DOI: 10.59277/PL.2025.28.01

DOES A TARSKIAN THEORY OF TRUTH OFFER A THEORY OF MEANING? A SELLARSIAN-TYPE EVALUATION AND CRITIQUE OF DONALD DAVIDSON'S TRUTH-CONDITIONAL SEMANTICS*

MIRCEA DUMITRU

University of Bucharest
Romanian Academy

Abstract. The paper examines and assesses how Davidson could be fixed through Sellars. I begin by presenting Donald Davidson's truth-conditional semantics for a natural language, viz. the program according to which the meaning of a language is to be given by a Tarskian truth-theory for that language. Against this background, I build a scenario in which a competent logician can give a truth-theory for sentences of a language that he/she cannot speak/read/understand without thereby giving/knowing the meaning of the sentences that he/she cannot comprehend. The logician knows that the sentences in the unknown (for him/her) language are true but, nevertheless, he/she does not know what they mean. In order to fix this drawback of the Davidsonian truth-conditional based theory of meaning, I present the main elements of Sellars' subtle views on meaning and truth, pointing at how the latter can circumvent the problems with the extensional Tarskian truth-conditional approach put forward by Davidson.

Keywords: Donald Davidson's truth-conditional semantics, meaning, Wilfrid Sellars' approach to meaning and truth, Tarskian truth-theory.

1. INTRODUCTORY REMARKS

In some quarters of philosophy, there is almost unanimous agreement that a theory of meaning for a natural language can be accepted only if it satisfies the following conditions: (i) the theory must give the meaning (semantic interpretation of) every sentence of the language; (ii) the theory must satisfy a corresponding principle of compositionality, i.e. it must show how the meaning of a sentence depends on the meanings of the parts of that sentence plus its logical structure[1]; (iii) the theory must show how (i) and (ii) work in a testable form.

Once this first step has been taken, the next one is not unproblematic, since it is far from clear how we should answer these questions. What form should a theory of meaning capable of satisfying the above three conditions simultaneously take?

---

* This paper was originally published in Willem A. deVries and Marc A. Joseph (eds.), Sellars and Davidson in Dialogue: Truths Meanings, and Minds. London: Routledge, 2025, 56–79.

[1] For a new ground-breaking approach to compositionality, see Fine.

And if such a theory is in general constructible, what criteria for recognizing the satisfaction of its performance do we need in order to know that we are in the presence of such a successful theory?

It is worth noting that any successful theory of meaning for a natural language will *ipso facto* give an explanation of why humans can learn that language. In other words, the theory will also show us how it is possible for a speaker to master the language, i.e., to produce and understand sentences with meaning whose number is infinite, given that what the speaker has learned, used and interpreted is a finite vocabulary and a finite set of rules for composing the lexicon of the language. Another way of expressing the same fact about a successful theory of meaning is this: what the speaker of a natural language needs to know in order to be able to produce and interpret meaningful sentences is something that can be rendered in a finite form, whereas the result of the speaker's use of this finite knowledge is that he/she can produce and interpret an infinite multitude of sentences.

Among the theories of meaning in the relatively recent literature in the field, the one proposed by Donald Davidson in several of his studies has held the attention of a large number of authors and has greatly influenced discussions in the semantics of natural languages. Simplifying things in order to make complex and subtle logical-semantic issues more accessible, Davidson's idea is that the kind of theory which fulfills all the aforementioned conditions for a good theory of meaning for a natural language *L* is a Tarski-type truth theory for that language *L*, a theory which is extended and adapted in a suitable way to make room also for the truth-conditional semantic treatment of the context-dependent structures and intensional constructions of that language. The need for the extension and adaptation of the Tarski-type theory of truth stems, of course, from the fact that Tarski's original theory has extensional formalized languages as its scope, i.e., those languages whose semantic interpretation does not exhibit the phenomenon of context-dependence or that of intensionality; Tarski's approach is aimed only at those languages in which "the sense of every expression is uniquely determined by its form"[2].

In a first provisional and approximate evaluation of Davidson's proposal, it should be noted that the success of this program, which aims to give a theory of meaning for a natural language through the Tarskian theory of truth for that language, depends decisively on the solution of problems whose source lies outside the program, but also of difficulties inherent in the program as such. By the former, I have in mind difficulties arising from the attempt to "tame" the intensional features of a natural language (propositional attitude attributions, modal constructions, indirect discourse) within the confines of an extensional framework, the only one recognized by Tarski in dealing with semantic problems. By the difficulties inherent in the program, I think of the problems faced by the project when it has to give meanings to types of sentences that do not have truth-conditional conditions (such as imperatives or questions), but, in order for the Davidson-Tarski program to work, would nevertheless have to find a place in a truth-conditional semantics.

The way in which I shall examine and evaluate Davidson's proposal will be articulated in a fourpartite structure. First, I shall discuss the extent to which

[2] Tarski, "Concept of Truth" 166.

Davidson's proposal that a theory of meaning be produced by means of a Tarskian theory of truth is compatible with the framework-presuppositions of Tarski's program of constructing a definition of truth for formalized languages. Second, I shall analyze how Davidson criticizes alternative theories of meaning, thereby preparing the ground for his own. Thirdly, I shall present and discuss an argument of my own, which is directed against a simplistic formulation of the idea that whenever we give a Tarskian definition of truth for utterances of a language, we thereby give meaning, interpretation, or understanding to those utterances. And fourthly and lastly, I shall explore how Sellars' profound and insightful philosophy of language can be put to bear on those issues.

The resources that Davidson's theory can bring into play to counter those pieces of criticism allow us to draw a distinction which, although it appears in a different context, may nevertheless give an account of the problem that my argument poses for Davidson's program. However, the dialectic of this argumentative move seems to indicate that the main assumption of Davidson's program must undergo a process of further elaboration and refinement, and the evaluation of *this* new re-folding move by the Davidsonians suggests that their extensionalist truth-conditional approach lacks certain virtues which, on the other hand, are present in a competing type of intensionalist approach to the problem of meaning.

2. DAVIDSON'S PROPOSAL

In order to understand what Davidson's proposal consists in, let us briefly present the main elements of Tarski's theory of truth.

Tarski shows how to construct a formally correct and materially adequate definition for the concept of a true sentence of a formal language *L*[3]. It is easy to see that the extension of the term "true sentence of a formal language *L*" depends on the language *L* we are considering. What we are defining, then, is not truth as such, but truth as a predicate that applies to the sentences of a particular language. The language for which the concept of a true sentence of a formal language *L* can be defined must be a formalized language whose characteristics are well known from the literature on the subject.

The crucial aspect of Tarski's method of defining truth is a distinction underlying the whole technical construction, which ultimately yields the sought-after definition of the concept of a true sentence. This is the distinction between an object-language *L*, for whose sentences the truth-concept is defined, and the corresponding metalanguage *L'*, in which the definition is constructed.

The definition can be correctly constructed only if the metalanguage possesses means of expression that are stronger than those of the object-language for whose sentences we define "true sentence in (language) *L*". More precisely, *L'* must contain names that denote all the component parts of *L*, as well as the operations for constituting well-formed expressions of *L*: the primitive expressions together with

[3] Cf. Tarski, "Concept of Truth"; Tarski, "Semantic Conception".

the syntactic operations that are used to construct sentences of L from its primitive stock of names, variables, predicate-letters and functional signs.

Although the generalized method of defining truth for natural languages does not necessarily require that the object-language be a (proper) subpart of the metalanguage – since we want to leave open the possibility of defining the concept of truth for a natural language in another natural language, distinct from the former – the requirement that the metalanguage has the expressive power to denote all the expressions of the corresponding object-language, together with the syntactic ways of composing the terms and sentences of the object-language, must nevertheless be respected. Thus, the expressive power of L' must be at least as strong as that of L.

The syntax of the first-order language with identity L is determined by two recursive definitions: one for *terms*, and one for *formulas*. The *closed sentences* or *formulas* of the language are defined in the familiar way, as the smallest set of utterances closed with respect to the well-formedness rules of the language and in which all variables are bounded by quantifiers.

The main technical obstacle to defining the concept of a true sentence in this language is that sentences are constructed from syntactic categories, such as names, predicate-letters and variables (or pronouns, which are the counterparts of variables in natural languages), none of which can be properly counted as true or false[4]. What is needed, therefore, is a way of defining the truth for a sentence of L in terms of other semantic notions, which can be applied, in a suitable way, to the syntactic categories whose composition results in that sentence.

To this end, Tarski uses the semantic concepts of *denotation* and *satisfiability*. A name denotes the object it stands for, and objects satisfy predicates (in the interpretation of that language). In order to define a truth-predicate, for the propositions of L, we need an interpretation for the propositions of that language. The basic idea on which Tarski's semantics rests is that truth must be characterized by means of assignments of objects, which are placed in certain positions in infinite strings of objects, to variables, which appear in formulas in positions corresponding to the objects in the infinite strings of objects. Since the individual-variables are infinite in number and are ordered in infinite strings, the objects must also be thought of as ordered in infinite strings, which are assigned to the variables of the language.

Variables are therefore to be regarded as a kind of temporary names, occurring in a certain fixed order, of the objects assigned to them. An open formula becomes true/false relative to an assignment of objects to its free variables: since the objects belong to infinite strings of objects, these strings satisfy or do not satisfy the open formula. Tarski gives an inductive characterization of the denotation relation, and then, by means of the concept of satisfaction, defines the concept of truth of an open formula, relative to an assignment of strings of objects to the free variables which occur in the formula. Finally, a closed sentence is (absolutely) true if and only if the formula is satisfied by every assignment (or, according to alternative versions of Tarski's construction, by some assignments or, still, by the empty assignment) and is false if and only if it is not satisfied by any.

---

[4] Cf. Forbes 14–20.

Tarski's *T-convention* plays a decisive role in Davidson's theory. Convention T requires that a definition of truth is considered adequate if all instances of the biconditional

$$(T)\ S \text{ is } \textit{true-in-}L \leftrightarrow P$$

are logical consequences of the recursive definition of truth. In (T), S is a sentence of L, and P is substituted by a sentence of the metalanguage L' which *has the same meaning* as S.

Let us pause for a moment and note that this formulation of the adequacy of the definition of truth resorts to the concept of meaning. The condition in question goes back in fact to the idea that in the theory of truth based on this definition, every T-sentence is provable, and that a T-sentence is a biconditional whose left-hand member *mentions* a sentence that *has the same meaning* as the sentence *used* in the right-hand member of the biconditional. This means, then, that Tarski presupposes the concept of meaning to be given in order to define the concept of truth.

From here, there is just one step to Davidson's theory. It is very plausible to hold that a sentence makes sense if and only if there is a condition whose satisfaction determines the truth of that sentence and whose non-satisfaction makes the proposition false. Davidson proposes to identify the meaning of a sentence with this condition, or – Davidson is rather unclear on this point – to take it that once we give the truth conditions of a sentence, we also give its meaning.

This proposal can be developed both as a theory of the production of meaning and as a theory of the interpretation of the meaning of a sentence in a natural language. In the former case, what is obtained is something like the following: a condition whose satisfaction leads to a true sentence (which in one of the ways of interpreting Davidson's proposal constitutes the meaning of the sentence) is obtained as the result of a kind of "conjugation" of other conditions, whose satisfaction in turn causes the component parts of a sentence (such as, for instance, names, variables, and predicates) to name/refer to an object or, respectively, to be satisfied by certain objects; then, by summing up, in reverse order, the satisfaction of these conditions governing the semantics of the component parts of a sentence, what is produced is a kind of "structured object", the attainment of which is equivalent to the realization of the condition that makes the sentence true. In the latter case (of meaning-interpretation), interpreting/knowing/understanding the meaning of a sentence is taken to be interpreting/knowing/understanding what that condition is whose realization makes the sentence true.

Thus, a truth-theory for a language L can give the meanings of every sentence of L and fulfill the task of a meaning-theory for L. (A brief remark is in order here. As the discussion of the case that I devise in Section 4 will show, this meaning theory cannot work unless we make a subtle distinction between the fact that one knows *that* the axioms and all their consequences in the truth-theory for L express truths, and the fact that he/she knows *which* truths the axioms and their consequences express or, in other words, which are the *contents* that those sentences express.)

Might it not be said, then, that the meaning of a language is obtained as a by-product of the elaboration of a Tarskian theory of truth for that natural language? This is how things seem to stand, but it is only after careful reflection that we become aware of the deep connection between the concept of truth and the concept of meaning.

The Tarskian theory of truth gives necessary and sufficient conditions for the truth of every sentence, and its condition of adequacy is formulated in terms of meaning. To give truth-conditions for each sentence amounts to giving its meaning, for to know what conditions must be satisfied for the sentences to be true is the same as to understand or give the meaning of those sentences.

Therefore, the condition of adequacy of a theory of meaning for a natural language is stated in terms of truth. Formulated in a simple way, Davidson's proposal sounds like this: "…a theory of meaning for a language *L* shows 'how the meanings of sentences depend upon the meanings of words' if it contains a (recursive) definition of the concept of truth-in-*L*"[5].

3. CRITICIZING MEANING-BASED THEORIES OF MEANING, AND JUSTIFYING DAVIDSON'S TRUTH-BASED PROGRAM

The development of the kind of theory of meaning sketched in the previous section was motivated, to a large extent, by the failure of competing approaches. In this section I shall trace some of the approaches, criticized by Davidson, based on intensional entities.

In order for the Tarskian semantic approach to produce valid results when applied to specific natural language topics and problems, the Davidson–Tarski program must be developed in a systematic way. What I have in mind here is the application of a strategy *of extensionalizing the intensional*[6]. This strategy involves two related aspects.

One aspect is the elimination of meanings as intensional entities. This leads to the need to show how certain features of natural language, which *prima facie* appear to be irreducibly intensional, can be dealt with within the bounds of a truth-conditional extensional semantics. The other aspect consists in confronting the problem of extending the extensional approach in such a way that, eventually, all intensional structures of language are dissolved in an extensional framework.

The challenge that Davidson is facing, therefore, is to show how his program can account for the problem of meaning better than a theory that assumes intensional concepts, and how it can successfully capture the intensional features of a natural language in an exclusively extensional framework.

To discuss these issues in turn, let us begin with what I think is the main feature of Davidson's program: "…my freewheeling use of the word 'meaning', for what I call a theory of meaning… [turns] out to make no use of meanings, whether of

[5] Davidson, TM 23.

[6] Forbes 23.

sentences or of words. Indeed, since a Tarski-type truth definition supplies all we have asked so far of a theory of meaning, it is clear that such a theory falls comfortably within what Quine terms the 'theory of reference' as distinguished from what he terms the 'theory of meaning'. So much to the good for what I call a theory of meaning, and so much, perhaps, against my so calling it"[7]. Thus, Davidson's truth-conditional semantics eschews the meanings of sentences as genuine entities.

The alternative account, based on the concept of meaning, owes much to the influence of Frege's doctrine of sense and reference and shares with Davidson's program the same interest in the compositionality of language, that is, in the mechanisms by which the meaning of a sentence is derived from the meanings of its component parts plus the logical structure of the sentence. Davidson, however, opposes the postulation of this kind of intensional objects for explanatory purposes: "Paradoxically, the one thing meanings do not seem to do is oil the wheels of a theory of meaning – at least as long as we require of such a theory that it non-trivially give the meaning of every sentence in the language. My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use"[8].

Before turning to the motivation that fuels this attitude towards meanings, let us see what lies at the core of this Fregean meaning-based theory of meaning. Its salient feature is the construal of the meanings of sentences and of the sentences themselves as a kind of function. The input of this function is taken to be the meanings ascribed to the parts of a sentence, but typically not to *all* of its proper sub-parts, in order for the resulting theory to be satisfactory. The output of the function is the meaning of that sentence. Of course, the details and technicalities of the functional relation need to be worked out and specified.

It is important to note that this theory, like Davidson's, seeks to solve the problem of semantic productivity. Therefore, what it seeks to show is how exactly the recurrent occurrence of a *finite* set of words, which are combined using a *finite* set of rules and which results in an *infinite* set of expressions, contributes to the meaning of each sentence belonging to this infinite set of sentences. The theory, of course, cannot claim that each and every word can be mapped to a meaning without paying a certain ontological "price", namely that of admitting that there are "unsaturated entities" corresponding to the words in the category of predicates, which do not have a "complete" sense when considered in isolation. To purge our ontology of such a dubious sort of unsaturated entities, Davidson recommends a minimalist design strategy for the principle of compositionality. According to this recommendation, it seems to be wiser to say that the meanings of sentences do not depend on the meanings of the words that figure as component parts of sentences, and that individual words have no meaning "in any sense that transcends the fact that they have a systematic effect on the meanings of the sentences in which they occur"[9].

The theory Davidson seeks to undermine comes in two distinct forms. The first is a naïve version, which claims that the meaning of a word is the object to which

---

[7] Davidson, TM 24.
[8] Davidson, TM 20–21.
[9] Davidson, TM 18.

the word refers. It is obvious that what follows from this assumption, that the meaning of a word is its referent, plus the principle of compositionality and the Fregean identification of a sentence's reference with its truth-value, is that "all sentences alike in truth value must be synonyms – [which is] an intolerable result"[10]. The avoidance of this result motivates the famous Fregean distinction between reference and sense. This sophisticated version of the theory can overcome the *cul-de-sac* into which the naïve form of meaning theory falls.

However, Davidson does not accept intensional concepts in his theory. As Loar points out, when Davidson objects to the presence of senses on the grounds that they have no obvious use, he seems to find his motivation in the fact that "meanings, as entities, are not specifiable independently of descriptions like the meaning of *Theaetetus flies*" and also in the fact that we cannot "specify relations among such entities in such a way as to be able informatively to say how the meaning of a sentence depends on the meanings of its parts"[11].

So, this new version of the meaning-based theory of meaning is not satisfactory, because it neither tells us in an informative way *what* the meaning of a sentence is, nor can it specify in an informative way relations between meanings[12].

At the same time, Davidson rejects the similar suggestion that a parallel functional explanation can be given for the *interpretation* of language by appealing to intensional entities (meanings): "It is not clear, of course, what it is for a theory to yield an explicit interpretation of an utterance. The formulation of the problem seems to invite us to think of the theory as the specification of a function taking utterances as arguments and having interpretations as values. But then interpretations would be no better than meanings and just as surely entities of some mysterious kind. So it seems wise to describe what is wanted of the theory without apparent reference to meanings or interpretations: someone who knows the theory [i.e., the recursive definition of truth for that language] can interpret the utterances to which the theory applies"[13].

We can now summarize Davidson's main worries with regard to the meaning-based theory of meaning. And, as will become apparent, his criticism of the intensional approach to meaning gradually turns into a justification of his own extensional theory.

We should note that the intensional functional approach sketched above aims to be similar to the recursive construction of syntax, by means of a recursive approach to semantics. But, as is well known, and as Davidson also points out, the recursive construction of syntax plus the dictionary that gives us the meaning of each syntactic atom do not lead to a recursive characterization of semantics. "Hopes will be dashed, however, if semantics is to comprise a theory of meaning in our sense, for knowledge of the structural characteristics that make for meaningfulness in a

[10] Davidson, TM 19.

[11] Loar 140.

[12] Loar engages in the project of showing that semantics is a component part of the psychology of propositional attitudes and that it should therefore admit intensional entities. He shows how the problems that Davidson raises can be addressed, but within an enriched intensional framework that "allows for a sufficiently rich crowd and possible worlds apparatus" (Loar 140). However, although I have certain reservations about Loar's suggestion and do not think that his proposal is unproblematic, I am not interested here in pursuing this line of thought.

[13] Davidson, RI 128.

sentence, plus knowledge of the meanings of the ultimate parts, does not add up to knowledge of what a sentence means. The point is easily illustrated by belief sentences"[14].

This failure leads us to look for a solution within the limits of a holistic approach to meaning, which rests on a noncommittal ontological view. The only way in which the theory makes any supposition about the meanings of the parts of a sentence is "in the ontologically neutral sense of making a systematic contribution to the meaning of the sentences in which they occur"[15].

Toward the same holistic conception also aims the adequacy condition that a theory of meaning must satisfy (a condition that parallels the Tarskian convention *T*): an adequate theory of meaning must have as consequences *all* sentences that have the form "*s* means *m*". To satisfy this condition we must solve two problems, the answer to the second problem being dependent on the answer to the first problem.

(1) We need a singular term in the position of *m* to stand for meanings; but if singular terms could refer to meanings, then wouldn't the latter be a kind of entity populating an inflationary ontology?

(2) Suppose we elaborate the condition we were aiming at in (1) in the following way: "*s* means that *p*", where *p* is a sentence. From this it follows either that a sentence refers to its meaning, which is not possible, however, because according to Frege the sentence refers to its truth-value, not to its meaning, or that the phrase "that *p*" refers to a meaning, which is also not possible, since "that *p*" is not a name. Furthermore, "means that" is a non-extensional operator and, as it turns out, this proposal creates more problems than it can solve.

But let's ask ourselves what work do we want this theory of meaning to do, after all? It turns out that it suffices for such a theory to match every sentence *s*, which belongs to the language we are considering, to a sentence that is required to give the meaning of *s*. If the sentence *s* in the object-language also belongs to the metalanguage, then *s* itself is a sentence that gives the meaning of *s*[16]; but if the object-language is not included into the metalanguage, then we need a translation of *s* into the metalanguage (in fact, in the former case, *s* is its own metalanguage translation).

The next move is to extensionalize the context created by "means that" by removing this phrase and changing the whole clause into another clause, governed by a biconditional whose right-hand side is the sentence itself or its metalanguage translation, and whose left-hand side consists in applying a predicate *T* to the name or designator of the sentence under consideration:

(*T*) *s* is *T* if and only if *p*. This is the crux of the Davidsonian stance of spelling out the meaning of standing sentences in natural languages in terms of their truth-conditions. This position has at least two aspects: (i) an ontological side according

---

[14] Davidson, TM 21.

[15] Davidson, TM 22.

[16] However, a caveat is in order here. Since most sentences of a natural language contain an indexical element (tense) – unlike the usual "snow is white" example, which is a standing sentence in Quine's sense – the used sentence on the right will usually be a gloss rather than the mentioned sentence itself. I owe this remark to Professor Marc Joseph.

to which meaning is truth-conditions, and (ii) an epistemic side which has it that to know/understand/comprehend the meaning of a sentence is to grasp its truth-conditions.

Thus, this is the step where Davidson pushes his critique of the meaning-based theory of meaning to the point where it merges with his own truth-based theory of meaning. He points out that the only way out of the mess is to give a theory of meaning that does not appeal at all to meanings or other intensional concepts that belong to the "theory of meaning"; on the contrary, a good theory, to which his program aspires, must belong to the theory of reference (in order to keep the discussion in the space described by Quine's dichotomy).

This concludes the justification of Davidson's proposal. What we shall discover, however, is that all the constraints we impose on the predicate "is *T*" in the schema (*T*) above make this predicate co-extensive with the predicate "true-in-*L*", as defined by Tarski's theory of truth[17]. And what we learn in this way is, as Davidson argues, in the nature of a discovery. "But whether explicitly defined or recursively characterized, it is clear that the sentences to which the predicate 'is *T*' applies will be just the true sentences of *L*, for the condition we have placed on satisfactory theories of meaning is in essence Tarski's *Convention T* that tests the adequacy of a formal semantical definition of truth"[18].

I conclude this section by saying a few words on extensionalizing the intensional features of natural language. As already noted, in order to extend Tarski's methods to natural language as a whole, we need to find a suitable way of dealing with the intensional structures of natural language. Therefore, the idea underlying this strategy is to "regiment" the natural language into the logical idiom of quantificational theory. This is tantamount to not accepting natural language sentences as such as the starting point of semantic analysis, and this is why there is a need for "a component of the theory that 'preprocesses' sentences of certain kinds, the theory of truth being required to deliver a *T*-theorem only for the sentence resulting from this processing"[19].

When Davidson formulated his program in "Truth and Meaning", he adopted, as he explicitly admits, "an optimistic and programmatic view of the possibilities for a formal characterization of a truth predicate for a natural language"[20]. At that time, the problems that counterfactual sentences, and sentences about probabilities, and causal relations, that adverbs and attributive adjectives, mass terms and verbs of action that imply purpose, that sentences about belief, perception, and intention, and finally all the sentences that seem not to have truth-value at all or at least that do not directly and obviously have it[21] – the problems that all of those were raising needed

[17] It is worth mentioning that the predicate "... is *T*" in the schema (*T*) above is undefined. So, what "discovery" means here is that it turns out that the predicate "... is *T*" is the truth predicate "true-in-*L*" for the language *L*. I owe this clarification to Professor Marc Joseph.

[18] Davidson TM 23.

[19] Forbes 23.

[20] Davidson, TM 35.

[21] The class of sentences that are not susceptible of taking truth values includes counterfactuals, sentences reporting probabilities and causal relations, sentences describing actions, sentences containing opinions, perceptions and intentions.

an approach that showed that they were tractable within the confines of an extensional framework. Davidson himself showed how to deal in an extensional framework with action-expressing sentences, adverbs, belief attribution, and that-clauses (i.e., sentences introduced by the preposition "that"); others, continuing in the same vein, have developed Davidson's program, bringing it to the point where his optimistic conception seemed to have good prospects.

However, such an approach should not be endorsed uncritically, and especially in its "preprocessing" component. There are good grounds for opposing not the extensionalization of the intensional as a general strategy, but certain regimentations into the first-order notation that distort the meanings of preprocessed sentences as we naturally grasp them, just as there is a risk that we will read far too much logical structure in the sentences we process, or there is a risk that by this regimentation we introduce entities that render our ontology dubious.

4. DOES A *MERE* TARSKIAN THEORY OF TRUTH OFFER A THEORY OF MEANING?

John is a (perhaps Davidsonian) monolingual English-speaking logician. He has been interested recently in the Davidsonian dictum that a Tarskian theory (definition) of truth for sentences of a natural language gives (allows one to interpret or to understand) the meanings of those sentences. John seeks to find out whether this dictum has empirical support no matter how the linguistic setting is devised. He picks out as his preferred theory of truth a homophonic theory[22].

As it happens, John knows the French equivalents for the English clauses that express the Tarskian apparatus through which one can recursively define truth for a language *L*, arbitrarily chosen. Hence, what his knowledge of French amounts to is the following: that French comprises the syntactic categories of names, pronouns (variables), and predicates.

Then he knows the French names for all the metalinguistic items that constitute the apparatus required by the Tarskian definition of truth: the quantifiers, the biconditional, the other truth-functional connectives, the French equivalents for "denotes", "satisfies", and "true", the form of the verb "être" at the third person in

[22] This, according to Grandy, is a theory "for a language *L* in a meta-language *ML* which exceeds *L* only to the minimum extent necessary for a definition of truth in *L*" (Grandy 179). Tarski's " 'Snow is white' is a true sentence if and only if snow is white" is a familiar example. John prefers this type of theory to other types, such as heterophonic or heteromorphic theories, due not to its simplicity but also to the fact that an assessment of the Davidsonian dictum gains if we focus on the *least* informative theory of truth.

When we say that truth appropriately defined gives meaning, we want to know whether meaning is given *only* by what the Tarskian recursive theory of truth reveals about the logical form of a sentence or else whether there are *other* ingredients that are added up to the theory and that together result in the meaning of that sentence.

So, what is at stake here is whether the theory of truth can and should serve *alone* as a theory of meaning, and then, accordingly, "[w]hatever we learn from such theories we can be sure that we learn from the truth theory itself and not from smuggled-in translations or independent analysis" (Grandy 181).

the singular, the French for "sentence", and "French language"; he knows as well how to express in French facts about infinite sequences of objects required in deriving each *T*-theorem within the theory of truth for French language.

Equipped with this knowledge, John asks his friend Pierre, who is a French logician who speaks English, for a French sentence. Let's say that the French sentence that Pierre comes up with is: "La neige est blanche".

Now, my claim is that John is able to come up with a homophonic definition of truth termed in French for the French sentence "La neige est blanche". It will read as follows:

(*T*) "«La neige est blanche» est une proposition vraie en français si et seulement si la neige est blanche".

But now it seems obvious that although John is able to give a correct Tarskian definition for the truth of a sentence he cannot, nevertheless, understand (give meaning) to that sentence[23].

However, being interested in the Davidsonian semantics, he conjures up the resources of Davidson's program and he defends its central tenet by drawing a distinction, viz. the distinction between somebody knowing *that* the axioms and all their consequences within the theory purportedly used in order to give meanings to sentences of a natural language express truths and somebody knowing *which* truths are expressed by those axioms and their consequences.

Simply put, John knows that his homophonic definition expresses a truth, but he doesn't know which truth it expresses. And because he doesn't know *which* is the truth that the homophonic definition expresses, although he knows *that* "«La neige est blanche» est une proposition vraie en français si et seulement si la neige est blanche" expresses a true French proposition, it follows that in order for him to grasp the meaning of the French sentence "La neige est blanche" and to correctly translate it into the English sentence "Snow is white", he has to know *which* truth does the homophonic definition expresses.

Hence to know or to understand or to give a meaning of a sentence doesn't amount to a sheer knowing (giving) of a truth definition for that sentence, but to knowing *which truth* does a true definition of truth for that sentence expresses.

Let's take a short look at this suggestion. If there is a distinction between the fact, if it is a fact, that the homophonic *T*-sentence is true and the truth expressed by that *T*-sentence, then it seems that we need another sentence *T′* that should refer to or designate the truth that the *T*-sentence expresses.

---

[23] John knows that his definition for the truth predicate attached to that French sentence is correct, since he knows that what the Tarskian definition of truth amounts to is that "the truth predicate is a device of disquotation" (Quine 12).

I invoke this syntactic feature of the theory because what I want to show is that John doesn't have to know that the meaning of the sentence mentioned on the left-hand side of the biconditional is the same as the meaning of the sentence used on the right-hand side.

However, it can be seen that the sentence that occurs on the left-hand side and the sentence that occurs on the right-hand side of the biconditional are series of signs of like shapes. From here, it can be inferred that the two occurrences of sentences are two tokens of a single type-sentence. And it is plausible that the two tokens have the same meaning. If, however, a doubt subsists as to this sameness in meaning between the two sentence-tokens, then we can settle that John asks Pierre for a sentence that lacks indexicals.

Now, under the arrangement just discussed concerning our monolingual logician John, for him to be able to understand *T′* it should be possible that he construct a homophonic definition of truth for *T′*. This could be done as he already did it for *T*. But by this definition all John is able to know is that *T′* is true, but not what is the truth that *T′* is expressing.

So, it is plausible to say that if John knows *that* *T* is true and if he doesn't understand *what* *T* is saying, then what John doesn't know is which is the truth that *T* is expressing. Hence, in order to understand or give the meaning of the sentence whose homophonic truth definition he just gave, he has to know which truth *T* is expressing.

Apparently, what follows is that for any standing sentence *T* of the French language, or of whatever other language John doesn't know, John can construct a homophonic definition Dh[*T*] such that John knows that the definition expresses a truth, but it is such that John doesn't know which is the truth that Dh[*T*] expresses[24].

Now, if there is a truth that Dh[*T*] is expressing, and this is neither an intensional entity, nor definable in terms germane to an intention-based semantics, then there should be another sentence *T′* such that *T′* designates or explicitly expresses the truth that Dh[*T*] is expressing, but John doesn't know it.

Does John understand the meaning of *T′*? Of course not, since *T′* designates the truth that Dh[*T*] is expressing, and John doesn't know the truth that Dh[*T*] expresses.

It won't do to say that a homophonic definition of truth for *T′* gives the meaning of *T′* and thereby we'll know the truth that Dh[*T*] is expressing. For *T′* is on a par with *T* as far as a homophonic definition of truth is constructed for it, and unless John already knows which is the truth that Dh[*T*] is expressing, which he presumably doesn't, he is not able to give the meaning of *T′* either.

It won't do either to say that we can stop this regress by fixing a sentence that not only *expresses* a truth through its homophonic definition, but it also *shows which* is the truth that it expresses. For then, what is the point of drawing the distinction from the beginning, if not for circumventing the problem that although John was able to define the term "proposition vraie" (and, hence, "true sentence") for a French sentence he was not, nevertheless, able to understand it?

Maybe something else can be said about this case, and I need to reflect more on it. Exploring something along the lines of a radical interpretation is a way out from this predicament[25]. But it seems to me that the would-be solution requires a complexity that will undermine both its plausibility and that of the approach that makes recourse to it.

---

[24] This ties to the concern that Davidson himself acknowledges when he imagines "the theory to be known by someone who had constructed it from the evidence" (Davidson, RF 173), which points to the centrality of observation sentences (as an interpreter's entering wedge) in radical interpretation. I owe this comment to Professor Marc Joseph.

[25] One piece of criticism which could be addressed to my example (and it was indeed raised by Professor Marc Joseph in personal communication) is that a *T*-theory is an empirical theory, and the *T*-sentences have the status of laws, so they are supposed to support counterfactuals, which depends on the theorems being implied by the evidence; that would not be the case in the way I have constructed my example. Now, as a way of defending my example, I can add that the empirical linguistic evidence and the content which John can have access to is partial (since he does not know the meaning of the French sentence), but nevertheless he relies on, and trusts the expertise of, his French counterpart Pierre, thereby the (partially contentful) *T*-sentences still have the status of laws, being thus implied by the partial evidence. In that sense the theory still supports counterfactuals.

Moreover, a case such as the one I discussed in this section might point to the fact that something relevant and important is missing from the truth-conditional approach and that to combine its viable and sound results with a type of analysis advocated by Strawson in his "Meaning and Truth"[26] is not quite out of track.

5. A SELLARSIAN FIX FOR DAVIDSON'S PROGRAM

I suppose now I can give an answer to the title-question of my paper. So, does a Tarskian theory of truth offer a theory of meaning? Apparently, the answer based on the analysis that I presented here would be a qualified "No!".

In a nutshell, the problem with the Davidsonian approach, as I see it, is that even if we accept in the least that a Tarskian truth theory would give the meaning of a language, still the users of that language would be in specific real situations in which they would know *that* what they uttered were true but, nevertheless, they would not be in the right position to know or understand the meaning of what they said. So, in a way, the meanings for the language might be there as a sort of transcendent semantic entities or properties, but the speakers would not be connected to them such that they would or could grasp them. However, this Platonist position of sorts regarding meanings would be odd for a Davidsonian; and this position would be even worse for a meaning theorist – such as Davidson himself – who would not believe that meanings had a real role to play in a theory of meaning for a natural language in the way that it were supposed to play, if at all, namely to show *in a testable form* how meanings would correlate to truth and reference in a compositional manner. More specifically, my thought here is that something is essentially missing in the purported account, if meanings would objectively be there, via a truth theory, but speakers could not grasp and/or follow them in their linguistic behavior. Such a theory of meaning would at least fail to show what it purports to show in a testable form.

However, since Davidson's approach has massively influenced the discussions in philosophy of language and linguistics, and since it also attracted lots and lots of comments and reactions, it is very important to see how we could fix some of the problems that the approach faces, such as the one presented above.

With this in view, I explore in the last section the explanatory virtues of Sellars' approach to meaning and truth. I submit that a fix to the specific failure of the Davidsonian approach which I pointed out can be worked out along the lines of some constitutive features of Sellars' philosophy of language (and of mind).

5.1 SELLARS' VIEWS ON MEANING

Wilfrid Sellars had a profound and nuanced perspective on meaning that challenged traditional views, particularly those rooted in empiricism. However, "like most Anglo-American philosophers of the twentieth century, Sellars's reflections on language sit at the heart of his philosophy. Understanding language is absolutely

[26] Strawson.

essential in metaphysics: it holds the key both to the nominalism–Platonism debate and to a proper understanding of the nature of mind, especially the intentionality of thought. Consequently, it is also crucial in epistemology and philosophy of science, for it is essential to understanding the mind's cognitive relation to the world. Sellars's treatment of meaning, in particular is… central to his thought…"[27]

Now, one feature of Sellars' philosophy is its systematic character. Thus, one would naturally expect that a view or a set of views of his on some specific topic would systematically correlate with consequential views on some other topics. For what concerns us here, his views on meaning are strongly connected to his grander philosophical project of harmonizing the scientific image of the world with the manifest, i.e., our day-by-day common understanding and image of the very same world. Sellars made this distinction between two aspects of human understanding: the "manifest image" and the "scientific image". Sellars saw both images as important for a complete account of meaning, as they reflect different ways of conceptualizing reality.

So, in order to harmonize the two images, Sellars develops what is his main contribution to the philosophy of meaning, which consists in his critique of what he called the "myth of the given". Sellars argued against the naïve positivistic view that knowledge is based on a foundation of non-conceptual, immediate, raw experiences[28].

Contrary to this, he maintained that all perception, which after all provides the contents of our empirical knowledge, is theory-laden, and hence shaped by prior conceptual frameworks and language. This has consequences for how we conceive meaning. Meaning cannot be grounded in raw sensory data alone, for indeed there are not such things. Meanings are not given. They are constructed within a linguistic and conceptual system, thereby being intrinsically linked to our linguistic and conceptual practices.

Sellars argued that our understanding of meaning is inherently tied to the linguistic practices and conceptual frameworks within which perceptions are interpreted. This perspective led him to advocate for a view of meaning that incorporates both the roles of language and the norms governing language use.

For Sellars, both the "manifest image" and the "scientific image" are essential, and they inform one another. Moreover, Sellars' account of meaning is univocal across the two frameworks. They differ, though, in what concerns the meanings of their terms, viz. roughly middle-sized and big objects, respectively, and very small objects which are postulated by our best theories and which could be observed in a given experimental set-up. Thus, what remains stable are the definitions and criteria of meaning, and what varies are the categories of objects that theoretical terms refer to. This also accounts for a sort of "complementarity" principle similar to the one framed by Bohr and others in the philosophical discussions on the foundations of quantum mechanics[29].

---

[27] deVries 23.

[28] Consistent with his critique of the "myth of the given", Sellars took an anti-foundational stance. He argued against the possibility of a secure, infallible foundation for knowledge, suggesting instead that knowledge and meaning arise through complex interactions within our conceptual and linguistic frameworks.

[29] I owe the input and motivation for this remark to a critical question raised by Professor Willem deVries.

Overall, Sellars' views on meaning emphasize the importance of language, social practices, and the interplay between different images of reality in understanding how we come to assign meaning to our experiences. To make a proper distinction which is important for my argument in this paper between Davidson's truth-conditional semantics and Sellars' pragmatist-expressivist views on language, it is worth pointing out some key aspects of Sellars' views on meaning.

Concerning the overall role of language, Sellars emphasized that language plays a fundamental role in shaping our understanding of reality and meaning. Contrary to model-theoretic approaches or truth-conditional approaches to meaning, he argued that meaning is not just a matter of reference or correspondence to the world, but involves the linguistic practices and the ways that terms are used in context. Consequently, he believed that meanings are not static entities, but are dynamic and situated within the contexts of language use. This idea positions language as a crucial mediator between our thoughts and the world. In line with later developments in ordinary language philosophy, he suggested that understanding the meaning of a term involves looking at how it is used within language. This perspective prioritizes the pragmatic aspects of language, and naturally leads to the idea that understanding a term or concept requires understanding how it functions within the language and the context of its use.

Furthermore, this position entails views on the functional role semantics, inferentialism, and not in the least on the intentionality of the mental. Sellars contributed to what later developed into inferential role semantics, the idea that the meaning of a term is derived from its role within a network of inferences and relationships among beliefs, intentions, actions, and statements. This means that understanding a statement is tied to understanding how it functions within reasoning and discourse. It further means that understanding a word, for instance, involves understanding the rules and norms which govern its use in linguistic and also non-linguistic contexts. Words do not have intrinsic essential meanings; rather, their meanings arise from the roles they play in specific "language games".

As a matter of consequence, he thought of meanings as a matter of interpersonal practice, thus considering meanings as inherently social. He believed that the meanings of terms are embedded within communal practices and that understanding meaning involves recognizing how terms are used and accepted within a specific linguistic community.

To deepen this view on the rule-governed, pragmatic, and social aspect of meanings, he articulated the position that we now dub inferentialism. In this regard, Sellars emphasized the formal and also material (contentful) inferential relations between statements. Thus, to understand the meaning of a statement, one is supposed to understand the inferences it supports and the inferences it follows from. This view changes our perspective from an atomic based understanding of the meanings of terms to the network and logical interconnections of sentences[30].

Within the same inferentialist framework, Sellars also engaged with the notion of intentionality, which is the capacity of thoughts to be about something. Sellars'

[30] This will turn out to be very important in fixing the Davidsonian truth-conditional (formal) theory of meaning.

views on the intentionality of the mind and the relationships between language and mind are following from his functional role semantics and inferentialism. Thus, for him, language is fundamental for thought. Thinking is a kind of inner speech. Hence, understanding meaning, which is a function and capacity of the mind, involves understanding the linguistic practices that are constitutive for thinking. According to Sellars, our cognitive capacities are profoundly intertwined with our linguistic practices.

One last important aspect of Sellars' views on meaning that I summarize here is the normativity of meaning. Regarding this feature, Sellars argues that to grasp the meaning of a term boils down to understanding the norms which govern the correct use of that term. So, meanings are not related only to descriptive facts, but also to prescriptive rules that constrain how are we supposed to use that term.

To sum up all those features that make up Sellars' profound and original views on meaning (of language and of mind) we can highlight the following: (i) the focus on the functional role of language, (ii) the functional/conceptual role semantics, and (iii) inferentialism, (iv) the rejection of foundationalist epistemology, (v) the interdependence of the language and thought, and (vi) the normative features of linguistic practices and meanings. Overall, the profound influence of Sellars may be not very easily noticeable to the naked eye. Actually, one may say that the current state of the scholarship does not do full justice to his philosophy; at least not to the extent that it should deserve if we take into consideration the depth, ingenuity and originality of Sellars' philosophy. However, maybe his time is yet to come. Some signs are already there. If we take a careful look, for instance, at the current trends in the study of cognition, one could discern his lasting influence along the lines of thought that I summarized in this sub-section. Thus, substantive critical engagement with Sellars' views on mental states, mindreading and mindshaping, propositional attitude ascriptions, conceptual role semantics, inferentialism, prompted the rise of this new, ground-breaking 4E cognition paradigm. Sellars appears as a forerunner whose massive systematic thinking on those issues gave an impetus for the clarification of this new conceptual paradigm shift about embodied, embedded, enactive, and extended mind and ways of looking at cognition. There you have a sample of this dialectical engagement with Sellars. "The whole question… [of mindreading, mindshaping, and the distinctiveness of human social cognition] turns on what we mean by 'sophisticated mindreading' and 'propositional attitude attribution'. Most philosophers of psychology follow Wilfrid Sellars' *Empiricism and the Philosophy of Mind*[31] when interpreting these concepts. Propositional attitudes are treated as states of an unobservable nexus responsible for an agent's behavior: the agent's mind. Furthermore, … their relations to observable circumstances and behavior are holistically constrained: what one does in specific circumstances depends on indefinitely broad networks of propositional attitudes; hence it should be difficult to determine an agent's propositional attitudes based on observations of finite bouts of behavior, and an agent's future behavior based on attributions of finite sets of propositional attitudes. Finally, if we take the Sellarsian picture seriously, and

[31] Sellars.

think of propositional attitude attribution on the model of scientific hypotheses about unobservable causal factors, then propositional attitude attribution should involve a strong appearance/reality distinction"[32]. Now, what we have here interwoven are most of the canonical Sellarsian themes in philosophy of mind, of language, and epistemology. And of course, people working today in 4E Cognition may distance themselves from this canon. But canon it is. And I do want to "take the Sellarsian picture seriously" when going back to the Davidsonian conundrum with the aim of fixing it.

5.2 FIXING DAVIDSON THROUGH SELLARS

The problem with Davidson's approach, as shown in Section 4, is that one can legitimately know *that* a Tarskian *T*-Convention gives the truth of a given sentence (and presumably its meaning) but one is not in the position *to know what* is the content or the meaning of that truth.

For me, this indicates firstly that the strategy of extensionalizing the intensional, which is part and parcel of Davidson's approach to giving the meaning of a language through the truth predicate for its sentences, fails as a *general strategy*. It does not produce the outcome that it is supposed to produce. More generally, this offers a ground for rejecting the view that the intensional (the propositional attitudes, the contents of the mental states and/or processes) could be in principle reduced or even eliminated in favor of extensional entities only. The intensional is fundamental and irreducible.

More specifically, this general intensional stance that I take here finds strong underpinnings in some of Sellars' arguments that go along with his general philosophy of meaning presented in Section 5.1. Thus, in contradistinction to the Davidsonian extensionalization of "means that"[33], Sellars advocates an anti-empiricist stance which targets the traditional position called "concept empiricism": "a proper analysis of meaning shows that we do not have to accept a traditional empiricist account of understanding according to which there are simple, basic words or concepts that get their meanings solely and directly from experience and that can be understood independently of any other concepts or linguistic expressions"[34].

Sellars leaves us here with two positive key ideas of his non-relational analysis of meaning-statements: (i) expressions have meanings in virtue their connections of the word–object and word–word types; and (ii) the connections of the word–object type consist in associations between words and physical objects, not sense data, sensations, or other private subjective occurrences (physicalism wins over phenomenalism). These two interrelated features are not enough to fully depart from

[32] Zawidzki 747.

[33] As observed before, the Davidsonian strategy of extensionalizing "means that" consists in removing this phrase and changing the whole clause into another clause, governed by a biconditional whose right-hand side is the sentence itself or its metalanguage translation, and whose left-hand side consists in applying a predicate *T* to the name or designator of the sentence under consideration: (*T*) *s* is *T* if and only if *p*.

[34] deVries, Triplett 59.

Davidson's strategy. What makes a real difference here are the connections at play which have normative status. Likewise, alternatively, one could say that Sellars turns things on their heads, because in his view meaning is prior to reference (not temporally, but in the sense that reference is a special case of meaning)[35].

There are two interrelated features of Sellars' views on meaning which should complement the Davidsonian structural account thereby filling the gap in the latter. The first feature is Sellars' analysis of the "thermometer view" of meaningfulness[36]. On this view, a linguistic expression having a particular meaning is essentially dependent upon a correlation with features of the environment – "just as thermometer indications are about the temperature because they correlate with it under normal conditions"[37]. However, it is noteworthy that whether or not there is a mental entity or any inner process of "understanding" correlated to the sentences is completely irrelevant for the linguistic expressions having a meaning. Only facts determine expressions' meanings, viz. facts concerning features that are robustly and reliably correlated to words and sentences[38].

---

[35] I owe this remark to Professor Willem deVries.

[36] Cf. deVries, Triplett 65–66.

[37] deVries, Triplett 65.

[38] However, at the end of the day Sellars rejects the thermometer analogy. So, it is incumbent upon the whole argumentation here to spell out in more detail what Sellars has in mind with his example and how one can use it with regard to the account of the relationship between language, mind, and knowledge. Sellars uses the "thermometer view" to describe a concept central to his philosophy of mind and epistemology, particularly in understanding how we relate perceptual data to conceptual knowledge. Sellars employs the analogy of a thermometer to explain that not all forms of response to environmental conditions constitute knowledge in the human sense. He uses this analogy to draw a line between mere causal responses to the environment and true perceptual or conceptual awareness.

In this analogy, a thermometer's response to temperature is purely mechanical or causal: it registers a temperature without any awareness or understanding. This response is non-conceptual because the thermometer does not "know" or "understand" what it measures – it simply reacts according to the physical properties of its components.

Sellars uses this model to make a point about non-human animals and even about humans' responses to stimuli before they apply any conceptual frameworks. For example, when we react to temperature changes instinctively or non-reflectively, this reaction might be compared to a thermometer's response. However, this reaction is not equivalent to knowing or perceiving temperature in a conscious, reflective sense.

Sellars' main point about the thermometer analogy is that true knowledge or perception (in the human, conceptual sense), requires more than a causal reaction; one needs a system of beliefs, concepts, and norms that can meaningfully interpret and integrate sensory data. For Sellars, a human being who perceives coldness is doing something fundamentally different from a thermometer, as the human can express, justify, and reason about the perception within a larger web of beliefs and language. This interpretive framework is essential for Sellars' concept of "the space of reasons" – a domain in which sensory information becomes meaningful only when situated within a network of inferential, normative relations.

Sellars uses the thermometer analogy to argue against simplistic views on perception and to advocate for a more sophisticated understanding of mental processes. He suggests that mere sensations or responses are insufficient for knowledge; instead, knowledge requires a capacity for concept application, inferential justification, and an understanding of rules and norms. In this way, Sellars positions human cognition as distinctive in its ability to transform raw perceptual data into knowledge by embedding it in a conceptual framework. I thank Professor Willem deVries for prompting this footnote which spells out the gist of Sellars' thermometer analogy.

The second feature opens up word meaningfulness to the whole sentence and language meaningfulness. This is the inferentialist rule-based and conceptual-role semantics which accounts for the meanings of terms and sentences through the understanding of the norms or rules which govern the correct use of them within the networks in which they occur. Historically, the idea that the meanings of various logical concepts are to be rendered via rules for introducing those terms into and respectively for eliminating those terms from the discourse is associated with the work of the German mathematician and logician Gerhard Gentzen.

Gentzen's successful project was to devise rule-based systems of deduction. For each and every logical constant in the language of the system one should come up with a pair of rules, one which allows us to introduce the constant and a companion rule which shows how we can eliminate the constant. It is the rules themselves which give the logical meanings of those logical words, and not the models or linguistic representations we might associate in the vernacular with those logical constants.

In addition to Sellars' work, the work of M. Dummett[39] and more recently of R. Brandom[40] have opened wide new horizons upon the connections and the order of explanation between representations and rules or, if we prefer an alternative set of terms, between model-theoretic and proof-theoretic categories.

According to inferentialism, each linguistic expression gets its meaning via the role that the expression plays in reasonings (inferences). The use of any linguistic expression or concept has two aspects: (i) the *circumstances* under which the expression is correctly applied, uttered, or used (which are governed by Introduction Rules); and (ii) the *consequences* of its application, utterance, or use (which are governed by Elimination Rules).

The explanatory virtue of this inferentialist semantics is encapsulated within the following *inferentialist principle*: the *content* to which a person is committed by using the concept or expression may be represented by the *inference* she implicitly endorses by such use, the inference from the *circumstances* of appropriate employment to the appropriate *consequences* of such employment.

To sum up this section and paper, my argument is that both the model-theoretic and the truth-conditional Tarskian semantics are not enough to account for the real dynamics of giving the meanings for a natural language and for explaining the way we get to grasp or understand those meanings. Important as they are for making sense of the meanings in terms of the compositional structure of the languages, they are not the whole story. And that is for a very good reason. Rather, an inferentialist conceptual-role semantics is apt to explain the word meanings and the sentence meanings supplementing the structural semantic explanation with a rule-based and rule-governed practice of how speakers get to use properly the terms and the sentences of a language.

Thus, to give and to grasp the meaning of a sentence within a language is to articulate its truth-predicate and additionally, and essentially, to know a significant

[39] Dummett.

[40] Brandom.

amount of *circumstances* under which the expression is introduced into the discourse, and is correctly applied, uttered, or used; and also a significant amount of the *implications*, and the *consequences* of its application, utterance, or use.

John, our monolingual English Davidsonian logician, would understand the meaning of the French sentence "La neige est blanche" not only through his knowing *that* the substitution in the *T*-convention "«La neige est blanche» est une proposition vraie en français si et seulement si la neige est blanche" expresses a true French proposition, but also if he knows a significant amount of circumstances in which the French words are correctly introduced in the discourse and likewise a fair amount of circumstances in which the French sentence is implied by other French sentences and implies in its turn many other French sentences. Only if and when he masters those introduction and elimination rules, respectively, he knows enough to grasp the meaning of the sentence and language.

BIBLIOGRAPHY

Brandom, R.B. *Articulating Reasons*. Cambridge, MA: Harvard University Press, 2001.

Davidson, D. "Truth and Meaning." *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984, 17–36.

Davidson, D. "Radical Interpretation." *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984, 125–139.

Davidson, D. "Reply to Foster." *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984, 171-179.

deVries, W., and T. Triplett. *Knowledge, Mind, and the Given: Reading Wilfrid Sellars's "Empiricism and the Philosophy of Mind" Including the Complete Text of Sellars's Essay*. Indianapolis: Hackett, 2000.

deVries, W. *Wilfrid Sellars*. Montreal: McGill-Queen's University Press, 2005.

Dummett, M. *Frege: Philosophy of Language*. Cambridge, MA: Harvard University Press, 1981.

Fine, K. *Semantic Relationism*. Oxford: Blackwell, 2007.

Forbes, G. *Philosophical Logic: Supplementary Notes*. Tulane University, 1995.

Grandy, R.E. "What a Truth Theory Need Not Tell Us." *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Edited by Ernest LePore. Oxford: Blackwell, 1986, 179–189.

Loar, B. "Two Theories of Meaning." *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Edited by Ernest LePore. Oxford: Blackwell, 1986, 138–161.

Quine, W.V.O. *Philosophy of Logic*. 2nd ed. Cambridge, MA: Harvard University Press, 1970, 1986.

Sellars, W. *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press, 1956, 1997.

Strawson, P F. "Meaning and Truth." *The Philosophy of Language*. Edited by A. P. Martinich. 2nd ed. Oxford: Oxford University Press, 1990, 91–102.

Tarski, A. "The Concept of Truth in Formalized Languages." *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press, 1956, 1983, 152–278.

Tarski, A. "The Semantic Conception of Truth." *The Philosophy of Language*. Edited by A.P. Martinich. 2nd ed. Oxford: Oxford University Press, 1990, 48–72.

Zawidzki, T.W. "Mindshaping." *The Oxford Handbook of 4E Cognition*. Edited by A. Newen, L. de Bruin, and S. Gallagher. Oxford: Oxford University Press, 2018.