Jaroslav Peregrin

LANGUAGE AND ITS MODELS: IS MODEL THEORY A THEORY OF SEMANTICS?

1. Introduction

Tarskian model theory is almost universally understood as a formal counterpart of the preformal notion of semantics, of the "linkage between words and things". The wide-spread opinion is that to account for the semantics of natural language is to furnish its set-theoretic interpretation in a suitable model structure; as exemplified by Montague 1974.

The thesis advocated in this paper is that model theory cannot be considered as semantics in this straightforward sense. We try to show that model theory is more adequately understood as shining light on considerations concerning the relation of consequence than on those concerning the relation of expressions to extralinguistic objects; and that it makes little sense to use model theory for the purposes of answering such questions as *what is meaning?* or *when is a sentence true?*.

The organization of the paper is the following: We start by considering various formal reconstructions of natural languages utilizing standard logic. Section 2 points out that the usual way of explicating the semantics of natural language, namely the way of Tarskian model-theoretic interpretation, is problematic. In Section 3 we propose a more adequate way: to explicate semantics via the delimitation of the space of possible truth-valuations of sentences, especially in terms of the specification of the relation of consequence (indicating that logic, as an account for consequence, is *eo ipso* a theory of semantics). Section 4 demonstrates that a plausible way to account for the space of possible truth-valuations is to locate a "basis", i.e. a class of mutually independent sentences each truth-valuation of which uniquely extends to a truth-valuation of the whole language; and it is noted that some of

Nordic Journal of Philosophical Logic, Vol. 2, No. 1, pp. 1–23.

the most usual traditional logical calculi, notably the first-order predicate calculus, may lack such a basis. In Section 5 we consider another example of a baseless language, namely the modal propositional calculus, and we indicate the general pattern which appears to invoke the idea of a model. Section 6 puts forward the thesis that the idea behind model theory is that of making such baseless languages based. In Section 7 we analyze the concept of quantifier, which appears to be crucial for the fact that predicate calculus is in general baseless; we point out that the seemingly indirect way in which a quantifier alludes to the infinite cannot be replaced by any direct way. In Section 8 we draw the general conclusion regarding the nature of model theory: we conclude that model theory can be understood as concerned with extending baseless languages to based ones; however, that doing model theory usually does not mean presenting such an extension, but rather only to stipulate its existence. Section 9 concludes that model theory should not be considered as a direct presentation of semantics, not only because it connects words with words (rather than words with things), but ultimately because it cannot succeed in explicitly connecting semantically non-transparent (baseless) languages with semantically transparent (based) ones. In Section 10 we point out that the common "ideology" behind model theory is closely connected with the doctrine of logical atomism as pursued at the beginning of this century by Russell, Carnap and Wittgenstein. Section 11 then lays out our conclusion that insofar as model theory can be considered a theory of semantics, it is not in force of its being an imitation of the "real" *denotandum/denotatum* relation, but rather in force of its being an account for consequence.

2. Language and the Ways of its Reconstruction

What is language? It is now commonplace to consider language as a bundle of generative and/or transformational rules, as some kind of algebra, as a set-theoretically interpreted logical calculus, etc. Such views need no substantiation as long as we stay within the realm of mathematics: there we are relatively free to define the concepts we use. But, outside of mathematics we must not forget, however banal this may sound, that language is first and foremost something to be found in our world, that it is something *factual.* This is what is stressed by Wittgenstein (1953, §494): "Der Apparat unserer gewöhnlichen Sprache, unserer Wortsprache, ist *vor allem* das, was wir 'Sprache' nennen; und dann anderes nach seiner Analogie oder Vergleichbarkeit mit ihr".[1]

It is obviously legitimate, and indeed helpful, to try to reconstruct factual language in terms of a mathematical theory. The attempt may be guided by diverse purposes—we may want to throw light on the workings of some restricted range of expressions (such as logical connectives), to regiment a part of language in a form manageable by computers, etc. What concerns us here is the purpose of *explication,* i.e. that usage of formal means which aims at fostering our *understanding* of the phenomenon of language. If some kind of a mathematical structure is to be considered a reconstruction of language relevant in this respect, then it has to share some substantial amount of the characteristic features of the factual language. Which features are considered relevant, and which not, is again partly a matter of viewpoint; but once we are clear about what functions of language we consider relevant we may, in terms of Quine's (1960, p. 259) classic definition of *explication,* "devise a substitute, clear and couched in terms of our liking, that fills the functions."

However, there seem to be also certain general principles governing adequate explication, principles that can be articulated independently of a viewpoint. One such principle can be called *the principle of persistence of individuation* (hereafter *PPI*): it states that different *explicanda* should obtain different *explicata* and vice versa, that the correspondence between the domain of the pre-theoretic entities that are to be explicated and that of those which are used to explicate them should be one-to-one. If we try to explicate the intuitive concept of ordered pair, then we cannot furnish an explication such that two intuitively different ordered pairs would be explicated by one and the same object, or that one ordered pair would be explicated by two different things. If we try to explicate the concept of language, then we should maintain the one-to-one correspondence between languages in the intuitive sense and the theoretic constructs we employ as their explications.[2]

One possible way of reconstructing natural language is to reconstruct it as a grammatical system, as a set of basic strings of letters (lexicon) plus a set of rules to combine strings of letters into strings

[1] "It is primarily the apparatus of our ordinary language, of our word-language, that we call language; and then other things by analogy or comparability with this."

[2] The factual boundaries between languages are, of course, both fuzzy and ambiguous (in one sense, we can regard, e.g., British and American English as two subspecies of a single language; in another sense we can regard them as two different languages). However, this is not in itself problematic—it is in the nature of *explication* that it replaces fuzzy lines with crisp ones and sticks to a single sense where ambiguity occurs.

of strings of letters (grammar). This is quite a useful reconstruction employed by syntacticians; however, it is based on the (purposeful) disregard of the semantic aspect, it clearly violates *PPI* (for two different languages may conceivably share one and the same syntax) and it is thus essentially incapable of being taken as a genuine explication of the concept of language. What makes language into more than a grammatical system is that its expressions are in some sense *significant,* that they *have meaning*—the reason two different languages with the same syntax may still be different is that they can differ in meaning, in what their expressions 'stand for'. So it may seem that to account for the whole of language we need a grammatical system plus an assignment of some kind of entities to expressions.

This intuition appears to be done justice to by the notion of language established by Carnap (with support from Morris, Tarski and others), which is nowadays taken almost for granted. According to this commonly accepted doctrine a language consists of (i) a list of basic expressions (words; the *lexicon*); (ii) a list of rules which produce more complex expressions out of simpler ones (extending the lexicon into an algebra generated by its words; the *syntax*); (iii) an assignment of some basic set-theoretic objects to basic expressions; (iv) an assignment of operations over the set-theoreticals to rules (extending the interpretation of words into the homomorphism of the whole algebra). As Montague claims, this notion of language is so encompassing that it bans any need to make a difference between natural language and formal ones.[3]

However, it is necessary to realize that to take this kind of reconstruction directly as an explication would still violate *PPI*; in this case because there would be more than one explicatum for a single explicandum, for a single language. First, it is quite clear that the grammar of a language can be articulated in various ways. If we take English, then, e.g., *John, to love* and *Mary* can be composed into *John loves Mary,* but we can account for it either by postulating a rule which combines a transitive verb phrase with a noun phrase into an intransitive verb phrase and a rule combining an intransitive verb phrase and a noun phrase into a sentence, or we can have a single rule combining a transitive verb phrase with two noun phrases into a sentence (not to speak about other possibilities). Second, also the semantics of a language can be accounted for by means of various set-theoretic interpretations. We know, for instance, that it makes no substantial difference whether we interpret a predicate by a subclass of the universe, or by the charac-

[3] Montague 1974, pp. 188 and 222.

teristic function of the subclass; however, a class and its characteristic function are two different things and corresponding interpretations thus make up different objects. Similarly, we know that as far as the usual principles of arithmetic are concerned we may consistently consider a number to be a primitive object, a set of sets of objects (as Frege in effect proposed), or the set of all the smaller numbers (as proposed by von Neumann).

This indicates that what makes up a language in the intuitive sense should be considered not as a definite grammatical system plus a definite set-theoretic interpretation, but something 'more abstract', something which is shared both by different but equivalent grammatical systems, and different but equivalent set-theoretical interpretations. This is to say that to furnish a real explication of the concept of language we should have to be able to decide when two different grammatical systems are mere alternative ways of capturing the syntax of the same language, and when they concern different languages; and similarly when two set-theoretical interpretations are mere alternative accounts of its semantics and when they differ substantially. In other words, we should have to specify identity criteria, a criterion of identity for grammatical systems and a criterion of identity for interpretations.

The situation is relatively straightforward in the case of syntax: two grammatical systems seem to be two different accounts for the same language if and only if they generate the same class of sentences. The class of sentences (or, more basically, the speakers' ability to tell a well-formed sentence from a mere haphazard string of letters) is what can be considered as genuinely empirically capturable, the grammar is our way of accounting for this class.[4] What two grammatical systems relating to one and the same language have in common is the class of sentences; hence we can identify the explication of the syntactical part of language with this set.[5]

However, the situation is more complicated in respect to interpretation; here no such straightforward criterion is in sight. To volunteer a reconstruction of semantics doing justice to *PPI*, we need to differentiate interpretations only up to some kind of "semantic equivalence" or

[4] Within linguistics a grammar is sometimes considered adequate if it generates the right class of sentences *and* if it associates the sentences with some kind of "right" inner structures. However, it is problematic on which empirical grounds it could be decided which inner structures are right and which not; it rather seems that the inner structure is a mere by-product of the reconstruction of the class of sentences and that hence the structure is right if and only if it is the product of an adequate reconstruction. This debate, though, is beyond the scope of the present paper (an interested reader is referred to Peregrin 1996b).

[5] Cf. Quine 1969, pp. 48–49.

"isomorphism" (or, to use a term coined by Quine, up to some kind of *proxy function*). Arithmetic is more than a bare class of sentences; but it is in the same time less than the class plus an assignment of definite objects. Language is a class of sentences plus that part of its interpretation which it shares with all interpretations which are "semantically equivalent" (*elementarily equivalent,* in terms of current model theory). If we managed to reify the structure which any two so "semantically equivalent" interpretations share, than we have that which must be added to syntax to make up a good explication of language.

3. Meaning and Truth

So what must two set-theoretic interpretations share in order to be understood as two representations of the same semantics, rather than as representations of two different semantics? Let us consider arithmetic. It is not relevant whether we consider numbers in von Neumann's sense and $<$ an inclusion, or numbers in Frege's sense and $<$ the relation $\lambda XY.\forall x \in X.\exists y \in Y.x \subset y$; all that counts is that the interpretation makes the right sentences true: that $1 < 2$ comes out true, while $2 < 1$ false. This is what Quine (1953, p. 45) seems to have in mind stating: "there is no saying absolutely what the numbers are, there is only arithmetic." In the same sense it is inessential whether a predicate P is interpreted by a set or by the characteristic function of the set; all that counts is the truth values which result from the combination of P with terms and quantifiers. This suggests that what might be added to grammar to make up a fully-fledged language is something like a specification of truth.

The situation is relatively straightforward in the case of arithmetic or any other mathematical language. For such languages truth coincides with eternal, necessary truth; what is actually true is always true. In such a case we can speak about *the* class of true sentences, or about *the* truth-valuation (assignment of truth values to sentences). However, in the general case there may be truths which are not necessary, truths which may become false at any moment. There is no definite set of truths, no definite truth-valuation. A sentence such as 'It is raining' may be true in the morning, but false in the afternoon.

In the general case it is thus not possible to speak about *the* class of true sentences, or *the* truth-valuation; we can at most speak either about the class of *actually* true sentences, (*the actual* truth-valuation), or about the class of *all the possible classes* of true sentences (truth-valuations). The enterprise of the account for truth for a general language thus may be thought as split into two distinct enterprises: first, the delimitation of the space of all the possible truth-valuations and, second, the specification of the actual truth-valuation. We can call the first of these enterprises, borrowing from Schlick (1932), *the pursuit of meaning,* while calling the latter *the pursuit of truth.* The former is a matter for philosophy and logic, the latter one for the sciences.[6]

A logical account of truth for a general language thus means the delimitation of the space of all the possible truth-valuations of the sentences of the language. For English, such a space would contain truth-valuations which would verify (assign *the truth* to) ‘Aristotle is human’ and it would also contain those which would verify ‘Aristotle is not human’, but it would contain no truth-valuation which would verify both of these sentences, and it would probably also not contain one which would falsify both of them.

However, the delimitation of the space of possible truth-valuations seems to be far from enough to really account for semantics. It seems to be insufficient to know that ‘Aristotle is human’ cannot be true simultaneously with ‘Aristotle is not human’. We have to know *when* ‘Aristotle is human’ is true; we must know its *truth conditions.* However, are we really in a position to articulate the truth conditions in a nontrivial way? We can surely say that ‘Aristotle is human’ is true just in case that Aristotle is human. There seem to be many logicians and philosophers who, since Tarski, consider this kind of ‘correspondence-theoretic’ explication of truth conditions satisfactory; but there are also more skeptical theoreticians of language, such as the late Wittgenstein or Quine, who consider this to say nothing nontrivial.[7] The statement ‘“Aristotle is human” is true if and only if Aristotle is human’ can never bring us nontrivial information, because it begs the question: to be able to understand it we have to know what it means for Aristotle to be human, and to know this is nothing other than to know when the sentence ‘Aristotle is human’ is true.[8]

[6] Of course it is just this division which was challenged by Quine’s famous attack on the analytic/synthetic distinction. However, as we have stressed before (see footnote 2), a theoretical explication consists, besides others, in positing sharp boundaries where there are in fact none. Thus, although Quine is certainly right that we cannot turn the distinction between *the pursuit of truth* and *the pursuit of meaning* into philosophical cash in the way Schlick proposed, we do need such a boundary when building a formal model of language.

[7] As Wittgenstein (1977, p. 27) puts it: “Die Grenze der Sprache zeigt sich in der Unmöglichkeit, die Tatsache zu beschreiben, die einem Satz entspricht (seine Übersetzung ist), ohne eben den Satz zu wiederholen.” [“The limit of language is shown by it being impossible to describe the fact which corresponds to (is the translation of) a sentence, without simply repeating the sentence.”]

[8] Caveat (due to P.Tichý): The statement ‘“Aristotle is human” means that Aristotle is human’ (or ‘“Aristotle is human” is true if and only if Aristotle is

This indicates that the only way theoretically to account for the semantics of a language over and above the demarcation of the space of its possible truth-valuations is to translate it into another language taking the semantics of the other language for granted. This claim may immediately raise suspicion: would it in such a case ever be possible to learn a first language? However, to learn a language and to articulate its semantics explicitly are two essentially different matters: learning a language is an enterprise that is essentially *practical* and the word-thing (or, better, sentence-circumstances) links which are established during it cannot be directly spelled out within a *theory*—we may teach someone to use the term 'dog' to refer to dogs, but by putting this into words we get the trivial '"dog" refers to dogs'. A theory can pair words with words; it is only practice that can literally "pair words with things".

However, do not modern formal theories of semantics aim higher than that: do they not explain the semantics of language in an absolute way (not only relatively to another language)? Do they not provide a direct access to meanings by way of their set-theoretic articulation? Of course not: such claims as that 'dog' stands for the class of dogs, or that the sentence 'Aristotle is mortal' is true in those possible worlds in which Aristotle is mortal are subject to the same objection as above; namely that to understand them we would have to know what they try to communicate to us in advance. Only if we know what 'dog' means can we understand what the class of dogs is; and only if we know when 'Aristotle is mortal' is true we can understand which are the worlds in which Aristotle is mortal.

This is why Quine (1953, p. 49) suggests that the problem of meaning boils down to the problem of synonymy.[9] However, both Quine's and, especially, Davidson's (1984) considerations also suggest that meaning can be somehow approached via truth: and it is this idea on which we elaborate here. The decoding of an unknown language involves, as Quine's and Davidson's thought experiments with radical translation or interpretation make plausible, finding out which

human') *is* informative because it is a literal English translation of the German statement '"Aristotle is human" bedeutet daß Aristoteles menschlich sei', which is evidently informative. However, if this were true, then the English translation of the *Englisch–Deutsch Wörterbuch,* containing statements to the effect that 'dog' means dog, etc., would have to be as informative as the original. Nevertheless, while the *Englisch–Deutsch Wörterbuch* is surely a bestseller, one could hardly believe that a single copy of the English translation could be sold.

[9] Elsewhere Quine puts this point even more vividly (1953, p. 50): "What makes sense is to say not what the objects of a theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another."

sentences are true when, and also how the truth of one sentence depends on that of other sentences. We suggest that if we allow ourselves of positing a sharp boundary between possible and impossible truth-valuations (which is, as we indicated, admittedly an oversimplification in respect to natural language, but a helpful one in respect to its theoretical explication), we can see semantic explication of a language as the delimitation of the space of its possible truth-valuations. What we tried to indicate above was that a set-theoretic interpretation can do no better than the delimitation of this space. In other words, our suggestion is that if we want to furnish a formal explication of a language, then a suitable one, not violating *PPI,* would be (the specification of) a class of sentences plus (the specification of) a space of possible truth-valuations of these sentences.

Hence the first thesis of the present paper: *A language is most adequately considered as a class of sentences plus a space of their truth-valuations; hence to account for the semantics of a language means to delimit the space of its truth-valuations.*

4. Entailment and Bases

The space of possible truth-valuations can be considered one side of the coin the other side of which is consequence. Indeed, consequence can be understood as a specification of the boundaries within which it is reasonable to consider alternative distributions of truth values among sentences: to say that the sentences $S_1, \ldots, S_n$ entail the sentence S is to say that it is impossible for S to be false if $S_1, \ldots, S_n$ are true; hence it is to say that a truth-valuation which would assign T to all of $S_1, \ldots, S_n$ and F to S is unacceptable. Hence *any account for consequence is* eo ipso *an account of the space of possible truth-valuations, and vice versa.* This establishes the important link between logic and semantics: to do logic means to work toward an account of consequence, and hence, in view of the conclusion reached in the previous section, towards an account of semantics.

Now the most straightforward way to account for consequence and hence for the space of the possible truth-valuations is clearly the axiomatic method. (The axiomatic method aims simply at a list of all the instances of consequence—or, which is, for a "reasonable" language, the same, of necessary truths; it is only because such instances are infinite in number that it is forced to provide, in proxy, merely a potential generator of the list.) We have seen that an axiom can be understood as a statement to the effect of the exclusion of some truth-valuations; hence an exhaustive system of axioms can be understood as a (negative) demarcation of the class of those truth-valuations which are possible.

However, the axiomatic method is not the only possible approach to such a demarcation. There is also another approach, an approach based on the idea of finding a "comprehensible core" of language the truth-valuations of which uniquely determine the truth-valuations of the whole language. To clarify the nature of this alternative approach let us first introduce some auxiliary concepts.

We shall say that a sentence S is *partially determined* by a class C of sentences, if there is a distribution of truth values among the sentences in C which forces S to have a definite truth value. (If C is $\{S_1, \ldots, S_n\}$, then to say that C partially determines S is to say that there is a necessarily true sentence $S_1^\star, \ldots, S_n^\star \rightarrow S^\star$, where each $X^\star$ is either X or $\neg X$). We shall say that a sentence S is *(totally) determined* by a class C of sentences, if *every* distribution of truth values among the sentences in C forces a definite truth value of S. We call a class of sentences of a language L a *quasibasis* of L if it totally determines every sentence of L.

The definition of the concept of quasibasis says that every truth-valuation of a quasibasis extends uniquely to a truth-valuation of the class of all sentences; hence that the truth-valuation of the whole class is uniquely determined by that of the quasibasis. This means that the study of truth-valuations of the entire language can be reduced to the study of their restrictions to the quasibasis. Every language obviously has some quasibases; this follows from the fact that the whole language itself is its own quasibasis. However, only a nontrivial quasibasis, a quasibasis whose complexity is in some sense remarkably smaller than that of the whole language, can be of interest from the viewpoint of the demarcation of the class of possible truth-valuations.

We shall call a sentence S *independent* of a class C of sentences if S is not partially determined by C. A set of sentences is called *independent,* if every one of its elements is independent of the other sentences of the set. This means that a set of sentences is independent if all the distributions of truth-values among its elements are possible. We call a quasibasis a *basis* if it is independent. Although every language has a quasibasis, there is evidently no guarantee that it has a basis. Let us call a language *based* if it does have a basis; and let us call it *baseless* if it has none.

Notice that a set which is independent can be considered as comprehensible from the viewpoint of truth-valuations: every distribution of truth values among elements is possible. The problem of demarcating the class of its possible truth-valuations is hence in fact—in a sense—solved by finding a basis: there is evidently a one-to-one correspondence between possible truth-valuations and subsets of the basis; the space of possible truth-valuations can thus be identified with the power set of the basis.

So to find a basis is one of the possible ways of solving the problem of the demarcation of the class of possible truth-valuations, and hence of the problem of the *pursuit of meaning.* Hence: *To specify a basis is to specify the space of the possible truth-valuations and hence to account for semantics.* It is an alternative way to handle the task which is otherwise handled by the axiomatic method.

5. Model Theory as the Completion of Basis

Now it seems that we somehow assume that a language must (or at least should) have a basis. If it has none, then we have the impression that the basis must be only somehow "hidden" and we feel urged to "fix up" the language to make the basis visible. The main claim of the present paper is that it is plausible to see precisely this kind of urge for a fix up as that which is constitutive of model theory. But before we discuss the thesis explicitly, let us consider the existence of bases within the most traditional logical systems.

There is evidently a nontrivial quasibasis within the classical propositional calculus: it is constituted by the atomic sentences. The truth value of every sentence is uniquely determined by the truth values of some atomic sentences (it is determined by its atomic subsentences; and this is the reason why the calculus is decidable). Moreover, the truth value of an atomic sentence is independent of those of any other atomic sentences, hence the quasibasis is a basis. The classical propositional calculus thus fulfills our expectations in respect of the existence of basis. The troubles begin when we pass to predicate calculus, i.e. when we introduce the apparatus of quantification.

The atomic sentences of the first-order predicate calculus do not in general constitute a basis: the truth value of a sentence such as $\forall x P(x)$ is not totally determined by those of all the atomic sentences. If all the atomic sentences (especially all the sentences $P(t)$ for any term t) are true, then $\forall x P(x)$ may be true as well as false. However, it is not possible to add $\forall x P(x)$ itself to the basis: it is not independent of the class of atomic sentences, because it cannot be true if some of the $P(t)$'s are false. The point is hence that $\forall x P(x)$ is partially determined by the class of atomic sentences (in particular by those of the form $P(t)$), but it is not totally determined by it.

Now having found out that there is no visible basis within a first-order theory, one feels somehow urged to say that there is a "latent" one: to say that the falsity of $\forall x P(x)$ "means" that there exists a "nameless" individual i which does not have the property P. If we embrace this intuition, then in fact we compensate the lack of a name for i by considering the "metaname" 'i', and the lack of the (false) sentence "$P(i)$" by postulating the (false) "metasentence" 'i has the property P'.

To see what is going on in such a case, imagine the situation where $\forall x P(x)$ is false, but every instance of $P(t)$ is true. The conclusion will be that the universe contains a nameless individual which is not P. This conclusion sounds like an empirical discovery which *explains* the falsity of $\forall x P(x)$: we have found a certain individual with a certain property which *makes* the statement false! However, in fact the existence of the individual results from mere reshaping of the information implicit in the falsity of $\forall x P(x)$ and the truth of all the $P(t)$'s. The conclusion that the universe contains an individual which is not P requires nothing more that the trivial transformation of $\neg\forall x P(x)$ into $\exists x \neg P(x)$; whereas the conclusion that the individual is nameless is the direct consequence of the fact that no $P(t)$ is true. Thus, it is misguided to say that the existence of such an individual in the universe *explains* the truth values of the formulas—for it is only another way of *stating* the truth values (see further Peregrin 1995, Chapter 5).

Now if we accomplish such a compensation (of the lack of names by a supply of metanames) systematically in frames of set theory, then what we reach is model theory. If a $\forall x P(x)$ is false although $P(t)$ is true for every term t, then the universe of the model is to contain such an i that $i \notin ||P||$ (where $||P||$ is the denotation of P, a subset of the universe). This nature of model theory is especially manifest if we consider the famous completeness proofs presented by Henkin (1949, 1950). What Henkin in fact showed is that every ("reasonable") language (or theory) can be conservatively extended to a language with respect to which "there are no nameless objects" (i.e. to a theory which has the property that whenever a $\forall x P(x)$ is false, then there is a t such that $P(t)$ is false).[10] What is important is that to understand Henkin's construction we need not consider interpretations and models at all; it is enough to consider embeddings of language into another

[10] Henkin's procedure incorporates two steps. First, we add names to guarantee that whenever there is a true existentially quantified formula $\exists x.F(x)$, then there is a name t such that $F(t)$ is true. Second, we equate names displaying no difference in behaviour with respect to consequence. (If our system allows for definite descriptions, then we can drop the first step; if it does not include the identity predicate, then we must drop the second one). Then it is possible to assume the one-to-one correspondence between equivalence classes of names and objects and so the relationship between the language and the model may be seen as transparent.

language. For what Henkin's proof really says is that every theory can be conservatively embedded into a theory the atomic sentences of which form a basis. If it were guaranteed that every language does have a basis, then Henkin's proof would lose its import.

This indicates that considering a model of a language may be seen as considering an embedding of the language into another, based language. (Notice that only the case discussed above makes the need for a model theory really urgent: if there were a false $P(t)$ for every false $\forall x P(x)$, then there would always be a trivial model constituted by the language itself.)

6. Modal Logic

The predicate calculus with its overt quantification is not the only case of a well-established calculus which lacks a basis. Modal propositional logic is another instance. The status of the formula $\Box S$ of $S5$ is similar to that of $\forall x P(x)$ of the classical predicate calculus: it is neither totally determined by, nor quite independent of the class of all the atomic sentences. If S is false, then $\Box S$ is forced to be false too, but if S is true, $\Box S$ may be true as well as false.

Whatever differences one may find between $\forall x P(x)$ of the predicate calculus and $\Box S$ of the modal propositional calculus, the model-theoretic solution is surprisingly similar. If S is true (in the "actual world") and $\Box S$ is not, then this is understood as "meaning" that there is a possible world w, different from the actual world, with respect to which S is false. $\Box$ is thus understood as latently quantifying over possible worlds in the same way in which $\forall$ quantifies over individuals. While within the classical predicate calculus the basis is restored on the meta-level as the class of all sentences of the form 'the individual i has the property P', within $S5$ it is restored as the class of all the sentences of the form 'S is true w.r.t. the world w'. 'Ideological' differences aside, $S5$ is quite analogous to the monadic predicate calculus.[11]

This indicates that the pattern underlying the 'pull towards models' is more general than that of the overt quantification. The relevant pattern is that of the 'semi-independent quasibasis': a quasibasis Q with an independent subset B such that B determines the elements of $Q \setminus B$ partially, but not totally. Model theory for such a language then amounts to the embedding of the language into a language in which

[11] If we consider other modal calculi, then the situation is technically slightly more involved (the truth value of $\Box S$ depends on those of the elements of the basis in a complicated way—requiring a relation of "accessibility"), the principle is, however, the same.

B gets extended to B' such that B' is still independent, and totally determines all the elements of $Q \backslash B$. The new items constituting $B' \backslash B$ are then naturally felt as being in a way implicit to those of $Q \backslash B$.

7. Model Theory as Relating Words to Things

This is, of course, an unusual way of conceiving model theory. The usual idea is that model theory depicts the relationships between words and things. Expressions of natural languages, the story goes, are expressions because they are somehow conventionally linked to extralinguistic objects, and model theory does the job of convention for formal languages—it guarantees that expressions of these languages are also linked to something extralinguistic, namely to certain sets. Model theory done for a (fragment of) natural language thus appears to be a direct account of its semantics.

However, it is naive to take this idea at face value. It would be tantamount to what Quine (1969) calls the *museum myth*—to the idea that expressions are stuck on objects as labels in a great museum; and this is far from how language really works. As we said above, it is clear that no theory could succeed in directly relating words to things; it can do nothing over and above relating words to other words. It can give meanings of expressions of a language; but only given a "meta" language which is taken as the unquestioned background. (You cannot put a real dog into your theory to show what 'dog' means—you inevitably have to employ such or another representation.) And our natural language is the *ultimate* metalanguage, we cannot step out of it. There is no hope of doing model theory for our natural language, and hence of model-theoretic answers to philosophical questions about language.[12] This is why we stress that model theory should be primarily understood as amounting to extending languages, to making baseless languages based.

On the other hand, philosophy is only enlightened common sense (as Popper once remarked), and therefore it would be unwise simply to ignore the intuition which equates model theory with semantics in the pre-theoretic sense: rather, we should seek the rational core of this intuition. Perhaps there might be a way of reconciling the traditional view of model theory (as the words-things question) with the one proposed here (as the words-words question). There are languages, we might say, which are "semantically transparent", whose relation to the world is straightforward, and model theory can be considered as relating lan-

[12] I discuss this predicament in detail in Peregrin (1995, Chapter 11 and 1996a) but see Hintikka 1988 for a defense of the contrary claim.

guage to the world because it relates semantically non-transparent languages with semantically transparent ones.

Moreover, it seems to be plausible to conclude that for a language to be semantically transparent means to have a basis. Henkin's proof shows that establishing a basis means having the job of model theory essentially done. So, after all, there is a sense in which relating a baseless language to a based one, embedding the former into the latter, can be understood as establishing the link between words and things, i.e. as doing semantics in the direct sense. Is not model theory, therefore, the "direct" theory of semantics after all?

The problem is that what is understood as furnishing model theory for a given language is in fact usually *not* furnishing a based language into which the language could be conservatively embedded. On the contrary: in the nontrivial cases it is impossible really to furnish a based extension, so model theory is left with merely *postulating* the existence of such an extension, and it is ultimately this point which precludes taking model theory as a direct theory of semantics. However, to articulate this intelligibly, we must examine more closely the point at which the nature of model theory comes into the open.

8. The Nature of the Quantifier

From what has been said it follows that the lack of basis in the predicate calculus is somehow connected with quantification; let us thus analyze the notion of the quantifier. An excellent study of the nature of quantification in formal languages and of its early development has been presented by Goldfarb (1979). Commenting on the views of Bertrand Russell the author writes (p. 354):

> ... the incorporation of quantifiers enables us to have principles that deal with infinity .... The meaning of our signs is only finitely complex (...). By the use of signs with such an intelligible meaning, we finite human beings reason about the infinite. But the nature of only finitely complex meaning of the quantifier cannot be given by some explanation from without. No such explanation can have any force. Rather, the meaning of quantification is shown by how we employ the signs, that is, by the logical rules of inference.

Hence for Russell, as for Frege before him, quantification was a means of reasoning about the infinite. It was an indirect means of such reasoning; however, this is not to be understood as claiming that Russell would have entertained the possibility of a reasoning about the infinite that would be direct. On the contrary, the conviction was, as Goldfarb aptly points out, that "the meaning of a quantifier cannot be given from without".

It was Tarski who took the essential point of departure in this respect. It is not by chance that for the argument Tarski has used to document the insufficiency of the Frego-Hilbertian proof-theoretic approach and to substantiate the need of an alternative, model-theoretic approach, quantification was crucial. In the paper which is usually considered as the point of departure of model theory Tarski (1936) argues that the proof-theoretic approach is not able to account for the fact that the sentences 'n has the property E' for every natural n together entail the sentence 'All natural numbers have the property E'. Let us reproduce the relevant part of Tarski's paper:

Schon vor mehreren Jahren habe ich ein Beispiel, übrigens ein ganz elementares, einer derartigen Theorie gegeben, die folgende Eigentümlichkeit aufweist: unter den Lehrsätzen dieser Theorie kommen solche Sätze vor wie:

A_0. *0 besitzt die gegebene Eigenschaft E,*

A_1. *1 besitzt die gegebene Eigenschaft E,*

u.s.w., im allgemeinen alle speziellen Sätze der Form:

A_n. *n besitzt die gegebene Eigenschaft E,*

wobei 'n' ein beliebiges Symbol vertritt, das eine natürliche Zahl in einem bestimmten (z.B. dekadischen) Zahlensystem bezeichnet; dagegen läßt sich der allgemeine Satz

A. *Jede natürliche Zahl besitzt die gegebene Eigenschaft E,*

auf Grund der betrachteten Theorie mit Hilfe der normalen Schlußregeln nicht beweisen. Diese Tatsache spricht, wie mir scheint, für sich selbst: sie zeigt, daß der formalisierte Folgerungsbegriff, so wie er von den mathematischen Logikern allgemein verwendet wurde, sich mit dem üblichen keineswegs deckt. Inhaltlich scheint es doch sicher zu sein, daß der allgemeine Satz A aus der gesammtheit aller speziellen Sätze $A_0, A_1, \ldots, A_n, \ldots$ im üblichen Sinne folgt: falls nur alle diese Sätze wahr sind, so muß auch der Satz A wahr sind.[13]

Tarski's proposal is to improve this via the introduction of his, model-theoretic, notion of consequence. His proposal runs as follows:

[13] "Some years ago I gave a quite elementary example of a theory which shows the following peculiarity: among its theorems there occur such sentences as: A_0. *0 possesses the given property P,* A_1. *1 possesses the given property P,* and, in general, all particular sentences of the form A_n. *n possesses the given property P,* where 'n' represents any symbol which denotes a natural number in a given (e.g. decimal) number system. On the other hand the universal sentence: A. *Every natural number possesses the given property P,* cannot be proved on the basis of the theory in question by means of the normal rules of inference. This fact seems to me to speak for itself. It shows that the formalized concept of consequence as it is generally used by mathematical logicians, by no means coincides with the common concept. Yet intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true."

Die Aussage X *folgt logisch* aus den Aussagen der Klasse K dann und nur dann, wenn jedes Modell der Klasse K zugleich ein Modell der Aussage X ist[14].

Let us analyze Tarski's argument in detail. What he is urging is in fact the reduction of the truth of the sentence 'All natural numbers have the property E' to the truth of the infinite set of sentences containing the sentence 'n has the property E' for every natural n. However, what does the truth of such an infinite set really amount to? To say that a definite sentence is true may be considered a kind of paraphrastic way of asserting the sentence itself. To state that the sentence '1 has the property E' is true is to say that 1 has the property E. However, to say that the sentence 'n has the property E' is true for every n is no paraphrase: nobody can succeed in asserting an infinite number of sentences directly.

Let us suppose we are to put down an instance of consequence. If the instance amounts to the fact that something follows from the truth of the sentence '1 has the property E' (i.e. from the fact that 1 has the property E), we may consider it to follow either from the ("meta")sentence '"1 has the property E" is true' or from the sentence '1 has the property E'. However, if we say that something follows from the fact that the sentence 'n has the property E' is true for every n, then there is no alternative: we have no choice but '"n has the property E" is true for every n'.

We cannot list all the sentences of an infinite set, so the articulation of the truth of all its members inevitably boils down to the truth of such or another "meta"sentence. What Tarski himself in fact says is that it is impossible to account for the fact that the *"meta"sentence* '"n has the property E" is true for every natural n' entails the *sentence* 'All natural numbers have the property E'. And the employment of the "meta"sentence is no shortcut: there is no way of avoiding it.[15]

However, what is the real difference between '"n has the property E" is true for every natural n' and 'All natural numbers have the property E'? To state '$P(x)$' *is true for every* x is nothing other than to state $P(x)$ *for every* x and hence, if we stick to the usual predicate-calculus formalism, to state $\forall x P(x)$; the only substantial difference is that (from the viewpoint of the calculus) the first sentence is a *meta*sentence, whereas the last one is a sentence proper. What makes the difference between '$P(x)$' *is true for every* x and $\forall x P(x)$

[14] "The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X."

[15] The whole problem is in fact only a special case of the general problem of the infinite; see Peregrin 1995, Chapter 9.

is that moving from the former to the latter we cross the boundary between that which we take as metalanguage and that which we take as language. However, to say that between two sentences, such as '$P(x)$' *is true for every* x and $\forall x P(x)$, there lies a boundary is not an empirical discovery, it is rather something we impose on the language to treat it in a way we want to; the concept of metalanguage is in fact a methodological, not a directly empirical one. It is an outcome of Tarski's conceptual framework, and we can hardly accept that two things should be different only because we have decided to classify them into different compartments.

9. The Nature of Model Theory

So let us return to our considerations concerning the nature of model theory. The point made at the end of the previous section can now be articulated as follows: model theory is (usually) not an embedding of a based language into a baseless "meta"language; it is only *about* such an embedding. And in fact, where the explicit presentation of the extended language actually is a possibility, i.e. where it suffices to introduce a finite number of new ("meta")constants, there the situation is trivial.

Let us consider the language of arithmetic. Furnishing model theory for this language amounts to introducing the infinity of numbers. But obviously we cannot introduce all the numbers explicitly (by listing them); we can at most *stipulate* the existence of the domain of all the numbers. Within the original language, there might be a valid sentence $\forall x P(x)$. Model theory is incapable of explicitly furnishing the infinite number of "meta"statements *1 has the property* P, *2 has the property* P, etc.; it can only cover all these "meta"statements employing the "meta-meta"statement *every* n *of the domain has the property* P.

Model theory for the predicate calculus can be seen as motivated by a desire to reduce quantificational truth to truth *simpliciter* of infinitely many instances. However, this cannot be accomplished by model theory; at least not in nontrivial cases. Model theory in these cases is not the embedding of a baseless language into a based "meta"language; but rather the translation of the baseless language into another baseless language, which is purported to be the "meta"language of a based "meta"language of the original language. Thus, far from actually reducing quantificational truth to truth *simpliciter* it is only stipulating such a reduction.

Does all of this mean that model theory is nonsensical or vacuous? Not at all. Model theory amounts to relating two theories—the object language theory and a particular theory of sets. Such a relation can be interesting and can give rise to a respectable mathematical discipline; however, it should not be considered as relating expressions of the object language with things. The ultimate reason is not only that it amounts to a relation between two languages and not between a language and reality—we know that no *theory* could do better. The decisive reason is that it is not capable of relating a baseless language, such as that of a first-order theory, with a language that would be based and hence "semantically transparent".

This is to say that the idea of model theory, however interesting a mathematical discipline it might give birth to, is misguided as soon as it is considered to be capable of directly accounting for semantics of natural language and of explaining such pre-theoretic notions as truth, consequence or quantification. The underlying error consists in the failure to see the real nature of the infinite. A sentence is assumed to allude to infinity if it entails an unlimited number of certain consequences, and this is quite sound; what is not sound is to assume that this allusion to infinity can be made explicit. In that we manipulate our finite signs in a certain way, we do in a sense attain to the infinite; however, such an indirect way is in fact *the only* way in which we can really attain to it.

10. Logical Atomism

Our considerations have been based on the assumption that it somehow belongs to the nature of language, or at least of a "semantically transparent" language, that it has a basis. Where does this feeling come from?

The reason seems to be that we take for granted that the world has something as a basis, and that hence the language that faithfully accounts for it must have a basis, too. We usually consider the world as a kind of structure made of some basic building blocks, of what Putnam (1984) aptly calls the *Furniture of the Universe.* This idea found its most straightforward expression within several influential philosophical doctrines of the present century, namely within Russell's *logical atomism,* Carnap's *logical empiricism* and the philosophical doctrine of Wittgenstein's *Tractatus.* All these approaches share the conviction that in the foundation of our language there is an underlying basis of "atomic", "protocol", or "elementary" sentences; and they postulate the principal coincidence of the order of language with that of the world and that of our knowledge, so that they see the existence of the basis of atomic statements from which all other statements can be deduced as tantamount to the existence of a basis of atomic facts from which all other facts can be constituted and to the existence of a basis of atomic pieces of our knowledge ("observations") from which all our knowledge can be inferred. Logical empiricists even discussed the nature of concrete bases;[16] but the fact is that no definite basis has been established for natural language.

Nevertheless, if we accept that our world actually is as Russell, Carnap and the early Wittgenstein saw it, then the conclusion seems to be inevitable, that our natural language really is based—for otherwise we would be left with the conclusion that there are some parts of the world we are principally not able to speak about. Being a language—in the genuine sense of the word—thus seems synonymous with being a *based* language.

However, there is no reason to believe that every part of language is necessarily itself a language; similarly we do not believe that every part of, say, an algebra must itself be an algebra. It seems to be the case that certain closure properties are essential for the concept of language, and that besides syntactic closure properties certain semantic or truth-valuational closure properties are also relevant. If this is right, then what looks like a language may sometimes be a mere pseudolanguage: it may be a fragment of a language whose boundaries are not chosen properly.

A formal language can be considered as metaphorically "cut out" from its metalanguage; all languages being ultimately "cut out" from natural language, from the "metalanguage of all metalanguages". Our conviction that there should be a basis within every language indicates that a language in which there is none may be in some sense be faulty, that it has been erroneously gerrymandered. Model theory can then be helpfully considered as an attempt to restore the 'right' boundaries, to cut out the language anew without the error, to round off the original language.

Hence model theory, viewed from this perspective, can be considered as an attempt at a readjustment of a language cut out in a wrong way, an attempt to realign the language. If this is acceptable, then model-theoretic semantics of natural language may be considered as an attempt to realign natural language; however, out of what could natural language be considered to be improperly cut out? Logical atomism may suggest the idea that it is not our natural language, but the world itself that should be taken as the ultimate metalanguage—but it is hard to see how this idea could be substantiated.

Indeed it is highly debatable that the atomistic view of our world is adequate. The linguistic turn taught us to see the structure of the world in terms of the structure of the language we use to cope with

[16] From Carnap 1928 on.

the world, rather than the other way round. Logic shows us that there are many sentences that can be seen as following from other, simpler sentences; and this may be understood to mean that there are many facts that can be seen as constituted out of other, simpler facts. But logic does not guarantee that there is a basis of "atomic" statements from which all other statements would follow—this is a philosophical doctrine; and it is worth considering whether it is really so plausible as it usually seems. Giving up atomism means embracing holism and consequently pluralism (if there is no definite basis, then the decision what to take as basic and what as inferred is ours), and this may seem to be incompatible with the rigor of logic; but scholars like Quine have made it seem plausible that logic may be of a piece with holism.[17]

11. Conclusion

We have tried to indicate that it is preferable to see semantics as primarily a matter of the consequence relation, rather than as a matter of Tarskian assignment of set-theoretic objects. This supports a perspective on logical calculi the reverse of that which is widespread nowadays: the perspective of viewing proof-theory as more basic than model-theory. From this vantage point it is more appropriate to see, for example, incompleteness not as a failure on the part of axiomatics to capture a model-theoretically delimited class of sentences, but rather as a failure on the part of model-theory. That second-order predicate calculus is incomplete does not mean that its axiomatics is unsatisfactory, but rather that its classical model theory is inappropriate, that it is Henkin's models that are the "right" ones.

An orthodox model-theoretician may consider the majority of such considerations irrelevant; what she is after is a mathematical theory and model theory has proven itself to be an interesting mathematical theory. Completeness and incompleteness results are mathematical results which need not invoke any considerations of what is primary or what is appropriate. This is, indeed, true; however, there are not many model-theoreticians consequently orthodox in this sense. The relation between an expression and its model-theoretic interpretation is often understood not as a relation between two abstract entities of a mathematical theory, but rather directly as a relation between an expression and its meaning. Model-theory is increasingly understood as semantics in the intuitive, pre-theoretic sense of the word. While Tarski was—despite some problematical claims—principally aware that it is tricky to see model theory otherwise than in terms of extending languages (or

[17] See also Peregrin (In preparation).

mapping languages on their metalanguages), many of his followers, be they mathematicians or theoreticians of language, lack this awareness and mistake model theory for metaphysics.

Contrary to Montague's popular claim that there is no substantial difference between natural and formal languages we are convinced that there is quite a fundamental difference which manifests itself also in connection with model theory: it makes sense to make a model theory of a fragment of natural language or of a formal counterpart of such a fragment (and so to possibly repair its "false alignment"), there is, however, no clear sense in applying model theory to an entire natural language.

The idea that model theory is semantics because it deals with assignments of objects to expressions and because semantics is also a matter of objects connected with expressions, is misguided. *Model theory can indeed be considered a theory of semantics, but not in force of being an imitation of the 'real'* denotandum/denotatum *relation, but rather in force of being an account for consequence.*[18]

References

Carnap, R. 1928. *Der logische Aufbau der Welt.* Berlin.

Davidson, D. 1984. *Inquiries into Truth and Interpretation.* Clarendon Press, Oxford.

Goldfarb, Warren D. 1979. Logic in the twenties: The nature of the quantifier. *Journal of Symbolic Logic*, 44, pp. 351–368.

Henkin, L. 1949. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14, pp. 159–166.

Henkin, L. 1950. Completeness in the theory of types. *Journal of Symbolic Logic*, 15, pp. 81–91.

Hintikka, J. 1988. On the development of the model-theoretic viewpoint in logical theory. *Synthèse*, 77, pp. 1–36. Reprinted in J. Hintikka, *Lingua Universalis vs. Calculus Ratiocinator (Selected Papers 2)*, Kluwer, Dordrecht, 1997, pp. 104–139.

Montague, R. 1974. *Formal Philosophy: Selected Papers of R. Montague,* edited by R. Thomason. Yale University Press, Yale.

Peregrin, J. 1994. Interpreting formal logic. *Erkenntnis*, 40, pp. 5–20.

Peregrin, J. 1995. *Doing Worlds with Words.* Kluwer, Dordrecht.

[18] See also Peregrin 1994, where a similar conclusion was reached starting from considerations of another kind, namely from considerations of different meanings of the term *semantic interpretation*.

Peregrin, J. 1996a. Logic and the pursuit of meaning. *Logic and Linguistic in the Pursuit of Meaning (Studies in Logic and Philosophy—Department of Logic preprints).* Available via the author's homepage: http://www.cuni.cz/∼peregrin.

Peregrin, J. 1996b. Linguistics and philosophy. *Logic and Linguistic in the Pursuit of Meaning (Studies in Logic and Philosophy – Department of Logic preprints).* Available via the author's homepage: http://www.cuni.cz/∼peregrin.

Peregrin, J. In preparation. When meaning goes by the board, what about philosophy? *Proceedings of ANALYOMEN 2.*

Putnam, H. 1984. After Ayer, after empiricism. *Partisan Review*, 51. Reprinted as "After empiricism" in *Post-analytic Philosophy,* edited by J. Rajchmann and C. West, Columbia University Press, New York, 1985, pp. 20–30.

Quine, W.V.O. 1953. *From a Logical Point of View.* Harvard University Press, Cambridge, Massachusetts.

Quine, W.V.O. 1960. *Word and Object.* MIT Press, Cambridge, Massachusetts.

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

Schlick, M. 1932. The future of philosophy. *College of the Pacific Publications in Philosophy*, 1, pp. 45–62.

Tarski, A. 1936. Über den Begriff der logischen Folgerung. *Actes du Congrés International de Philosophique Scientifique*, 7, pp. 1–11. Translated as "On the concept of logical consequence", in Tarski 1956, pp. 409-420.

Tarski, A. 1956. *Logic, Semantics, Metamathematics.* Clarendon Press, Oxford.

Wittgenstein, L. 1953. *Philosophische Untersuchungen.* Blackwell, Oxford. Translated as *Philosophical Investigations*, Blackwell, Oxford, 1953.

Wittgenstein, L. 1977. *Vermischte Bemerkungen,* edited by G.H. von Wright. Suhrkamp, Frankfurt. Translated as *Culture and Value*, Blackwell, Oxford, 1980.

Department of Logic, Institute of Philosophy
Academy of Sciences, Czech Republic