On the ontology of model-theoretic semantics

Roger VERGAUWEN

1. *Introduction: Intentionality*

The use of model-theoretic methods in linguistics and philosophy has given rise to new perspectives in the study of 'traditional' philosophical problems, among which is the problem of *intentionality.* Roughly speaking, an intentional phenomenon "is one that is *about something else,* even if that something else does not exist" (Bealer '82, 12).

Natural language and its use can be considered to be such an intentional phenomenon. Words and sentences are meant to be *about* something, e.g. a state of affairs in the world. Although it is the basic goal of model-theoretic semantics "to specify how language connects with the world, in other words to explicate the inherent "aboutness" of language" (Dowty '81, 5), this 'aboutness' or intentionality of language is not so easily accounted for by the theory.

In "Minds, Brains, and Programs" (Searle '80) Searle sets up a thought experiment, the 'Chinese-Room Experiment' to show that intentionality cannot be programmed into a computer viewed as an instantiation of a formal system. The aim of this paper is to point out how a model-theoretic semantics considered as a formal system can be said to account for the intentionality of natural language and how this affects the overall ontology of the theory. Whereas for Searle intentionality is a biological property (Searle '80, 424), I think that in the case of natural language a logical counterpart for this phenomenon can be constructed.

2. *Model-theoretic semantics and the Grelling paradox*

The *model* in the model-theoretic semantics developed by Richard Montague (Montague '73, 258) is defined as a quintuple ($\mathfrak{A}$)

(1) $\mathfrak{A} = (A, I, J, \leqslant, F)$

The *ontology* of (1) contains a set of individuals (A), a set of possible worlds (I), a set of moments of time (J), an ordering relation on (J), $\leqslant$, and a function F assigning *meanings* to the 'non-logical constants' or words of the language under consideration. These 'meanings' are *intensions*, functions from possible worlds and times to denotations or extensions of the appropriate kind. The intentionality of language however can, not be accounted for by the intension-assignment function F. For, if one would want to identify the 'aboutness' of a word or a sentence with its intension considered as a procedure to determine the extensions unambiguously in every possible world, the *Grelling paradox* would get in the way: suppose we define an adjective to be *autological* if and only if it can truly be applied to itself and stipulate that otherwise the adjective is to be called *heterological*. Then, the paradox that emerges when we want to know where the adjective 'heterological' itself belongs in the 'autological-heterological' distinction shows how the extension of 'heterological' cannot unambiguously be defined for each and every of its arguments.

Formally, this situation can be translated into an incomputability result by which it is possible to show that the function F from (1) cannot be constructed so as to fully determine the intensions of the non-logical constants (Vergauwen '86, 298).

A complete specification of what linguistic elements are about is therefore impossible and the assignment of a 'meaning' to the meanings (intensions) cannot be effected through the model. This does, however, not mean that there would be no intentionality in the sense of 'the meaning of meaning(s)'.

3. *The logical basis of intentionality*

Gödel's first incompleteness theorem posits the existence of a true, though undecidable, number-theoretic formula. Roughly speaking, this formula asserts about itself that it cannot be proved, and its existence causes the arithmetic to be incomplete if it is consistent. The *incompleteness of model-theoretic semantics* is in part related to the fact that the semantical metalanguage used is a *higher-order intensional logic* and does therefore contain elementary arithmetic, but it can also be motivated from a more 'semantical' point of view in the following way. In the intensional logic there is a formula expressing that intentionality cannot be accounted for by the model. This formula says that there is no 'meaning of meaning' in the sense of intentionality. It expresses a true property of intensions, viz. that they cannot be proven to be about anything. In other words, it expresses the *absence of intentionality* for intensions within the model. As it stands, however, this formula appears to be an *undecidable* one (Vergauwen '86, 302), and therefore the semantics is incomplete if it is consistent. Let this formula be represented by σ and let Σ be the set of formulae from the intensional logic. Then the following holds:

(2a) If $\sigma \nvdash \Sigma$ then $\Sigma \cup \{\neg \sigma\}$ is consistent

(2b) Every consistent set of sentences has a model.

This means that it is possible to add the negation of the formula expressing the absence of intentionality to the system and to find a model for the new system thus constructed. σ is true in the *standard model* for the semantics, but by (2a-b) there is a *non-standard model* in which it is false. What is false in this non-standard model is that there is no intentionality for the intensions. The standard model for a semantics of natural language can be viewed as 'the world' in any good sense of the word. But, of course, it is the world as *qualified* by a particular definition of intentionality in which the nature of the correspondence relation between language and the world is fixed. A non-standard model such as the one in which the absence of intentionality is denied, is defined by a different specification of intentionality. So, what is considered to be intentionality in the standard model, can only be talked about in a non-standard model. The ensuing incompleteness and non-categoricalness of the semantics as a logical theory *simulates* the intentionality-phenomenon. For any model, the specification of what is to count as intentionality within this model belongs to a model that is non-standard to the first one. It is the logical gap created by the impossibility of defining the intentionality for a model within the model itself and the consideration that such a definition can be given in a non-standard model which act as an analogue for the 'aboutness' of language. In other words, if the words and the world would be in a relation of strict isomorphy, words could, in a sense, not even be said to be *about* the world. Our theory has to predict that such a relation of isomorphy does not exist, which it does by showing how the meaning-relation itself in which a definition of intentionality is contained can never be fixed once and for all. Metaphorically speaking, the theory says that 'there is always something more' that it does not fully grasp.

There is an infinite hierarchy of models that can be shown to be connected by their respective intentionality-relations in the sense that one model can 'talk *about*' what it is for the preceding 'lower' model to exhibit intentionality. The 'aboutness' of the intensions is an effect of the incompleteness and non-categoricalness of model-theoretic semantics and can only be fixed in a relative way, because it is tied to a hierarchy of models (Vergauwen '86, 316).

4. *Intentionality-predicates and their semantics*

Intentionality is first and foremost an inherent characteristic of natural language, but the specification of the *kind of intentionality-relation* for a particular fragment of natural language may be effected by using specific predicates. These predicates can be called *intentionality-predicates* and are the expression of intentional states. Among these predicates, *verbs of propositional attitude* such as 'believe' or 'think', have been bothering linguists and philosophers alike for many years. It is not clear what is to count as the semantic objects of these verbs and, moreover, these verbs are known to cause serious problems for the model-theoretic enterprise.

In the 'traditional' Montague-type semantics, verbs of propositional attitude are defined as relations between propositions and sets of individuals, or 'individual concepts' for that matter. With these verbs, at least two kinds of problems crop up. First, elements such as words or sentences are called *logically equivalent* if they have equal intensions, and are therefore intersubstitutable *salve veritate*. However, this generalization breaks down in contexts with verbs of propositional attitude. Consider the following sentences:

(3a) Thales of Miletus believed that *two plus two equals four.*
(3b) Thales of Miletus that *the square root of two is irrational.*

The complement sentences in (3a-b) express mathematical tautologies, so they are always true in every possible world. Therefore these sentences express identical propositions (sentence-intensions), but it is obvious that they are not interchangeable salva veritate. It is not because Thales believed (3a) that he therefore would also believe (3b). Second, it can be shown that with these verbs intensionally equivalent words such as 'eye-doctor' and 'oculist' are not any longer synonymous and rigid designators loose their 'rigidity' (Vergauwen '86, 308).

In the aforementioned article I tried to cope with problems such as these by taking advantage of the notion of a *supervaluation* developed by Van Fraassen: supervaluations can be used to distinguish between elements that are otherwise intensionally equivalent. Alongside with intension and extension *supervaluated meanings* are introduced. Specifically, in the supervaluation approach one considers the set of 'classical' or 'general' valuations (v_g) over the model in which extensions are assigned to the elements in the different possible worlds 'as they should be', and one adds 'idiosyncratic valuations' (v_i) induced by the verbs of propositional attitude which may differ from the general valuations. Next, a 'supervaluation' (v_s) is defined over the set of classical and idiosyncratic valuations and the *truth-value gaps* that turn up in the supervaluation are used to prevent illicit substitution of otherwise equivalent or synonymous element (Vergauwen '86, 311).

The more profound cause of the problems of logical equivalence, rigid designation and synonymity with verbs of propositional attitude resides, in my opinion, in the *nature of the function* that corresponds to them in the semantic theory. The principle upon which it rests can be found in Cantor's idea of a diagonal function.

To illustrate this idea, let us imagine the set of all one-place number-theoretic functions over the positive integers. We shall say that a function is *effectively calculable* if there exists a definite algorithm that enables us to compute the functional value corresponding to any given value of its arguments. These algorithms can be expressed as sets of instructions and can be listed in a certain way (Davis '73, xvii). With each positive integer, i, there is associated in this list the i-th set of instructions, E_i, and the function associated in this way with E_i is called $f_i(x)$. Now it is possible to define the following number-theoretic function:

(4) $g(x) = f_x(x) + 1$

Though (4) is a perfectly good function it is not effectively calculable, because, if it were, this would mean that $g(x) = f_{i_o}(x)$ for some integer i_o:

(5) $f_{i_o}(x) = f_x(x) + 1$

This function would have to hold for all values of *x*, and particularly for $x = i_o$, yielding

(6) $f_{i_o}(i_o) = f_{i_o}(i_o) + 1$

But this is impossible, as no number equals itself plus one.

The diagonalisation technique can also be used in model-theoretic semantics in the following way:

words can be viewed as functions over possible worlds in which their extensions are assigned according to the valuations v_g, v_i, and v_s over the model. The following diagram (7) may help to visualize this situation:

(7)

	PW_0/V_0	PW_1/V_1	PW_2/V_2	....	PW_{n-2}/V_i	PW_{n-1}/V_s	PW_n/V_n
words							
$f_0(x)$	$f_0(PW_0/V_0)^*$	$f_0(PW_1/V_1)$	$f_0(PW_2/V_2)$	.....	$f_0(PW_{n-2}/V_i)$	$f_0(PW_{n-1}/V_s)$	$f_0(PW_n/V_n)$
$f_1(x)$	$f_1(PW_0/V_0)$	$f_1(PW_1/V_1)^*$	$f_1(PW_2/V_2)$	.....	$f_1(PW_{n-2}/V_i$	$f_1(PW_{n-1}/V_s$	$f_1(PW_n/V_n)$
$f_2(x)$	----------------	----------------	$f_2(PW_2/V_2)^*$	.....	----------------------	----------------------	----------------
.							
.							
.							
$f_{n-2}(x)$	----------------	----------------	----------------	.....	$f_{n-2}(PW_{n-2}/V_i)^*$	----------------------	----------------
$f_{n-1}(x)$	----------------	----------------	----------------	.....	----------------------	$f_{n-1}(PW_{n-1}/V_s)^*$	----------------
$f_n(x)$	----------------	----------------	----------------	.....	----------------------	----------------------	$f_n(PW_n/V_n)^*$

(Vergauwen '86, 314)

In (7), the *extensions* are the rows corresponding to the functional values of the words $f_n(x)$ in the distinct possible worlds PW_n. In the above diagram, the extensions of the n-th words in the n-th worlds are marked with an asterisk. This is done to show that, in these worlds, it is the *set of extensions* that functions as the denotation of the word. The reason for including the *meaning of a word* – considered as the set of its extensions within the model – as a separate entity among the extensions, is the following. Verbs of propositional attitude can change the usual denotations of the words or sentences in their scope. They *generate relations* between linguistic units and things these elements are about thay did not exist before and therefore affect the *meanings*. To be able to talk about these meanings, they are included in (7) in the way indicated.

The function corresponding to the verbs of propositional attitude is defined over (7) in the following way:

(8) $g(x) = f_x(x)$ and R

This function instructs us to take the x-th word, to apply it to its x-th argument, the x-th possible world, and then to *reshuffle* (R) the value of the function on this argument. From the manner of construction of (7), it is clear that the function defined by (8) reshuffles the *meaning* of the words, as it acts on the x-th argument of the x-th word, and this is exactly where the extension is marked with an asterisk. In the same way as was the case with (4) it can, now, be shown that the function represented in (8) is not effectively calculable, because no *word's meaning equals itself and at the same time something else induced by the R-operation* (Vergauwen '86, 315). Therefore, metaphorically speaking, this diagonalisation operation brings us out of the usual meaning-assignment and its accompanying intentionality-relation.

5. *Model-theoretic ontology: the case for non-standard meanings*

Coming back to what has been said in part 3 on the nature of intentionality, it could, now, be argued that there is, within the semantics, a formula expressing the verbs of propositional attitude. It would represent a true property of word-meanings, viz. that *there is no proof for the fact that some word's meaning equals itself and at the same time something else induced by R.* This formula would be undecidable because, loosely speaking, it asserts of itself that it cannot be proved and neither can its negation, if the semantics is to be consistent.

In a non-standard model, however, there would be objects that satisfy the negation of this formula and these could be called *non-standard meanings.* Verbs of propositional attitude redifine models and meanings in all kinds of ways, creating new intentionality-relations and giving rise to a hierarchy of models. Non-standard meanings can be viewed as *ordered pairs* consisting of the denotations assigned by the standard-model and the ones assigned by a non-standard model (Vergauwen '86, for an application).

More importantly, the foregoing forces us to change the ontology of the model presented in (1). By the nature of the operations and functions described by intentionality-predicates such as verbs of propositional attitude, it turns out that the semantic objects of these predicates are, in a sense, the models themselves as a representation of the meanings. With these verbs the meaning is *folded back upon itself* and this permits changes in meaning and switches from one model to another. The upshot of all this is a modification in the ontology of the model in the following way:

(9) $\mathfrak{A} = (A, I, J, \leqslant, F, \mathfrak{A}^{*})$

The necessity of including a *representation of the model*, $\mathfrak{A}^*$, within the ontology of the model itself should be evident from what has been said on intentionality-predicates and their semantics. Moreover, it also derives its interest both from the opportunity it gives to describe the semantics of a large class of linguistic elements in which the verbs of propositional attitude are included and from the philosophical advantage to be gained from a *logical* elucidation of the concept of intentionality.

Research Assistant at the national Fund for Scientific Research
Katholieke Universiteit Leuven — Roger VERGAUWEN
Hoger Instituut voor Wijsbegeerte
Kardinaal Mercierplein 2
3000 Leuven
Belgium

REFERENCES

[1] Bealer G. (1982): *Quality and Concept* (Oxford).
[2] Davis M. (1982): *Computability and Unsolvability* (New York).
[3] Dowty D., Wall R., Peters S. (1981): *Introduction to Montague Semantics* (Dordrecht).
[4] Montague R. (1973): "The proper Treatment of Quantification in ordinary English". In: Thomason (ed.), pp. 247-270.
[5] Searle J. (1980): "Minds, Brains, and Programs": *The behavioral and Brain Sciences* 3, pp. 417-457.
[6] Thomason R.H. (ed.) (1974): *Formal Philosophy: selected papers of Richard Montague. Edited and with an introduction by Richmond H. Thomason* (New Haven-London).
[7] Vergauwen R. (1986): "How to do Things with Worlds: Intentionality and the Ontology of Model-Theoretic Semantics": *Logique et Analyse*, 115, pp. 297-320.