.

7. Classical and "Standard" Semantics

There seem to be some limits to what kinds of meaning can be conferred by rules; and it is sometimes claimed that we also need meanings of a kind that is beyond the reach of semantics generated by inferences.

The fact is that if we understand the term inference as we did so far, we even do not get classical propositional connectives in the inferential way. For though there is an inferential pattern which inferentially characterizes the classical conjunction, there is none which would so characterize the classical disjunction, negation or implication. For though we can inferentially express that the conjunction is true if one of its conjuncts is true, we cannot express that it is false whenever both the disjuncts are false.[11]

However, as I have pointed out elsewhere (see Peregrin forthcoming), this obstacle can be solved by reconsidering the concept of inferential pattern, namely by admitting that an inferential pattern can be read as providing a kind of an exhaustive listing. Claiming

$$A \vdash A \vee B$$
$$B \vdash A \vee B$$

we claim that $A \vee B$ follows from A and from B; but presenting the two inferences as a self-contained *pattern* indicates that this list of premises be *exhaustive*; i.e., that there is nothing else which would entail $A \vee B$ and at the same time be entailed by both A and B.

This can be accounted for by letting inferential patterns incorporate an implicit exhaustivity claim: if I give an enumeration of premises each entailing a given conclusion, then we assume that the list is to be exhaustive. (Why? For this is how we usually understand enumeration: if I say "I own a VW Beetle and a Pontiac," the normal understanding is that these are *all* my cars.) As Koslow (1992) has shown, the admission of such "extremality conditions" into the concept of inferential pattern leads directly to an inferential capturing of all of classical logic.

And as I have indicated elsewhere (see Peregrin forthcoming), it could even lead to the inferential accommodation of what is usually called the *standard* semantics of higher-order logics (as contrasted to the Henkin one, see, e.g., Shapiro 1991). Its peculiarity can be located into the incompleteness

[11] We must not think that the completeness proof for the classical logic shows that the usual truth-functional meanings of the logical operators are conferred on them by the standard axioms: the axioms are compatible even with certain non-truth-functional interpretations.

of arithmetic – whichever axioms and rules satisfiable by the standard number sequence we accept, they will be also satisfied by some nonstandard ones. (The existence of the non-standard interpretations of higher-order logics then can be seen as a matter of the fact that the corresponding semantic structures are bound to incorporate the natural numbers sequence pattern.) Now if we construe the "exhaustivity condition" which we take to be implicit in inferential patterns broadly enough, we might even think of it as blocking the nonstandard models: the standard axiomatization of Peano arithmetic says that zero is a number and the successor of a number is a number, and the exhaustivity says that nothing else is a number, hence that we intend only the *minimal* model.

Of course the hints I have just given cannot *establish* that any kind of semantics can be seen as ultimately grounded in inferences (They are clearly in the need of further elaboration, which is, however, beyond the scope of the paper and which I hope to produce in the future.) What I hope they *can* do is shatter the preconception that some common varieties of semantics are, as a matter of principle, unavailable for the inferentialist. This means that nothing appears to stand in the way of our embracing Sellars's *role semantics*, i.e., the view that meanings are, as a matter of principle, roles of the expressions which are said to express them.

8. "Natural" and "Formal" Meanings

To conclude, let us return to the previous claim that natural languages are themselves one sort of Euclidean system. Of course the difference between a game like chess or a formal language, the rules of which are explicit, and our natural language games with their rules implicit in our linguistic conduct is not unimportant. In fact, Wittgenstein's *Philosophical Investigations* (1953) can be seen to a great deal as wrestling with the question of what it means for a rule to be implicit in behavior and what it takes to follow such a rule. But precisely the fact that it is possible to explicate parts of natural language in terms of formal systems constituted by means of explicit rules gives further credence to the conviction that the rules *are* implicitly constitutive of our language games.[12]

---

[12] Brandom (1994) claims that this kind of making implicit rules explicit is the crucial force behind the development of our language, behind logic, and in general behind our reason.

As we have also already noted, in the typical case the relation between a natural meaning and the corresponding formal one is that between an *explicandum* and its *explicatum*. (Only in cases of those parts of natural language which do not deal with empirical reality, like the language of arithmetic, might it be possible to directly *replace* the former by the latter.) Natural meanings are *empirical phenomena*; we do not voluntarily constitute them, we *encounter* them.[13] Formal meanings are, in contrast to this, directly of our making – constituted by our deliberate definitions. However, just because of this it is useful to explicate the former by the latter (which then results in the Lakatosian "quasi-empirical" view of the latter).

As I have discussed in greater detail elsewhere (Peregrin 2000; 2001, Chapter 9), the way we humans often approach complicated empirical phenomena consists in isolating their structure which then can be studied by non-empirical, i.e., mathematical means. Thus we make mathematical models of the weather, of economic developments or of geological layers; and we make also such models of language. The point of this kind of *mathematization*, which is characteristic of a great deal of modern natural science, is the resulting beneficial division of labor: the task of the mathematician is to study the model, whereas that of the natural scientist is to build it and see to it that it remains adequate.

I think that the situation with logic and semantics is quite similar: as soon as Boole, Frege and others achieved what they saw as the basic logical structure of natural language, Hilbert and others subjected the structure to mathematical treatment, giving birth to what has come to be called *mathematical logic*. But the study of mathematical structures cannot tell us anything about natural phenomena *unless we see to it that the structures can reasonably be seen as structures* of *these phenomena*. (To avoid misunderstanding, recall that the rules articulated within logic do not merely reflect regularities of general linguistic traffic. They reflect the regularities of *normative attitudes* of participants in the traffic and thereby the *rules* of language.)

What is fundamentally important is that this kind of mathematization always involves (smaller or greater) idealization: as the explicatum (a natural meaning) and the explicandum (a formal one) are entities of different "orders,"

[13] There is an important difference between two possible modes of such an encounter – we can either view them "from outside" of language, so to speak, where they appear as something directly on a par of other phenomena of the world, or "from inside" of it, where they appear rather as "transcendental" presuppositions of our dealing with the world. But this distinction is not important for us at the moment.

passing from the former to the latter always means a kind of a "bettering": we lead sharp boundaries where there are merely fuzzy ones, we extrapolate and round off, and sometimes we even try to repair or upgrade.

All of this means that, in contrast to Sellars, I do not believe that the inferential nature of meaning would prevent us from engaging in the enterprise of Carnapian formal semantics. We only have to realize that the set-theoretical denotations with which the systems of formal semantics work are not to be regarded as representations of entities on which expressions of natural language really stuck, but rather as recapitulations of the inferential roles of the expressions (see Peregrin 2001; 2003).

Jaroslav Peregrin
Dept. of Logic, Institute of Philosophy
Academy of Sciences of the Czech Republic
Jilská 1, 110 00 Prague
Czech Republic
www.cuni.cz/~peregrin
e-mail: jarda@peregrin.cz

REFERENCES

Brandom, R. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.

Harman, G. (1970). Sellars's Semantics. *Philosophical Review* **79**, 404-419.

Koslow, A. (1992). *A Structuralist Theory of Logic*. Cambridge: Cambridge University Press.

Lakatos, I. (1976). A Renaissance of Empiricism in the Recent Philosophy of Mathematics? In: I. Lakatos, J. Worrall, and G. Currie (eds.), *Mathematics, Science and Epistemology: Philosophical Papers*, vol. 2, pp. 24-42. Cambridge: Cambridge University Press. Reprinted in Tymoczko (1998), pp. 29-48.

Marras, A. (1973). Sellars on Thought and Language. *Nous* **7**, 152-163.

Peregrin, J. (1999). The Pragmatization of Semantics. In: K. Turner (ed.), *The Semantics/Pragmatics Interface from Different Points of View*, pp. 419-442. North-Hollard: Elsevier Pub.

Peregrin, J. (2000). The "Natural" and the "Formal." *Journal of Philosophical Logic* **29**, 75-101.

Peregrin, J. (2001). *Meaning and Structure*. Aldershot: Ashgate Pub.

Peregrin, J. (2003). Pragmatismo e Semantica [Pragmatism and Semantics]. In: P. Ghirardelli, Jr. and J. Zabatiero (eds.), *Revista Filosofia Americana*. CEFA & FTSA. English original available at *http://jarda.peregrin.cz/mybibl/PDFTxt/482.pdf*.

Peregrin, J. (forthcoming). Semantics as Based on Inference. In: J. van Bentham, G. Heinzmann, M. Rebuschi, and H. Visser (eds.), *The Age Of Alternative Logics*. Dordrecht: Kluwer.

Prior, A. N. (1960/61). Roundabout Inference Ticket. *Analysis* **21**, 38-39.

Seibt, J. (1990). *Properties and Processes (A Synoptic Study of Wilfrid Sellars's Nominalism)*. Atascadero, CA: Ridgeview Pub.

Sellars, W. (1950). Language, Rules and Behavior. In: S. Hook (ed.), *John Dewey: Philosopher of Science and Freedom*, pp. 289-315. New York: Diall Press.

Sellars, W. (1953). Inference and Meaning. *Mind* **62**, 313-338.

Sellars, W. (1954). Some Reflections on Language Games. *Philosophy of Science* **21**, 204-228.

Sellars, W. (1963). Abstract Entities. *Review of Metaphysics* **16**, 627-671.

Sellars, W. (1969). Language as Thought and as Communication. *Philosophy and Phenomenological Research* **29**, 506-527.

Sellars, W. (1974). Meaning as Functional Classification. *Synthese* **27**, 417-437.

Sellars, W. (1992). *Science and Metaphysics*. Atascadero, CA: Ridgeview Pub.

Sellars, W. (1997). *Empiricism and the Philosophy of Mind* (with an introduction by Richard Rorty and a Study Guide by Robert Brandom). Cambridge, MA: Harvard University Press.

Shapiro, S. (1991). *Foundations without Foundationalism*. Oxford: Clarendon Press.

Tymoczko, T., ed. (1998). *New Directions in the Philosophy of Mathematics. Revised edition*. Princeton, NJ: Princeton University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Basil Blackwell Pub.

Wittgenstein, L. (1977). *Vermischte Bemerkungen.* Edited by G.H. von Wright. Frankfurt: Suhrkamp. English translation: *Culture and Value* (Oxford: Basil Blackwell Pub., 1980).