MEANING AS
A SET-THEORETIC OBJECT

A Gentle Introduction to the Ideas
Behind Formal Semantics

Jaroslav Peregrin
Institute of Philosophy
Czech Academy of Sciences

Table of Contents

1 Meaning as an object

1.1 Formal semantics

Some fifty years ago, semantics of natural language started a new era. The harbingers of the new period were people operating on the boundary between logic, philosophy and linguistics, such as Montague (1974), Lewis (1972) or Cresswell (1973), who went for the reconstruction of meanings as set-theoretic objects. This provided for a very neat picture of the system of meanings reflecting (or being reflected by) the system of linguistic expressions. How was all of this accomplished, and was this a genuine breakthrough in semantics? (It is fair to note that by far not all people engaged in the semantics of natural language were impressed by this kind of formal semantics.)

Nowadays, this breakthrough is far behind us and we can look back at it with the benefit of hindsight. Not that the paradigm of formal semantics would fade away; analyses of various aspects and phenomena of natural language carried out within its framework keep appearing (regularly, for example, in journals like *Linguistics and philosophy* and *Natural language semantics*). But the heated debates of its conceptual foundations are long over now, formal semantics is taken as an established (though sometimes slightly heterogeneous) framework that can provide space to discuss semantics of natural language, not as something to be questioned or challenged. This distance makes it possible to see the foundations of formal semantics in a clear light and to assess its merits.

The roots of this approach to meaning go back to the early analytic philosophy with its conviction that natural language is treacherous and that we must see through its apparent surface structures to its logical forms which determine the contents of its sentences. The paradigmatic analysis of the sentences with definite descriptions due to Russell (1905) tell us that a sentence like

The king of France is bald

does not necessarily tell us that an individual, the king of France, has a property, baldness. Rather it tells us something much more complicated, which Russell captures by the following formula:

$$\exists x(\boldsymbol{KF}(x)\wedge\boldsymbol{B}(x)\wedge(\forall y(\boldsymbol{KF}(y)\rightarrow(y=x))))$$

It is only if we realize this, Russell insists, that we are clear about what the sentence really says - about its meaning. It follows, according to Russell, that the existence of the king of France is not a condition of the meaningfulness of the sentence, but just part of what sentence asserts; so if this is not the case the sentence is simply false. This analysis also provides for the ambiguity of the negation of the sentence, while the negation sign can be put in front of the whole sentence, but also in front of the predicate ***B*** etc.

Whatever we think about the merits of this analysis, we see that it does not tell us explicitly what the meaning of the sentence is, not to speak about the meanings of its parts, such as the definite article *the*. One possible approach to a further development of this approach to language was to try to capture the meanings explicitly. And some ideas of Frege, buttressed by the development of set theory, appeared to offer a way of how to accomplish this. The way was hacked through especially by Carnap (1942; 1947), followed by Montague & comp.

The outcome of this new approach was, for example, capturing the meaning of a sentence as a class of possible words in which the sentence is true, or capturing the meaning of the definite article as something like the function from sets to individuals which maps all singletons on their single members and all other sets on nothing. So here we have the promise of answering the question about the nature of meaning, wholly explicitly, by presenting meanings as set-theoretical objects.

This approach to meaning has smuggled in the assumption that meanings are objects. In fact, this assumption appears quite trivial, for what else could a meaning be than a kind of object?

Imagining linguistic meanings (*viz*. that which makes a mere sound or scribble into a meaningful expression) as objects comes naturally. Is not the relationship between a word and its meaning akin to t relationship between a proper name and its bearer? Yet when we start to think about words other than nominal phrases, this idea becomes ever blurrier. What kind of objects, represented by expressions, could meanings be?

Several answers spring to mind:

One possibility is that meanings are objects of the spatiotemporal world. This reinforces the parallel between proper names and linguistic expressions in general. But it seems that here we encounter what can be called *the problem of scarcity*: the spatiotemporal world does not contain entities capable of serving as meanings for all our expressions. We need not even broach words such as *always* or *notwithstanding*, for already for such ordinary words as *dog* or *run* there do not seem to be single spatiotemporal objects which could be seen as their meanings. (Something like the mereological sum of all the actual dogs or runners would obviously not do.)

Another possibility is that meanings are objects of the mental realm. This appears to solve the problem of scarcity, for the mental realm can be seen as containing a bottomless fount of entities. However, here we face another kind of problem which can be called *the problem of intersubjectivity*: meanings are essentially intersubjective, they fulfill their role only as much as they can be shared among participants of communication, hence they cannot reside in private minds. (As Davidson, 1990, p. 314, puts it, "that meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language".)

Then there is a further possibility: Meanings are objects of a realm of ideal entities. This tries to avoid both the problem of scarcity and that of intersubjectivity, for the objects of this realm are considered objective like spatiotemporal things, and yet are not limited by space and time. Frege (1918, p. 69) was adamant about this being the only viable possibility for meanings: "A third realm must be recognized. What belongs to this corresponds with ideas, in that it cannot be perceived by the senses, but with things, in that it needs no bearer to the contents of whose consciousness to belong."

The problem with this, however, is the status of the realm of the ideal entities. Its whereabouts seem to be slightly enigmatic and in any case its denizens seem to be causally inert, which makes it hard to explain their impact on the spatiotemporal world. But here there is an opening for sets: they are ideal entities of the above kind, and yet they are backed up by a host of respectable mathematicians, hence to call them "enigmatic" would seem inappropriate. (The causal inertness is not yet addressed, but hopefully this too could be worked around ...) So perhaps sets may help us make semantics ultimately explicit. This would explain the open arms with which many theoreticians of language welcomed the set-theoretic semantics.

However, let us first return to the time before set theory came to the full function.

1.2 Frege on ideal objects

The general question is how we can explicitly grasp ideal objects. Frege, the ur-father of formal semantics, wrestled with this problem when he faced the question *What is a number*?,the answer to which was, in his eyes, key to the understanding the foundations of mathematics. He came to the idea that such objects must be grasped in terms of their names; however, he insisted, it is not enough to have (alleged) names of numbers, we also must have a way to determine when two names name the same number. Frege (1980, p. 73) writes: "If we are to use the symbol a to signify an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply this criterion. ... When we have thus acquired a means of arriving at a determinate number and of recognizing it again as the same, we can assign it a number word as its proper name."

To indicate that this approach does not concern only numbers, but abstract objects more generally, he first turns his attention, as a warm-up, to the abstract objet *direction*. The direction is something that can be, in geometry, represented by a straight line, we can therefore use the names such as *the direction of the line a*. Hence we have names for the directions, but in order to be able to use them as genuine names, we must supply a method which determines when two of the names name the same direction. Hence we need a method which lets us decide, for any claim of the form *the direction of the line a is the same as the direction of the line b*, whether it is true. Such a statement, however, can be clearly seen as a mere paraphrase of the statement *the lines a and b are parallel*, the truth value of which we can find (at least in principle). This lets us understand directions as objects (to recognize the same direction under different names).

But all this doesn't really tell us much about what a direction *is*. But this is normal, according to Frege: "The definition of an object does not, as such, really assert anything about the object, but only lays down the meaning of a symbol." (p. 78). To constitute an abstract object, then, is to know nothing more than to establish the meaning of a certain kind of sign: we cannot establish meaning except by determining how that sign is used in certain sentences (especially +certain equations), and more specifically what truth-values those sentences have. (Frege states bluntly, "It is only in the context of a proposition that words have any meaning"; p. 73.) All we can say about direction is that it is something that is common to two parallel lines. Thus the object *direction a* is identical with the object *direction b* precisely if *a* and *b* are parallel; both *direction a* and *direction b* in this case simply denote 'what *a* and *b* have in common (in this respect)'.

But if this is the case, then, according to Frege, we can identify the direction of a line without scruple with any suitable object that is associated with the line and that satisfies the condition that it is the same for two lines precisely when they are parallel. Such a convenient object, according to Frege, is, in the case of a line *a*, the domain of the notion of *line parallel to a* - that is, we would say today, the set of all objects which fall under this notion, i.e., the set of parallels of *a*. For the domain of the notion *parallel to a* (the set of parallels of *a*) and the domain of the notion *parallel to b* (the set of parallels of *b*) are obviously really identical precisely when *a* and *b* are parallel. Frege thus proposes to identify the subject direction a with the set of all parallels of *a*.

Does this mean that the direction is the set of parallels? (Isn't this counter-intuitive? - It doesn't seem that if we were thinking of the direction of a line, we would be thinking of a set!) There is no simple answer to this question. Frege's reasoning here is essentially this: since we can say nothing more about direction than that it is what is common to all parallel lines, we can pretty well identify it with anything that has this property - and disregard that the direction so grasped may well take on other, non-intuitive properties. This suggestion foreshadows the method that Carnap (1947) later called *explication*: replacing some abstract and hard-to-grasp entity with something that shares all its characteristic properties and is in some sense easier to grasp, usually some mathematical construct like a set (more about it in Section 7.3).

By analogy, Frege now wants to approach the notion of number (or what he calls *Anzahl*). He states that just as direction is something that belongs to a line, number is something that belongs to a concept (so, for example, to the concept of *planets of the solar system* there belongs the number eight). If, as we have already done, we call the set of all the objects falling under a concept *the domain* of that concept, we can say that the object *number of F* is identical with the object *number of G* precisely when the domain of the concept *equinumerous*[1] *to F* is identical with the domain of the concept *equinumerous to G*. In this case, then, we can - by analogy with what we have done in the case of direction - identify the number of

[1] Frege's German term is *gleichzählig*, which is often translated as *equal* or *equivalent*. This is, I think, a little bit misleading. Two equinumerous concepts are characterized by the fact that there is a one-one mapping between their domains and at least in the case of very small domains, we can just *see* this as we can *see* that two lines are parallel

F with the domain of the concept *equinumerous to F*, that is, with the set of all sets *equinumerous* to F.

Let us summarize the principles of Frege's approach to abstract entities. According to him, we can speak of entities of a certain kind, as we have said, if we have sigs for them, and if we can say when two of these signs designate the same entity, i.e. if we have a certain equivalence between the signs. We can then look at the relevant entity as that which all such equivalent signs, as signs, have in common. In both of Frege's examples, moreover, the sign of the abstract entity is uniquely tied to some more "concrete" entity - the direction sign to a line and the number sign to a concept. Thus, direction can be thought of as what all parallel lines have in common, and number as what all equinumerous concepts have in common. What Frege proposes next can be understood as identifying what the elements of a set have in common with the set itself: identifying what all parallel lines have in common with the set of all parallel lines, and what all equivalent concepts have in common with the set of all equivalent concepts.

In this way, then, Frege reduces the notion of number, which is at the basis of arithmetic, to the notions of a concept and of a domain of a concept (a set), which he regarded as purely logical notions. And when the distinction between a concept and its domain was later blurred, it was possible simply to say that numbers are sets of sets.

1.3 Sets

What, after all, is a set? It is notoriously difficult to say - though sets play a crucial role in the foundations of modern mathematics. The modern theory of sets originated with Georg Cantor[2], who originally considered sets of numbers, i.e. of points on the number axis. He studied certain functions and find out that if they have certain properties at every point of their domains, something

[2] See Cantor (1932).

important follows; but then he realized that it keeps to follow if the functions have the properties at every point of their domain *with the exception of some of them*. (For example: it was proven, already before Cantor, that if a function is representable by a uniformly convergent trigonometric series, then the series is unique. Cantor proved not only that uniform convergence is not necessary, but that even simple convergence might fail at some exceptional points, providing the set of such points is finite or even infinite if it has a certain structure.) It turned out that he needed a theory to characterize the sets in question - hence the original version of set theory.

But the theory acquired a life of its own and started to overspill its original confines. Thus, for example, Russell took sets to derive from properties: a property, according to him, determined the set of all objects having the property. (Thus, we cannot have a set all the elements of which do not share a property[3].) It followed that there was no reason to consider numbers as the only potential elements of sets. The property of (*being*) *hot* determines the set of all hot objects; the property of *being born in Istanbul on* 1.1.1111 determines the set of all individuals born in Istanbul on 1.1.1111, the property of *being a set of sets* determines the set of all sets whose elements are sets.

Frege's considerations presented in the previous section led Frege and Russell conclude that numbers are nothing else than certain sets. Number five, for example, is the set of all sets with five elements (oversimplifying a little bit). This led them to the conclusion that mathematics can be embedded into logic - the view known as *logicism* (but this is a story for a different occasion)[4]. And if numbers could be so easily embedded into such theory, why not other abstract objects? Functions, the abstract

---

[3] An anecdote ascribed to Bertrand Russell says that while we can have the set of all left shoes (because these can be distinguished from right ones), we cannot have that of left socks.

[4] See Demopoulos (2013).

objects crucial for mathematics, came to be grasped as sets of ordered pairs. Thus, the function of square (of natural numbers) would be the (infinite) set $\{<1,1>, <2,4>, <3,9>, ...\}$, where $<a,b>$ is a shorthand for $\{\{a\},\{a,b\}\}$.

Hence as long as mathematics was taken to be "a science of number", the sets it was interested in were points on the number axis, but the more mathematics moved from these confines to becoming the general theory of structures, the more it could be seen as a theory of sets, because all structures came to be reconstructed as set-theoretical objects. This was because set theory slowly became a general framework for reconstructing abstract and ideal objects, so that reconstructibility within such theory became a hallmark of being a genuine object.

The problem with sets in mathematics was that even after sets started to acquire such an important role in its foundation, nobody was able to say very clearly what a set is. The solution of this problem came with the axiomatizations of set theory. Several such axiomatizations were proposed and the general feeling was that a successful axiomatization gives us the answer to the question about the nature of sets. True, there were the disputes between various versions of set theory and several technical problems concerning these theories, but it was felt that generally these theories enlightened the nature of sets to a satisfactory extent[5].

The most basic axiom of every set theory states that sets are uniquely determined by their elements - that sets with the same elements are identical. Then there are axioms stipulating the existence of sets: of an empty set, of the set of all subsets of a given set, a set of all elements of all subsets of given set, etc. Then there are some more complicated axioms the need for which you will not understand unless you seriously submerge into set theory. And they are not important for us here.

What is important for us is that plus/minus all entities that are

[5] See Lavine (1994); Grattan-Guinness (2000); Potter (2004).

addressed by modern mathematics can be grasped as various sets. (We will discuss, in Section 5.5, the claim of the logician Pavel Tichý that he needs a kind of entities that cannot be accommodated within set theory, but this claim is controversial.) Crucial for this is the appropriation of the concept of function, which is central for many areas of mathematics.

1.4 Frege's maneuver

Frege, however, did not believe that all meanings are objects. Objects are meanings of *names* (in the broadest sense of the word - even sentences, by Frege's lights, are sort of names, meaning peculiar objects - truth values), but predicates express *concepts*, which, according to Frege, are certain functions; and functions, Frege maintains, are *not* objects. The thing, however is that subsequently functions come to be ever more identified with what Frege called their courses-of-values and which *are* certain sets. Hence even in this way Frege paved the way to grasping meanings as set-theoretical objects. And it is extremely important to understand how functions got into the picture.

Concepts, according to Frege, are functions which map objects on truth values. Think about how we use a predicate phrase: we attach to it a subject phrase to produce a sentence. Thus we may take the phrase *conquered Gaul* and attach it to *Caesar* to get the (true) sentence *Caesar conquered Gaul* (Frege's famous example). Similarly, I can do it with other subject phrases:

Caesar + *conquered Gaul* = *Caesar conquered Gaul*.

Aristotle + *conquered Gaul* = *Aristotle conquered Gaul*.

Cartman + *conquered Gaul* = *Cartman conquered Gaul*.

…

We can summarize the behavior of the predicate as the following function:

Caesar → *Caesar conquered Gaul*.

Aristotle → *Aristotle conquered Gaul*.

Cartman → *Cartman conquered Gaul*.

…

This all concerns expressions; but we can shift this whole consideration from their level to the level of meanings (using $\|\ldots\|$ to designate a meaning):

$\|\textit{Caesar}\| \rightarrow \|\textit{Caesar conquered Gaul}\|$

$\|\textit{Aristotle}\| \rightarrow \|\textit{Aristotle conquered Gaul}\|$

$\|\textit{Cartman}\| \rightarrow \|\textit{Cartman conquered Gaul}\|$

…

Now in view of the fact that the meaning of a name, according to Frege, is the individual named by it, and the meaning of a sentence is its truth value, this boils down to (let me use pictures of individuals instead of names to stress that these are *not linguistic* objects)

 → ***Tr***

 → ***Fa***

 → ***Fa***

...

I propose to call this encapsulation of the functioning of an expression into a function *Frege's maneuver*. It has been copiously repeated by the formal semanticists in the building of their semantic models of natural language. Consider, for example, the adverb *quickly*: it maps predicative phrases on predicative phrases:

conquered Gaul → *conquered Gaul quickly*

killed Cesar → *killed Cesar quickly*

robbed a bank → *robbed a bank quickly*

...

By Frege's maneuver we can transform its semantic behavior into a function that maps functions from objects to truth values on the same kinds of functions. And in a similar way we can accommodate expressions of plenty of other categories within the semantics.

Something close to Frege's maneuver also played a crucial role in the transfer from the extensional to the intensional model of meaning. The idea behind this was that to understand an expression, I need to know not only its actual extension, but also its potential extension in contrafactual situations. Thus, to know the extension of a sentence is to know its truth value, to know its intension is to know in which circumstances it is true (*viz*. to know its truth conditions).

Now the maneuver close to Frege's here encapsulates this again into a function – a function whose arguments are no longer meanings of linguistic expressions, but rather possible worlds. (What exactly the possible worlds are supposed to be has been a notorious source of dispute, which makes the foundations of this very approach somewhat shaky.) We discuss this in detail in Chapter 4.

1.5 Semantic models of language

If meanings are ideal objects, hence if they neither exist within space and time, nor do they exist merely in the subjective mental world of a speaker, they are difficult to get hold of and to explain their workings. Frege, in effect, proposed to embed these objects into the realm of mathematics which was later established as a realm supervised by set theory. This has turned out to be a fruitful way; but has this movement resolved the general questions concerning the nature of meanings in the sense that meanings *are* sets? Well, we leave the answer to this general question to the last chapter of this book; and so far, we only conclude that it has turned out to be useful to *reconstruct* meanings as sets.

The notion of reconstruction lets us maintain a certain distance between the possibly "ineffable" meaning as such and its set-theoretic reconstruction. Therefore, I think it is best to see the situation where we reconstruct meanings as set-theoretical objects as building logico-mathematical *models* of language, especially of its semantics.

Let me, before I characterize the kind of models we are going to build, point out some of their general structural features. We want there to be one and only one meaning for every meaningful expression (this is obviously an oversimplification, we disregard ambiguity); moreover, we want that the meaning of a complex expression be produced out of the meanings of their parts. Hence our models will incorporate what is called the *principle of compositionality*[6]. This principle states that the meaning of a complex expression is always uniquely determined by the meanings of its parts plus the mode of their combination. In particular, it states that if we denote the meaning of E as $\|E\|$, then for every syntactic rule R there must exist an operation R^* such that

$$\|R(E_1,...,E_n)\| = R^*(\|E_1\|,...,\|E_n\|)$$

for every expressions $E_1, ..., E_n$ that can be combined by the rule R into a complex expression. The principle of compositionality is thus a constitutive feature of our semantic models. The principle of compositionality is equivalent to what can be called the principle of intersubstitutivity of synonyms:

[6] See Werning, Machery, and Schurz (2005). See also Peregrin (2001a, Chapter 4) for a discussion of the motivations and consequences of the principle.

if $\|E_i\|=\|E_i'\|$

then $\|\mathrm{R}(E_1,...,E_i,...,E_n)\|=\|\mathrm{R}(E_1,...,E_i',...,E_n)\|$

Indeed, if the principle of compositionality holds and $\|E_i\|=\|E_i'\|$, then $\|\mathrm{R}(E_1,...,E_i,...,E_n)\| = \mathrm{R}^*(\|E_1\|,...,\|E_i\|,...,\|E_n\|) = \mathrm{R}^*(\|E_1\|,...,\|E_i'\|,...,\|E_n\|) = \|\mathrm{R}(E_1,...,E_i',...,E_n)\|$. Conversely, the R^* claimed by the principle of compositionality does not exist iff there are $E_1,...,E_n,E_1',...,E_n'$ so that $\|E_1\|=\|E_1'\|,...,\|E_n\|=\|E_n'\|$ and $\|\mathrm{R}(E_1,...,E_n)\| \neq \|\mathrm{R}(E_1',...,E_n')\|$; but it is easy to see that this is excluded by the principle of intersubstitutivity of synonyms.

As Janssen (1986) pointed out, given the principle of compositionality, we can see a semantically interpreted language as a many-sorted algebra of expressions mapped - by a homomorphic mapping - onto the many sorted algebra of meanings.

Speaking about such general structural semantic principles, we can mention one more, interconnecting meaning and truth. It states that if two sentences differ in truth value, they cannot but differ in meaning. If we denote the truth value of the sentence S as $|S|$, we have[7]

if $|S_1|\neq|S_2|$, then $\|S_1\|\neq\|S_2\|$,

for every sentences S_1 and S_2.

This principle together with the principle of the intersubstitutivity of synonyms yields what can be called the principle of the intersubstitutivity of synonyms *salva verirate*:

if $\|E_i\|=\|E_i'\|$,

then $|\mathrm{R}(E_1,...,E_i,...,E_n)| = |\mathrm{R}(E_1,...,E_i',...,E_n)|$,

---

[7] Elswhere I called it *the principle of verifoundation* (see Peregrin 1994; 2001b). Cresswell (1982) considers it to be the most certain principle of semantics.

whenever $R(E_1,...,E_i,...,E_n)$ and $R(E_1,...,E_i',...,E_n)$ are sentences. The inversion of this principle is sometimes called the Leibniz principle:

if $|R(E_1,...,E_i,...,E_n)| = |R(E_1,...,E_i',...,E_n)|$ for every sentences $R(E_1,...,E_i,...,E_n)$ and $R(E_1,...,E_i',...,E_n)$, then $\|E_i\| = \|E_i'\|$.

1.6 The four kinds of models

Let me explain, in greater detail, what kinds of models of language formal semantics introduced and we are going to reonstruct. Each such model will consist of two parts, of *syntax* and *semantics*. The syntactical part will reconstruct the set of well-formed expression with which it will work. It will consist of two parts: a vocabulary and a set of syntactic rules. The vocabulary will be a list of words, of elementary expressions divided into syntactic categories (e.g. the words <u>*Eco*</u> and <u>*Schwarzenegger*</u> of the category *term* and the words <u>*writer*</u> and <u>*actor*</u> of the category *predicate*). The rules will prescribe how to build more complex expressions out of simpler ones (e.g. that it is possible to combine a term with a predicate into a *statement*: <u>*writer*</u>(<u>*Eco*</u>)). In sum, the syntactical part will determine what counts as a well-formed expression.

The semantical part of the definition of the model will equip every well-formed expression with a set-theoretical object, reconstructing the meaning of the expression (e.g. the words <u>*Eco*</u> and <u>*Schwarzenegger*</u> with the persons Eco resp. Schwarzenegger and the words <u>*writer*</u> and <u>*actor*</u> with the sets of all writers resp. actors). It will consist of two parts copying the two parts of the syntactic component. The first part will assign a set-theoretical object to every word of the vocabulary. The second part, then, will tell us, for every syntactic rule, how to construct an object assigned to a complex expression built according to this rule from the objects assigned to its parts (e.g. that a statement consisting of a term and a predicate is assigned the truth value ***Tr*** if the object assigned to its term is an element of the set assigned to its predicate; and the value ***Fa*** otherwise).

What we call models are thus a kind of artificial languages such as those we know from formal logic. And indeed I think that the artificial languages of logic can be seen as this kind of models of natural language (Peregrin, 2020). The difference is that logical languages concentrate on a logical part of their vocabulary, usually leaving the extralogical part completely aside; whereas here we are interested in all parts of the vocabulary, without a difference.

In this book, I discuss four kinds of models delivered by formal semantic. In Chapter 3, I discuss what I call the *extensional model of meaning*. This is the model which was mostly developed within formal logic thanks to people like Frege, Tarski and many others. It is the model that is sufficient if we want to discuss mathematics or just leave aside the empirical dimension of language. Many logicians did concentrate on mathematics, and so they were content with this kind of semantics; however, for philosophers and linguists who took natural language seriously this kind of semantics was simply a non-starter.

The model that I call *intensional*, then, was a breakthrough. It was foreshadowed in the writings of Carnap and it came to full fruition in the hands of Montague and his followers. It showed how to model the meanings of even the empirical expressions, using the concept of possible world. This concept was hinted at by Carnap, who felt that extension is not a fair model of meaning in the intuitive sense of the word and that we need to get a grip on the notion of intension; independently of this, it was also arrived at by Kripke, who looked for the semantics of modal logic. I discuss this semantic model in Chapter 4.

As intentional logical model turned out to be not wholly waterproof (It was challenged especially with the propositional attitude reports), there appeared amendments, which led, quite quickly, to models which I classify as *hyperintensional*. This is a bundle of different kinds of models, which share the assumption that meaning has a kind of structure that is related to the syntactic structure of their expressions. These models are discussed in Chapter 5.

Finally, in Chapter 6, I discuss the models that I call *dynamic*. They appeared most recently and they took into consideration the fact that language is a means of discourse which is a dynamic enterprise. Their proponents maintained that if we want to analyze natural language with all its peculiarities like pronouns or the phenomenon of anaphora, we must build different semantic models than the intensional or the hyperintensional ones. In particular, we need to incorporate the concept of context.

In the last chapter of the book I return to the general problem of capturing meaning as object. I conclude that the question whether this is reasonable must be kept apart from the question whether meaning really *is* an object (where there may be no clear answer to the latter question). I maintain that we should see the relation between the models of meaning and their target phenomena as that of *explication*. And that this explication is extremely fruitful because it equips semantics with a huge toolbox of set theory.

2 A very short history of formal semantics

2.1 Frege

The roots of formal semantics, we already saw, can be traced back to the writings of Gottlob Frege (1848-1925), the German mathematician, logician and philosopher who laid the foundation not only of modern formal logic but also of what has later become known as analytic philosophy (Dummett, 1996). He was the first to clearly realize that semantics has little to do with psychology, and that it could be usefully explicated in mathematical terms (Dummett, 1981a; 1981b).

Frege's depsychologization of semantics followed from his depsychologization of logic. Frege understood how crucial it was for the development of logic to draw a sharp boundary separating it from psychology: to make it clear that logic is *not* a matter of what is going on in some person's head, in the sense that psychology is. The reason is that logic is concerned with what is true and consequently what follows from what – and whether something is true, or whether something follows from something else, is an objective matter independent of what is going on in the head of a particular person.

As a consequence, Frege realized that if logic must be separated from psychology, then the same is true for semantics – at least insofar as semantics underlies truth and entailment. It is clear that the truth value of a sentence depends on the meaning of the sentence: the sentence "London is in England" is true not only due to the fact that London is in England, but of course also due to the fact that the words of which it consists mean what they do in English. Hence, if meaning were a matter of what is going on in somebody's head, then truth would have to be too – hence meaning must not, in pain of the subjectivization of truth, be a psychological matter. But what, then, *is* meaning?

Frege started from the *prima facie* obvious fact that names stand for objects of the world. (Unprecedentedly, he assimilated sentences to name as well: he saw them as specific kinds of names that denote truth values. The reason for this move was that he divided expressions into two sharply separated groups: into "saturated" – i.e. self-standing – and "unsaturated" – i.e. incomplete – ones. He took names and sentences as species of the former kind, whereas he took predicates as paradigmatic examples of the latter one; and he came to use the word "name" as a synonym of "saturated expression".) His most brilliant contribution to the explication of the concept of meaning then was the way he accounted for the meanings of predicates in terms of what we have dubbed Frege's maneuver. He called them *concepts*, as usual; but he rejected the usual way of seeing concepts as something mental and, in effect, he suggested explicating them by means of studying how the expressions which express them – i.e. predicates – function within language.

What is the function of a predicate, such as "to think"? Well, the predicate is attached to a subject to form a sentence. Hence if we assume that the meaning of a complex expression is the result of combining the meanings of its parts (i.e. that meanings are composed in a way paralleling that in which the expressions expressing them are), then the meaning of the predicate together with the meaning of a subject, which is the object stood for by the subject, yields the meaning of a sentence, i.e. a truth value. Hence a concept is something that together with an object yields a truth value – and this led Frege to identify concepts with functions, in the mathematical sense of the word, taking objects to truth values. In effect, this meant the identification of the meaning of an item with the semantic function of the item captured as a function in the mathematical sense of the word; and this opened the door for a mathematical treatment of semantics. Thus, we can say that Frege married semantics, which he had earlier divorced from psychology, to mathematics.

Notice that Frege's maneuver has two substantial presuppositions. There is the presupposition that the meaning of a complex expression is yielded by (or 'composed of') meanings of the parts of the expression. This is the *principle of compositionality* we discussed in §1.5. How do we know that this principle holds? Some theoreticians appear to think that it is an empirical thesis that must be verified as empirical theses are: by means of inspecting as many cases as possible. However, such a view presupposes that meanings are independently identifiable objects whose combinations can be studied in the way we study, e.g., combinations of molecules in a solution: that we can empirically verify (or falsify) the thesis that, say, the meaning of a sentence is yielded by the meaning of its subject and that of its predicate, by means of finding the meanings and finding out what they yield if they are put together. In contrast to this, we saw that for Frege the principle was rather a way of articulating what it takes to be meaning: the principle was co-constitutive of the notion of meaning in the sense analogous to that in which, say, the principle of extensionality is co-constitutive of the concept of set. And just as it makes no sense to try to find out whether sets are extensional (for this is simply part of what it takes to be a set), it makes no sense to try to find whether meaning is compositional.

The other presupposition of Frege's maneuver is of a different kind: it concerns the behavior of the particular expressions to which it is applied. The presupposition is that the role of the expression within language is exhausted by, or at least in some sense reducible to, its role within the kind of syntactic combination which is taken into consideration. We explicated the meanings of predicates by considering the way they combine with names into sentences; but predicates also do other things, e.g. combine with adverbials into complex predicates. We must always be sure that this is taken care of – that it is proven that it is somehow substantiated to treat some part of the functioning of an expression as representative of the whole functioning.

Is Frege's way of explicating the concept of meaning acceptable? In fact not: what Frege called *meaning* cannot be taken as a plausible explication of the pre-theoretic notion of meaning. After all, who would want to claim that all true sentences have the same *meaning*? And Frege soon came to realize the implausibility of such an explication. Therefore he complemented his theory of meaning by what he called a theory of *sense*[8]. Every name, he claimed, has not only a meaning, but rather also a sense, which is the 'way of givenness' of the meaning. And it is then Frege's concept of *sense*, rather than his concept of *meaning*, which is to be taken as his explication of the intuitive concept of meaning.

Frege's own instructive example is that of the terms "morning star" and "evening star". As we now know, these two terms refer to one and the same celestial body, the planet Venus. Hence they share the same meaning (in Frege's sense of the word), or (in the current jargon) the same *referent*. However, although the sentence "The morning star is the morning star" is obviously trivial, "The morning star is the evening star" does not appear to be such. The reason, Frege claimed, is that the terms differ in their senses, i.e. in the ways they present their referent: "the morning star" presents it as the most attractive body in the morning sky, whereas "the evening star" as the most attractive one in the evening sky.

Hence we have the general picture according to which the relation between a name and what the name refers to is mediated by the sense of the name:

NAME

$\downarrow$

SENSE (i.e. meaning in the intuitive sense of the word)

$\downarrow$

MEANING (i.e. object referred to)

[8] See Frege (1892b).

Another of Frege's path-breaking contributions to the development of logic and mathematics was his establishment of a logical langua

ge which, despite the fact that it looked very different, was structurally almost identical to what we now call predicate calculus. His basic achievement was the introduction of what we know today as quantifiers (although Frege's own notation was idiosyncratic)[9].

When introducing quantifiers, Frege said roughly this: Imagine a sentence decomposed into two parts, and imagine one of the parts 'abstracted away', the sentence being thereby turned into an "unsaturated", gappy torso. Then imagine the gap in this matrix being filled with various things[10] and consider the truth values resulting from the individual saturations. In some cases it can happen that however we fill the gap, we will always reach a true sentence; which Frege abbreviated by means of the general quantifier. Thus, the sentence that in modern notation reads

$$\forall x \mathrm{F}x$$

was, by definition, his shorthand for "whatever replaces the *x* in F*x*, we gain a true sentence". Similarly, Frege's equivalent of the modern

$$\exists x \mathrm{F}x$$

was introduced to shorten the claim that there is at least one thing that can replace the *x* in F*x* so that we gain a true sentence.

---

[9] See Frege (1879).

[10] According to modern standards, this is ambiguous: it can mean either (i) imagine that the gap is literally filled with an *expression*, or (ii) imagine that the formal sign marking the gap, the variable, is made to refer to an *object*. These two interpretations (ultimately resulting into what is nowadays called the *substitutional* and the *objectual* notion of quantification, respectively) come out as equivalent only if we assume that every object (within the relevant universe of discourse) has a name.

In this way, Frege introduced the machinery of quantifiers and variables in the shape that has underlain almost all formal logic since.

2.2 Tarski

Frege's distinction between *meaning* and *sense* has influenced virtually all subsequent theories of meaning; however, his "mathematical" explication of concepts and meanings has not been absorbed as quickly as it might have deserved. Thus it is somewhat peculiar that when Carnap (who was familiar with Frege's teaching) in 1934 wrote his important book about the logical formalization of language, *Der Logische Syntax der Sprache*, he claimed that the only aspect of language that is susceptible to formalization is syntax; that semantics is ineffable. It was only slightly later, after he had absorbed the teaching of Alfred Tarski (1901-1983), that he admitted that there was a way of formalizing semantics that was as rigorous as the formalization of syntax.

A great deal of Tarski's theory of semantics looks like re-discovering Frege's ideas and putting them into the context of a more developed theory of formal logic. His formalization of semantics was a kind of by-product of his theory of truth. What he was after was the fixation of the meaning of "true" by putting together some *axioms* governing it; just like some of his colleagues had fixed the meaning of "set" by means of axiomatic set theories before. Tarski[11] realized that what would fix the meaning of "true" were all sentences of the shape

(1) **true**(....) ↔ __,

with the dots replaced by a name of a sentence and the underscore by the very sentence. Sentences of this form are now generally called T-sentences (you may opt for interpreting the "T-" as standing either for "Tarski", or for "truth"). However, the set of all these sentences was infinite, and hence could not be taken as

[11] See Tarski (1935; 1944).

the desired theory. So the problem of the explication of the concept of truth appeared to boil down to the problem of finding a 'reasonable' (preferably finite) set of axioms that would entail the infinite set of all the T-sentences[12].

If we assume that the language which we are considering is the language of standard logic, we can divide its sentences into three classes: *atomic* (sentences consisting of a predicate applied to the appropriate number of terms), *logically complex* (those consisting of a logical operator applied to one or two sub-sentences) and *quantified* (those consisting of a quantifier binding a variable in a formula). Now it is clear that using a few simple principles we can deduce the T-sentences for logically complex sentences from those for the other sentences. The point is that using the principles

[12] Could we make do with a single axiom produced by 'closing' the schema (1) by means of variables and quantifiers? It is easy to see that $\forall x \forall y(\mathbf{true}(x) \leftrightarrow y)$ would not do; but what about $\forall y(\mathbf{true}('y') \leftrightarrow y)$? A moment's reflection reveals that it would work only if '*y*' referred to a name of the sentence *y* – but in fact the name which arises from enclosing a symbol in quotes notoriously refers to the very symbol. (Hence '*y*' does not refer to the name of the sentence *y*, but rather to the penultimate letter of the alphabet.) Similarly $\forall y(\mathbf{true}(\mathbf{name}(y)) \leftrightarrow y)$: even if we disregard that it would require *y* to be a sentential variable and **name** to form terms out of sentences (and hence would not be accommodable within the framework of the predicate calculus, which Tarski took as the ultimate framework of logic), there does not appear to be a *function* taking denotations of sentences to their names (which is to be denoted by **name**) – the relationship between the former and the latter appears to be one-to-many. And even if we assumed that for every denotation of sentence there is one 'canonical' name to be yielded by such a function, to make the functor denote the right function (taking a denotation of a sentence to its canonical name) would amount to inventing an axiomatic theory doing precisely what Tarski urged: entailing all the T-sentences. This becomes obvious when trying to engage the converse functor **denotation** forming sentences out of terms and denoting a function taking sentences to what they denote. In this case the T-scheme would yield us $\forall y(\mathbf{true}(y) \leftrightarrow \mathbf{denotation}(y))$; and it is clear that then **denotation** becomes simply *equivalent* to **true**, and providing a theory for it *is* providing a theory of **true**. See also Kirkham (1995, Chapter 5).

(where $Neg(x)$ refers to the negation of the sentence x and $Con(x,y)$ refers to the conjunction of x and y[13])

(2) $\mathbf{true}(Neg(x)) \leftrightarrow$ not $\mathbf{true}(x)$,

and

(3) $\mathbf{true}(Con(x,y)) \leftrightarrow (\mathbf{true}(x)$ and $\mathbf{true}(y))$,

we can deduce the T-sentence of any conjunction and any negation from those of its immediate subsentence(s) (the same clearly holds for the other standard logical operators); and applying this recursively, we can eventually deduce it from other than logically complex sentences.

Now what about quantified sentences, of the kind of $\forall xF$? In general they do not contain any subsentences, but only (possibly open) subformulas, so the previous method is not applicable. Here is where Tarski came with an ingenious idea[14]: he put the concept of truth to the side for a while and instead turned to the notion of *satisfaction*, a relation between formulas and objects of the world. Intuitively, satisfaction is the relation which holds between the formula $\mathbf{brothers}(x,y)$ and a pair of persons iff the two persons are brothers; but as a formula can contain an unlimited number of different variables, things are simplified if we define it as a relation between formulas and infinite sequences of objects. (Variables are thought of as linearly – e.g. alphabetically – ordered and hence corresponding to the objects of the sequences in a one-one way. Hence if we assume that x is the first and y the second variable, in the case of $\mathbf{brothers}(x,y)$ only the first two elements of such a sequence matter.)

---

[13] Note that talking about the sentences of a language requires that the sentences are contained within our universe of discourse. As the assumption that we are capable of treating of a language within the very language might be dangerous (by being liable, as Tarski showed, to inducing inconsistencies), we assume that we are operating within is what Tarski called *metalanguage*, i.e. a language capable of treating of an *object language* by containing names of its expressions.

[14] See Peregrin (1999).

It is easy to see that satisfaction of quantified formulas is reducible to that of their unquantified subformulas. To show this, we must first introduce some notation. Let us assume that the variables of the language we deal with are ordered, and let Var(i) refer to the ith one. Moreover, let All(x,y) denote the formula constituted by the concatenation of the general quantifier, a variable x and a formula y; and let Ex(x,y) denote the formula constituted by the concatenation of the existential quantifier, the variable x and the formula y[15]. Finally, if S is an infinite sequence of objects, then let S[i,a] refer to the sequence which is identical with S save for the only possible difference that its ith constituent is a. Then it clearly holds that

(4) **sat**(All(Var(i),y),S) $\leftrightarrow$ **sat**(y,S[i,a]) for every object a of the universe

(5) **sat**(Ex(Var(i),y),S) $\leftrightarrow$ **sat**(y,S[i,a]) for at least one object a of the universe

and satisfaction for quantified formulas is indeed reducible to that for their subformulas. (As it is more perspicuous to work directly with functions assigning objects to variables instead of the object-sequences which effect such assignments indirectly, we will do

---

[15] Note that the sign "x", as employed within the previous paragraph, is not a variable (of the object language), but a *name* of a variable (a symbol of our metalanguage). This may easily lead to a certain chaos, for the following reason. When we speak about a non-linguistic object, we have no choice but to use a sign standing for it – we cannot put, e.g., an apple itself into a sentence speaking about it. Not so, however, in case of *linguistic* objects, signs. We can, as is often done, put the object itself, instead of its name, into a sentence. Thus suppose that "α" is a variable of the language we are investigating. Then we can say *the formula... contains the variable "α"*, or, if "α" is the first variable according to the relevant ordering and we use the notation introduced in the previous paragraph, we can equivalently say *the formula ... contains the variable Var*(1), but it is often also written *the formula ... contains the variable α*. The last formulation is literally incorrect, but it is often used instead of the first one.

so; and we will call a function assigning an object to every variable a *valuation of variables*.)

Moreover, satisfaction for logically complex formulas is reducible to that for logically simple ones along the lines wholly analogous to those along which truth is, *viz*.

(6) **sat**(*Neg*(*x*),*S*) ↔ not **sat**(*x*,*S*)

(7) **sat**(*Con*(*x,y*),*S*) ↔ (**sat**(*x*,*S*) and **sat**(*x*,*S*)).

This implies that for a language with a finite number of atomic formulas we can have a *finite* theory of satisfaction (the theory would be constituted by the sentences stating the satisfaction conditions for all the atomic formulas plus by (4)-(7)). Now the point of this maneuver is that for sentences (formulas with no free variables), truth is clearly reducible to satisfaction:

(8) **true**(*x*) ↔ (**sat**(*x*,*S*) for every sequence *S*).

Hence the finite theory of satisfaction yields us the desired finite theory of *truth*.

Tarski's investigation thus seemed to suggest that the concept of truth has to be attacked by means of the investigation of a language-world relation such as satisfaction – therefore this theory has come to be called the *semantic* theory of truth[16]. Moreover, for many logicians (notably for Carnap) it acted as a revelation of the fact that semantics was not as inaccessible to a formal treatment as it had appeared up to the point.

2.3 Carnap

Tarski's relation of satisfaction gestures towards a formalization of the relation expression-meaning (or expression-referent); but it is not really a formalization of it. In fact, from the viewpoint of natural language it is slightly unnatural – for it presupposes the

[16] See Kirkham (1995, Chapter 5); Peregrin (1999).

existence of open formulas, which have no counterpart in natural language[17].

Let us consider a formal language with no variables and quantifiers, but with an infinite number of atomic sentences. Let us assume that the category of terms of the language is productive, i.e. that we have functors capable of taking terms to terms (such as '+' or '×' of arithmetic, which join pairs of terms into complex terms). How might a truth definition for such a language look? Instead of Tarski's **sat** we would need the relation **des** relating an expression to an object, to what it 'designates'. Let us provide for the reducibility of **true** to **des** (which gives the introduction of **des** its point) by assuming that the objects that are designated by sentences, 'propositions', are capable of 'being true'. As the analogue of T-sentences we now have what can be called D-sentences, namely sentences of the form

des(...., __),

where the dots are replaced by a name of an expression and the underscore by the expression itself. Thus the D-sentences include

des("John",John)

des("to be bald", to be bald)

des("John is bald", John is bald)

the second arguments of which are supposed to stand for the individual John, the property of being bald and the proposition that John is bald (whatever properties and propositions might be supposed to be), respectively. In this way we reach the view of semantics developed in the forties by Rudolf Carnap (1891-1970).

Now if the concept of designation is to yield us a theory of truth as the concept of satisfaction did, we must provide for two things: (1) a finite theory of designation (a finite number of axioms

[17] Cf. Peregrin (2000b).

entailing all the D-sentences); and (2) the reduction of the concept of truth to the concept of designation. Let us start with the latter.

Carnap (1942, p. 51) claims that the reduction works as follows[18]:

$\mathbf{true}(x) \leftrightarrow \exists prop\ (\mathbf{des}(x,prop) \wedge prop)$

i.e. a sentence *x* is true iff it expresses a proposition *prop* and *prop* is the case. Hence, a sentence is true if, e.g., it expresses the proposition that it is raining and it is the case that it is raining. This amounts to assuming that there is a way from propositions to their truth values. The trouble is that it is not clear what propositions (and properties) are really supposed to be. Alternatively, we could stick to the Fregean approach and to assume that what sentences designate are directly the truth values, which would yield us

$\mathbf{true}(x) \leftrightarrow \mathbf{des}(x,\boldsymbol{Tr})$.

Then the reduction of all the D-sentences to the D-sentences for logically non-complex sentences is effected by the principles of the following kind

$\mathbf{des}(Neg(x),\boldsymbol{Tr}) \leftrightarrow \mathbf{des}(x,\boldsymbol{Fa})$

$\mathbf{des}(Con(x,y),\boldsymbol{Tr}) \leftrightarrow (\mathbf{des}(x,\boldsymbol{Tr})$ and $\mathbf{des}(x,\boldsymbol{Tr}))$.

Now, however, we need not stop here, for designation is defined not only for sentences, but also for their parts; and for the sentence $p(s)$ consisting of a subject *s* a predicate *p* we can stipulate

$\mathbf{des}(p(s),\boldsymbol{Tr}) \leftrightarrow \exists i \exists r(\mathbf{des}(s,i)$ and $\mathbf{des}(p,r)$ and the individual *i* has the property *r*)

If we accept the Fregean identification of properties with functions, we can turn this further into

$\mathbf{des}(p(s),\boldsymbol{Tr}) \leftrightarrow \exists i \exists r(\mathbf{des}(s,i)$ and $\mathbf{des}(p,r)$ and $r(i))$

[18] Note that if we define **designation**(x) as the only *prop* such that **des**(x,*prop*), this becomes tantamount to the last proposal discussed above in footnote 12.

Moreover, for complex names of the shape $f(t)$ we have

$$\mathbf{des}(f(t),i) \leftrightarrow \exists g \exists i'(\mathbf{des}(t,i') \text{ and } \mathbf{des}(f,g) \text{ and } i = g(i'))$$

Carnap (1947) also realized that if what we are after is meaning in the intuitive sense of the word, then we should not be interested so much in meanings in the sense of Frege, but rather in Fregean senses. However, as Frege did not explicate the concept of sense to Carnap's satisfaction, Carnap proposed replacing the Fregean twin concepts of meaning and sense with the concepts of *extension* and *intension*. Roughly speaking, the *extension* of a term is what the term shares with all terms that are equivalent to it; whereas its intension is what it shares with all the terms that are *logically* equivalent to it.

Of course this definition becomes non-trivial only after we give a rigorous account of the concept of equivalence on which it rests. For the basic categories of the predicate calculus this is not difficult: two individual expressions I_1 and I_2 are equivalent iff $I_1 = I_2$, two n-ary predicates P_1 and P_2 are equivalent iff $\forall x_1 \ldots \forall x_n(P_1(x_1,\ldots,x_n) \leftrightarrow P_2(x_1,\ldots,x_n))$; and two sentences S_1 and S_2 are equivalent iff $S_1 \leftrightarrow S_2$. This explication leads to the concept of extension almost indistinguishable from Frege's concept of meaning: the extension of an individual expression being the object for which it stands, that of a predicate the function assigning the truth value ***Tr*** to those n-tuples of objects of which the predicate is true (or, equivalently, the class of the n-tuples) and that of a sentence its truth value.

The concept of intension is far more problematic, but Carnap indicated a way to approach it, namely via the concept of (possible) *state-of-affairs*. This concept has later been replaced, especially thanks to the seminal results of Saul Kripke (1963b) concerning the model theory for modal propositional calculus, by the concept of *possible world*.

2.4 Standard Logic and its Semantics

Some logicians during the twentieth century, especially those inclining to mathematics, came to the conclusion that there is something as *the* language of logic – that it is the language of what has come to be called the first-order predicate calculus[19]. The vocabulary of a language within the framework of this calculus falls into three categories:

1. logical constants ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\exists$, $\forall$)
2. extralogical constants (individual, predicate, functor)
3. variables (individual)

The syntax of such a language is then as follows:

An individual constant and an individual variable is a *term*; moreover, if F is an n-ary functor and $T_1,...,T_n$ are terms, then $F(T_1,...,T_n)$ is a term (and nothing else is a term).

If P is an n-ary predicate and $T_1,...,T_n$ terms, then $P(T_1,...,T_n)$ is a *formula*; moreover, if F_1 and F_2 are formulas and x a variable, then $\neg F_1$, $F_1 \wedge F_2$, $F_1 \vee F_2$, $F_1 \rightarrow F_2$, $\forall x F_1$ and $\exists x F_1$ are also formulas (and nothing else is a formula).

The semantic treatment of a language of this kind that has become standard needs some elucidation, for it rests heavily on the logical/extralogical boundary, which has not as yet played a principal role in our exposition. We have seen that Tarski's semantic theory of truth was based on the concept of satisfaction, which in turn necessitated the employment of the valuations of variables. Tarski's explication of the concept of logical consequence was based on an analogous idea, only applied at a higher level. While a sentence S is a consequence (*simpliciter*) of $S_1,...,S_n$ iff S cannot be false unless at least one of $S_1,...,S_n$ is, it is their *logical* consequence iff this is the case independently of

[19] Probably not many logicians would subscribe to the existence of "one true logic" explicitly (as e.g. Priest (2001) would), but in majority of contexts where the term "logic" is used without a qualification, it means first-order predicate calculus.

what is stood for by the non-logical words in $S_1,...,S_n$ and S – in other words if this is the case for every assignment of values to these words (This, of course, presupposes that we are able to classify the vocabulary of our language into the logical and the non-logical part, which is far from non-controversial, but let us neglect this problem for now.) This means that the fact that *John is unmarried* is a logical consequence of *John is a bachelor* and *Every bachelor is unmarried* because *X is* (*a*) *Y* is a consequence of *X is* (*a*) *Z* and *Every Z is* (*a*) *Y* whatever may be stood for by *X*, *Y* and *Z*, schematically

$$\forall X \forall Y \forall Z\, (X\ is\ (a)\ Z,\ Every\ Z\ is\ (a)\ Y \Rightarrow X\ is\ (a)\ Y).$$

Now this is a 'meta-level' quantification (the 'quasiformula' just presented is not to be understood as belonging to the language in question, but rather as being 'about' it), which cannot be mixed with the 'object level' one (such as would result, e.g., from the standard regimentation of *Every bachelor is unmarried* as $\forall x(\mathbf{B}(x)\rightarrow\mathbf{U}(x))$). Therefore, we need *two* separate kinds of variable words and *two* separate sets of valuations: we keep calling the 'object level' variables simply *variables*, whereas we call the 'meta-level' ones *parameters*. Whereas the former underlie quantification of the object language, the latter underlies (explicit or implicit) quantification of the metalanguage, of statements *about* the object language.

Thus, the resulting semantics is based on two sets of assignments of objects to expressions: the *valuation of variables* and the *interpretation of parameters*. This variety of formal semantics has been developed especially within the framework of what has come to be called *model theory* and which has developed out of Tarski's later work. Hence from the model-theoretic perspective, a language of the kind discussed has three basic kinds of expressions: The semantics of logical constants is taken to be fixed; and they are usually not taken to designate objects (although it also is possible to take them so). Parameters or extralogical constants are taken to be assigned a denotation by the interpretation, which maps individual constants on elements of a universe, predicate constants on relations over the universe, and functor constants on functions over the universe. Variables are then taken to be interpreted by the valuation, which maps them on the elements of the universe. Given an interpretation I and a valuation V, every individual term T is assigned a denotation $\|T\|_{I,V}$ based on I and V in the following way:

If T is an individual constant, then $\|T\|_{I,V} = I(T)$

If T is an individual variable, then $\|T\|_{I,V} = V(T)$

If T is $U(T_1,...,T_n)$ for some functor U and terms $T_1,...,T_n$,

then $\|U(T_1,...,T_n)\|_{I,V} = I(U)(\|T_1\|_{I,V},...,\|T_n\|_{I,V})$

An interpretation I and a valuation V also render each formula true or false. The usual inductive definition goes as follows:

If F is $P(T_1,...,T_n)$ for some predicate P and terms $T_1,...,T_n$, then F is true w.r.t. (or satisfied by) I and V iff $<\|T_1\|_{I,V},...,\|T_n\|_{I,V}> \in I(P)$

If F is $\neg F'$, then F is true w.r.t. (or satisfied by) I and V iff F' is not

If F is $F_1 \wedge F_2$, then F is true w.r.t. (or satisfied by) I and V iff both F_1 and F_2 are

If F is $F_1 \vee F_2$, then F is true w.r.t. (or satisfied by) I and V iff either F_1, or F_2 is

If F is $F_1 \rightarrow F_2$, then F is true w.r.t. (or satisfied by) I and V iff either F_2 is or F_1 is not

If F is $\forall xF'$, then F is true w.r.t. (or satisfied by) I and V iff F' is true w.r.t. (or satisfied by) I and V' for every valuation V' which differs from V at most in the value it assigns to x

If F is $\exists xF'$, then F is true w.r.t. (or satisfied by) I and V iff F' is true w.r.t. (or satisfied by) I and V' for some valuation V' which differs from V at most in the value it assigns to x.

The fact that this system of logic is sometimes accepted as the "standard", "classical" or "normal" logic should not hide the fact that there exist lots of both alternatives and extensions. Especially in recent decades there have emerged an immense number of new logical systems; some of them having been initiated by impulses from formal semantics.

2.5 Chomsky

Both Tarski and Carnap saw an unbridgeable gap between natural language and the formal languages of logic they investigated. They claimed that natural languages, not being exactly defined, cannot be directly studied by the mathematical means developed by logicians; and they tacitly assumed that the formal languages they dealt with were what natural language should ideally be replaced by if we want to do serious science. Moreover, later Tarski and his followers developing model theory were increasingly delving deeper into pure mathematics and losing sight of natural language.

However, at the same time and quite independently of the development of logic, a revolution within the approach to natural language, which was to result into a large scale 'mathematization' of linguistics, was being started by Noam Chomsky (1928-). Chomsky's original goal was a rigorous description of the syntax of natural language. In his 1957 path-breaking *Syntactic Structures* (Lakoff, 1971) he introduced a general framework for such a description. It is based on the concept of *generative grammar*, in effect a collection of rules understood as generating all well-formed sentences of the language being described.

The basic idea behind Chomsky's generative grammar is the idea of a *rewrite rule*. A rewrite rule simply instructs us to rewrite a sequence of symbols by another sequence of symbols. Thus, the rule

$S \rightarrow NP\ VP;$

instructs us to rewrite "*S*" by "*NP VP*". Now the idea of a generative grammar for a language *L* is the idea of a set of rewrite rules, working with the vocabulary of *L* plus some set of auxiliary symbols, such that the set of all strings which can be produced by means of the (repeatable) application of the rules to the symbol S and which contain no auxiliary symbols coincides with the set of all the well-formed sentences of L. Thus, if L were to consist of the four sentences "John walks", "Mary walks", "John whistles", "Mary whistles", one of the possible generative grammars for it would be

$S \to N\ V$

$N \to$ John

$N \to$ Mary

$V \to$ walks

$V \to$ whistles

or, in an abbreviated form,

$S \to N\ V$

$N \to$ John | Mary

$V \to$ walks | whistles

(Clearly as long as the number of sentences of the language in question is finite, there always is the trivial grammar consisting of the rules instructing us to rewrite S by every particular sentence. However, as the number of sentences of natural language is potentially infinite, grammars for them cannot be that simple.)

Chomsky then supplemented rewrite rules by the so-called transformation rules, and subsequently introduced plenty of extensions, modifications and innovations of his model which refashioned its nature several times, but the basic idea remained unchanged: the formal grammar should provide for the generation of all and only well-formed sentences of the language under consideration.

What has changed is Chomsky's interpretation of the generative and transformative rules. At first they looked as merely utensils of his theory, which did not correspond to anything real, later they ever more looked like descriptions of something to be found in human mind/brain, namely in that its part which Chomsky called the *language faculty*[20]. Thus, what originally looked like mostly an abstract mathematics, came to resemble an empirical theory of h ow language is realized in human brains.

Does Chomsky's approach go beyond what we know from the logical theories of formal languages? Not really. Consider the grammar of standard logic as summarized in the previous section. Its syntax can be, and usually is, defined in the following way:

An individual constant and a variable is a term.

If F^N is an n-ary functor and $T_1,...,T_n$ are terms, then $F^N(T_1,...,T_n)$ is a term.

If P^N is an n-ary predicate and $T_1,...,T_n$ are terms, then $P^N(T_1,...,T_n)$ is a formula.

If F_1 and F_2 are formulas, then $\neg F_1$, $F_1 \wedge F_2$, $F_1 \vee F_2$, $F_1 \rightarrow F_2$, $\forall x F_1$, $\exists x F_1$ are formulas.

This yields us, rather straightforwardly, the following generative grammar:

$S \rightarrow P^N(T,...,T) \mid \neg S \mid S \wedge S \mid S \vee S \mid S \rightarrow S \mid \forall VS \mid \exists VS$

$T \rightarrow V \mid C \mid F^N(T,...,T)$

$P^N \rightarrow ...$

$F^N \rightarrow ...$

$V \rightarrow ...$

$C \rightarrow ...$

Hence from this viewpoint, the formalization of syntax proposed by Chomsky is only a minor variation on the theme of the specification of a formal language standardly entertained in logic.

[20] See, e.g., Chomsky (1986; 1993; 2000).

However, the path-breaking import of Chomsky's approach did not consist in the shape of his grammars, but in the fact that he did not propose them to define formal languages, but rather to account for natural ones; and that he managed to persuade a substantial part of the scientific community that the syntax of a natural language can be usefully captured by formal means.

Chomsky's successful attempt at the rigorization of natural language syntax was followed by attempts at the rigorization of semantics along analogous lines. Probably the first was the so-called *generative semantics* (Lakoff, 1971); Chomsky himself then extended his theory to cover not only syntax, but other 'levels' of language as well. The idea behind such attempts was to capture 'semantic structure' in a way analogous to the one in which the syntactic structure was captured. Many linguists did take this as an acceptable approach to semantics of natural language; however, there were also protests that theories of this kind do not amount to theories of semantic interpretation. The most substantial argument, leveled e.g. by Lewis (1972), appeared to be that nothing can aspire to being a theory of *semantics* unless it yields a theory of truth conditions.

2.6 Montague and since

So around the sixties, on the one hand, there was a developed mathematical theory of natural language, but only of its syntax; and, on the other hand, there was a developed theory of semantics, but only for the standard predicate calculus, which appeared to be too simple to provide for an interesting model of natural language.

Of course there were logicians, linguists and philosophers who thought about bridging the gap. The most famous of them became the American logician Richard Montague (1930-1971) who proposed a logical system which, on the one hand, had a rigorously defined, Tarsko-Carnapian semantics, and, on the other, provided for a much more realistic model of natural language than any previous logical language. As a consequence, some theoreticians of language realized that model theory might therefore provide for an interesting explication of the semantics of natural language.

One important ingredient of this development was the construction of the semantics for modal logics due to Kripke (1963a; 1963b; 1965). Here it is where the all-important concept of *possible world* appeared as a pillar of the semantic theory. Modal logic is, roughly, the logic of necessity and possibility; and Kripke realized that to account for its semantics, we must not let the content of a sentence be exhausted by its truth-value (in the actual world), that it needs to contain also the information about its (potential) truth-values in worlds that are only possible, not actual.

Montague realized that if the model-theoretic means is to be engaged for the purpose of explication of the semantics of natural language, then it cannot stay on the level of extension. Drawing on the ideas of Carnap and Kripke, he formalized the concept of intension as, in effect, the relativization of extension to possible worlds (Montague, 1974). This is to say he presupposed a given set of possible worlds (representing ways our world might also be) and understood an intension of an expression as the function taking a possible world to the extension of the expression within the world. Thus if the extension of the singular term "the president of the USA" is the (current) president of the USA, its intension will be the function taking every possible world to the person who is the president of the USA in that world (if any); if the extension of the general term "horse" is the set of actual horses, its intension is the function which takes every possible world to the set of all the horses of the world; and if the extension of the sentence "The president of the USA is a horse" is the truth value ***Fa***, its intension is a function which takes every possible world to the truth value of the sentence in that world.

Montague furnished each expression of his model of language with a *denotation* (extension) and a *sense* (intension); and assumed that although what is essential are denotations, there are contexts in which the sense of an expression somehow assumes the place of its denotation. In particular, he introduced the operator ^ such that for any expression *E* the denotation of ^*E* is – by definition – the sense of *E*. (The dual operator ˅ then worked the other way around: the sense of ˅*E* is defined as the denotation of *E*.) The idea was that a natural language expression E could, on the level of Montague's logic, be interpreted as either *E* or ^*E*, depending on its character and the context of its occurrence.

Moreover, Montague introduced a very general framework for the formalization of languages. We have seen that whereas standard model theory assumed a discriminating stance towards the vocabulary of its language (only words of some categories were taken as designators), the Chomskian theory of syntax took expressions of all categories alike. Montague assimilated model theory for his intensional logic to the indiscriminative stance. (This was not unprecedented: this stance was adopted long ago by people studying the semantics of lambda calculus, such as Church (1940), or Henkin (1949); on which Montague's intensional logic also drew.)

This led Montague, in effect, to the vantage point from which an (uninterpreted) language appeared as a finitely generated algebra, its carrier being constituted by the well-formed expressions of the language, its generators being the words, and its operations the syntactical, formation rules. The interpretation (meaning-assignment) for such a language then appears as a homomorphism of the algebra into another algebra (of 'denotations' or 'meanings'), the requirement of homomorphism reflecting the principle of compositionality (Janssen, 1986).

The Montagovian vantage point has proved itself fruitful; and Cresswell (1973) intensional model of language has been followed by a number of elaborations and modifications. There were also alternative intensional models (due to Tichý (1978), Tichý (1978) and others) proposed partly or wholly independently of Montague's approach. Then there followed modified, 'hyperintensional' models of semantics attempting to improve on the intensional model especially to make it capable of adequately analyzing the so-called 'propositional attitude reports' (see Bigelow (1978), or Lewis (1972)). To these we can count systems based on the so called structured or Lewis-type meanings (Cresswell (1985), Tichý (1988)), Barwise & Perry (1983) theory of constructions, Barwise & Perry (1983) situation semantics etc. There further followed models reflecting the 'dynamic' aspect of natural language, such as Kamp (1981) DRT or various models based on dynamic logics (van Benthem, 1997). The early state of the art was excellently surveyed by van van Benthem & ter Meulen (1996).