An advanced introduction to

Tarski's theory of truth
&
Davidson's theory of meaning

Davor Josipović
3e bachelor wijsbegeerte
2007-2008

Abstract: I show here in what context and how Tarski defines truth. My main concern is to explicate the main components of his theory, and to give the reader a sense of what truth means in formal languages. Then I show that the notion of meaning was all but new in the formal logical tradition, and how Davidson tries to introduce it to a broader group of linguistic philosophers by putting some of basic concepts of semantics in form of a program for constructing theories of meaning for natural languages.

Prologue

My main concern in this paper is twofold and arose mainly from my logic and analytic philosophy courses. On the one side one has the concept of truth that Tarski explains in his article from 1944 to an elaborate audience. Reading it, one can feel that there is some structure, unexplained to the reader, which guides the whole article. It feels like reading a scientific report without any mention on the details of the setup, nor on how the results were obtained, only on the significance of them. All that the reader is left with are the empty word of his convention T: “x is true if, and only if p”. One can imagine that trying to explain in this way to someone what one learned from Tarski can only cause a frown on his/her forehead, and bring up the question: “oh, and that is philosophy?” One the other side one has Davidson who seems to have taken Tarski’s definition of truth, and made it into some sort of theory of meaning. So, can it be more puzzling?

My main goal in this paper is twofold. On the one hand, I want to sketch what Tarski really meant with a definition of truth, and lift the mist around the vagueness of his Convention T. On the other hand, I want to show how it can be possible to build up a theory of meaning from Tarski’s definition of truth. Before I started to write this paper, these were the unexplained facts in most introductory textbooks, which I hope will become *clear* after reading through this paper. And for the asthenic reader, it might be a refreshing thing to know in advance, that I am *not* going to use, not even once, the fact that ‘snow is white’, apart of course from this single mention, which I hope you have survived.

The way I am going to do this is with help of Kirkham’s book on theories of truth, a few Davidson’s early articles, and Tarski’s original work.

Introduction to truth and meaning

Maybe the best way to start this paper is to first introduce the reader to the subject I will be dealing with within the next pages. The sole reason for this introductory chapter is the fact that one could know a lot about a small and almost insignificant thing, without seeing its place in the big picture. That’s why it may be instructive to read about other ways of seeing truth and meaning and dealing with them. Giving a complete overview of all the possible uses of meaning and truth would require the volume of a book, rather than a paper. That is why I will restrict myself here to a brief overview of the development of these two concepts in the past hundred years of analytic philosophy. But before I do that, I will give a short preliminary exposé on truth.

Truth

The concept of truth is very vague. Throughout the history, it has been viewed from different perspectives: roughly the mythological, theological and scientifical. One of the properties of myths is that they were seen in ancient times as true explanations of some socio-political order or ritual practice. Theological conception of truth was most explicitly used during the middle ages when Christian, Muslim and Jewish theologians all had their holy scriptures that were taken to hold the ultimate truth and basis for all their reasoning. And lastly, the scientific concept of truth, which is vastly empirical and mathematical in nature, is most explicitly found in our modern age. To put the philosophical concept of truth as the fourth 'type' of truth would be a mistake as it would cut across the very nature of philosophy, which is to expound and give an account to those different conceptions rather than to create a fourth one.

What a philosopher needs to explain is what truth is (i.e. what it consists in when we say that something is true), giving (preferably) account to all the different perceptions. And an explanation on quiddity presupposes a theory. There are a few philosophical projects that try to do just this. Kirkham categorizes them as taking part in three different projects: the speech-act project, the metaphysical project and the justification project. The speech-act project "attempts to describe the locutionary or illocutionary purpose served by utterances that by their surface grammar appear to ascribe the property of truth to some statement (or belief, etc.), for example, utterances like 'Statement *s* is true'." (Kirkham, p.21) The metaphysical project (i.e. the one elucidated in this paper) "attempts to identify what truth consists in, what it is for a statement (or belief or proposition, etc.) to be true." (Kirkham, p.20) And finally, the justification project "attempts to identify some characteristic, possessed by most true statements and not possessed by most false statements, by reference to which the probable truth or falsity of the statement can be judged." (Kirkham, p.20) It is obvious that the *validity* of the project depends highly on the common conception of truth, which as mentioned, is time dependant.

Now, for our further discussion, only the metaphysical project is relevant. As already mentioned, this project tries to define is some way the predicate 'is true'. Note that I am speaking here of a definition in its broadest sense. Its purpose is to describe a term by the use of terms that are commonly understood and whose meaning is clear. So the *definiens* obviously should be informative and noncircular. The relation between the *definiens* and *definiendum* is one of equivalence. Kirkham distinguishes four kinds of equivalence relations.[1] Only the extensional equivalence is relevant for our current discussion. The extension of 'is red' is the set of all red things. Generally, the extension of 'some predicate' "is the set of all objects to which the predicate truly applies." (Kirkham, p.4) An extension (and thus extensional definition) of 'is true' would be the set of all true things. But note that this is not very informative as it is circular. "The attempt to produce such a noncircular description of the set of all true things (the extension of the predicate 'is true') is what [we] shall call the extensional project." (Kirkham, p.4), which is a subbranch of the metaphysical project.

The atmosphere in analytic philosophy in which we have to situate the logician Tarski is one of Logical positivism, preceded by Wittgenstein's *Tractatus Logico-Philosophicus* and Russell's Atomism, founded on logical works of Boole and Frege during the very beginning of the 20th century. In this atmosphere Tarsi sets out "to construct — with reference to a given language — *a materially adequate and formally correct definition of the term 'true sentence'*" (Tarski 1933, p.152) More specifically, he showed how to give an informative, unambiguous and noncircular extensional definition of truth for the languages of symbolic logic. "Main lines of research developed [during the 40s and 50s] involved enriching the formal languages of symbolic logic amendable to Tarski's techniques. [… and] by the early sixties it had become possible to imagine the day in

---

[1] Intensional eq. → essential eq. → natural eq. → extensional equivalence, where '→' stands for material implication. So it follows that if x and y are intentionally equivalent, they are necessary also essentially, naturally and extensionally equivalent. (Kirkham, 'Chapter 1')

which natural languages like English would be treatable in something close to their entirety by descendants of the logical techniques initiated by Tarski." (Soames, p.294) Still, some were skeptical about it.

Meaning

Two different movements emerged during the early 1950s, both from Wittgenstein's *Philosophical Investigations*. The first was that most philosophical problems are language problems. The second one was that meaning was not to be studied from a theoretical perspective, but rather through careful analysis of particular words in particular contexts. During the 1950s and early 1960 "it became increasingly clear that what is, and what is not part of the meaning of a word, is not a transparent matter. Because of this, questions about meaning can be investigated profitably only with the help of some systematic theory." (Soames, p.292) But as Soames notes:

> […] it was unclear whether such a [systematic] theory was possible, or, if it was, what it should look like. At the time, skepticism on the matter was fueled by Quine's highly influential arguments in *Word and Object* and *Ontological Relativity and Other Essays*, which reject our ordinary notions of meaning and reference as scientifically hopeless, while proposing radically deflated substitutes. Meaning, as Quine conceived of it, was not the center of anything, certainly not philosophy. However, his was not the only voice. In the early 1960s an important development took place. Philosophers working in a different tradition, growing out of the development of formal logic, came up with a philosophical conception of meaning that many found irresistible. The conception was formulated by Donald Davidson, who conceived of a theory of meaning as a systematic theory of the truth conditions of the sentences of a language. To many this seemed like precisely the thing that was needed to fulfill the conception of philosophy as the analysis of meaning — no matter that the conception of meaning employed was a descendant of one that the later Wittgenstein and the ordinary language philosophers who followed him had earlier rejected as irrelevant. (Soames, p.xiv)

According to Soames, one of Davidson's main contributions was that he made some of the simple techniques that were employed with formal languages accessible and relevant to those philosophers who had long studied the natural language, but who were also relative outsiders to the formal, logical tradition. This approach that, as I will sketch, proves fruitful in the specialized domain, he has put in form of a program for constructing empirical theories of meaning for natural languages. "In so doing, he helped establish significant contact between two different philosophical subcultures in analytic philosophy." (Soames, p.295)

On truth and meaning

Davidson's search for an adequate theory of meaning[2]

Davidson's main concern for a theory of meaning was to explain the mystery of natural language, which is the fact that we can produce and understand an infinite number of sentences, even though we are finite beings able to handle finite number of axioms and rules. In his understanding of language, (1) this implies knowing the grammar[3] and a finite amount of semantical primitives. So a theory of meaning should be able to give a constructive account of the meaning of sentences in a particular language. The reason why he disapproves other theories, and creates his own one, is because "a number of current theories of meaning do either conflict or ignore the condition of being learnable". (Davidson, p.3) He gives a few examples that I can only mention here, such as the impossibility to define the quotation marks as a rule, which makes them unlearnable.[4]

First and most obvious question would be: what does 'meaning' mean to Davidson? This is probably a wrong question. Just as a particular theory of nature will tells us what

---

[2] I will base myself here mainly on Davidson, "Truth and Meaning", where one finds the basic tenets of his theory of meaning.

[3] i.e. a set of rules by which we can constitute meaning through the use of semantic primitives. Later, this notion, which does not simply refer to school grammar, will be better explained.

[4] This is probably the reason why he doesn't allow the sentences (of object language) that are being referred to in his theory, to be names of these sentences, but only structural descriptions.

‘nature’ is, so will a particular theory of meaning tell us what ‘meaning’ is. So the concept of meaning has some particular sense in a theory of meaning, and the sense of it will largely be determined by the purpose for what the theory is made for and by the constraints that judge whether the theory is acceptable or not. Davidson does have a theory of meaning. Its purpose is to explain our understanding, and one of its constraints, as noted, is that it must account for the learnability of languages.

Second thing we need to know before we go further into the details, is what he conceived as the task of a theory of meaning. As already mentioned above, the main task of it is to explain the possibility of meaningful sentences in a particular language. This leads him to put to his theory two extra goals that it must satisfy if it is ever to be called a theory of meaning.

> (2) “It should provide an interpretation of all utterances, actual and potential, of a speaker or group of speakers” and
>
> (3) it should be “verifiable without knowledge of the detailed propositional attitudes of the speaker.” (Davidson, p.xv)

As competent speakers of some language we can construct infinite number of meaningful sentences from a finite number of words. So, he holds that one must show how the meaning of sentences (whole) depends on meaning of words (atomic parts). This follows from his constraint of learnability of languages. Hereby the words (atomic parts) do not need to have meaning on their own[5]. Further on, as we want to define the meaning of an arbitrary sentence, we want a theory of the following form: “’s means m’ where ’s’ is replaced by a structural description of a sentence and ‘m’ is replaced by a singular term that refers to the meaning of that sentence; a theory, moreover, that provides an effective method for arriving at the meaning of an arbitrary sentence structurally described.”

---

[5] Davidson argues that even if we presuppose a dictionary where we can find all the meanings of words, we could not extract from them, and an adequate syntax, the meanings of belief-sentences. (Davidson, p.21) Another, more worked out example he gives, is that the complex term ‘the father of’ doesn’t need to correspond or mean anything. It gets it’s meaning only if there is some name attached to it: ‘is the father of x’.

(Davidson, p.20) This form also implicitly supposes that we have a theory of syntax for our language that can tell us whether the structural description is meaningful or not. Further on, Davidson also clarifies the form 's means m' by replacing it with 's means that p' (with p a sentence) to make sure that meanings are not named.[6] So, what we have is that the sentence p in metalanguage gives the meaning of s. (If English was our object language, then our metalanguage could consist of English and the structural description connector, such that p would be an English sentence and s the structural description that corresponds with some English sentence in the object language.) Therefore, if we presuppose an adequate syntax (i.e. a system of rules for providing meaningful sentences) and a finite amount of semantic primitives, we can give for every structural description s some kind of a pattern consisting of these rules and semantic primitives, which constitutes the meaning of s. Also, every competent speaker should have implicit knowledge of these rules, and every speaker explicitly knowing the rules and semantic primitives would be able to understand the language if the equivalences between sentences and his knowledge were given to him (and he could reason fast enough).

The final bold step, as he himself calls it, is to try treating the position occupied by 'p' extensionally. The problem is the 'means that' because it presupposes the semantic concept of meaning, and that is just what we want to explain[7]. So het chooses to replace the 'means that' with 'is T ≡'[8] (I will call this latter form the Convention M). In Kirkham's words, "the semantic content has thus been pulled out of 'means that' and put

[6] Sentences cannot name meanings, they get their meanings in and through the language in which they function. So sentences such as p, prefixed with word 'that' are not names at all.

[7] Here is what Davidson has to say about it. "It looks as though we are in trouble on another count, however, for it is reasonable to expect that in wrestling with the logic of the apparently non-extensional 'means that' we will encounter problems as hard as, or perhaps identical with, the problems our theory is out to solve. The only way I know to deal with this difficulty is simple, and radical. Anxiety that we are enmeshed in the intensional springs from using the words 'means that' as filling between description of sentence and sentence, but it may be that the success of our venture depends not on the filling but on what it fills." (Davidson, p.22) What follows is a piece of rhetoric that amounts to Tarski's Convention T.

[8] I follow Kirkham in his notation of extensional equivalence as '≡'.

into the predicate 'is T', which thereby allows p to be an extensional *definiens* of the semantic property T.", for all sentences of L. (p.226) By doing so Davidson lets his theory of meaning meet Tarski's material adequacy condition, so "that the sentences to which the predicate 'is T' applies are just the true sentences of L". (Davidson, p.22-23) Therefore 'T' can be replaced by 'true'.

The reasoning behind this last equivalence (Convention T ≡ Convention M) might be the following. Davidson requires that for every meaningful sentence, one can construct Convention M: 's is T ≡ p'. Tarski, as we shall see, requires that for every correctly formulated sentence s, one can construct Convention T: 's is true ≡ p'[9]. It would follow, if as we shall see there is some same correlation between p and s for both conventions, that T ≡ true. This means that if one knows the semantic concept of truth for a particular language, one automatically understands it, and *vice versa*. As there might be many p's with which s is extensionally equivalent, giving them all represents the full meaning of s. Same holds for truth.

Truth for Tarski

Any program that people devote themselves to, puts some constraints to existing theories that they are willing to incorporate in that program. In case of Davidson's program, the learnability of language is of utmost importance. So, on basis of the fact that "a number of current theories of meaning do either conflict with or ignore this condition for being learnable" (Davidson, p.3) he rejects them. Let us call such constraints 'criteria of adequacy'. Now, "Tarski wanted a theory that would meet what he calls the material adequacy condition. […] The condition asserts simply that any good theory of truth has to *entail* all sentences on the pattern of the following:

'The wall is red' is true ≡ the wall is red

[9] Note here that for Tarski it doesn't matter whether the sentence being referred to in the object language by s is a name or a structural description. Note also for later that I will write 's' to explicitly denote the structural description in metalanguage of sentence s of the object-language.

‘Snow is slippery’ is true ≡ snow is slippery […]”[10] (Kirkham, p.143)

On the left side, the sentence is mentioned; on the right side the sentence is used. Further, the equivalence is extensional in nature.[11] The pattern, ‘x is true if, and only if, p’, he calls the Convention T. The reason why he puts this constraint on his theory of truth seems pretty obvious: any theory that doesn’t comply with Convention T would be odd, as it would seem to exclude something we feel is should not, and any theory that doesn’t comply with material adequacy condition, would seem incomplete. In other words, Convention T is a test on basis of which we accept or reject a proposed definition of truth.

Before I go further, I’ll have to explain what it means for a sentence to be mentioned on one side, and used on the other, in Tarski’s terms. Doing this in a formally correct way, as Tarski did, would extend this paper beyond the required scope. That’s why I will restrict myself to the most important aspects of his work, and give a brief outline of the most important subjects in his work. In doing so, I will try to declare the meaning of every technical word on first use, either as footnote or between brackets, as to preserve clarity.

Tarski’s main concern for a definition of truth is the formal language. So first we have to specify what is meant with a formal language. In general, a formal language consists of:

(i) a non-empty set of primitives

(ii) a set of statements about the primitives, the axioms

(iii) a means of deriving further statements from the axioms[12]

---

[10] One can find the definition of Convention T in Tarski 1933, p.187-188. It is also clear that he perceives truth as a property of a sentence. Truth is in effect a class for him where a sentence x is true ≡ x is an element of the class of all true sentences of some language.

[11] He mentions on few occasions, that he wants to define truth extensionally. This is suggested for example in his introduction (1933, p.152) and also analyzed by Kirkham (Kirkham, p.33)

[12] Which consist either in “(a) an explicit set of recursive rules of derivation; or (b) appeal to a background logic for the language in which the axioms are stated, usually first-order logic [which itself can be axiomatized]; or (c) no explicit means of derivation; one is to derive "whatever logically follows" from the

Take, as for an example, the following set of primitives: ‘p’, ‘q’, ‘$\neg$’, ‘$\vee$’, ’(‘, ‘)’.

Take the following statement as the only axiom:

(A_a) $\neg p \vee (p \vee q)$

And the following two inference rules:

(R_a) Rule of Substitution: For a variable in any statement Q we may substitute a statement P, provided that P is substituted for every occurrence of that variable in Q, where Q and P are variables or statements of arbitrary complexity.

(R_b) Modus Ponens: From X and $\neg X \vee Y$ infer Y, where X and Y are variables or statements of arbitrary complexity.

We can use axioms and inference rules to deduce other valid statements. For example, if we take P as a theorem (i.e. statement established by means of a proof from the axioms, with inference rules), and Q as any statement, then $P \vee Q$ is also a theorem. Proof:

(1)	P	(Premise)
(2)	$\neg p \vee (p \vee q)$	(Axiom A_a)
(3)	$\neg P \vee (P \vee Q)$	(R_a: p by P and q by Q in (2))
(4)	$P \vee Q$	(R_b: from (1) [= X] and (3) [= $\neg X \vee Y$] we get (4) [=Y])

If we extend our axiom system with three others (cf. Partee, p.220), then we have axiomatized statement logic (i.e. proposition logic), and one can show that all the known tautologies (like $\neg P \vee P$) can be deduced from them (i.e. the logical axioms) in the same way as we did in the previous example.[13] In itself, the primitives, axioms and inference rules have no meaning. This means that we can deduce plenty of statements from the axioms making use of only the given inference rules (this is called formally deducible or provable), like the one above, without knowing what those statements mean. [14]

---

axioms.” (Partee, p.92) (cf. Tarski 1933, p.166, which is approximately the same characterization of a formal language)

[13] For more examples, see Partee, ‘8.6.1 An axiomatization of statement logic’.

[14] The used signs $\vee$ and $\neg$ might seem familiar, but I could have used any other like ╢ and ╬, which do not seem familiar, and have no associated common meaning.

Meaning one gets by assigning a specific model to a language. More specifically, a model consists of a set *D* of objects (i.e. that about which the language speaks) and a function F (= assignment-function) which assigns, among other things, to each constant a member of *D* and for example to each predicate-symbol of the theory a subset of the model.[15] Now we have rules that tell us how to evaluate sentences from the theory in this model. This is done with the interpretation function I, which is consistent with F, but extends it by, among other things, also defining the truth-table (i.e. *truth*) of logical vocabulary and thus sentences. In logic, this function I is usually taken as granted and has exactly defined truth for all logical vocabulary.[16] It should be clear now that *truth* of a statement is dependant on a specific interpretation (i.e. on functions F and I). In other words, the concept of a true statement doesn't exist in syntactic activity, only in semantic activity.[17] To return to our previous example; saying that our theorem $P \vee Q$ is true under provisional truth of P, is a semantic argument, and presupposes the truth-table of the $\vee$-operator (i.e. its interpretation), while deducing it from our system of axioms is a syntactical argument. Also, saying that $\neg P \vee P$ is a tautology (i.e. statement that is true in every possible model, i.e. true simply because of the meaning of the connectives $\neg$ and $\vee$)

---

[15] More elaborate characterization of function F one can find in Partee, p.143 or Horsten, p.27, both for first-order logic. Informally, the assignment function relates all non-logical vocabulary with the model, for example: every constant of the theory with an object of the model, every predicate-symbol with a set of ordered n-tuples of elements from D, where n is the number of predicate-placeholders, every function-symbol with a function of the model,…

[16] For an example of how this interpretation function may look like, see Horsten, p.28. Informally, this function tells us that P(a) is true if, and only if, the object to which 'a' refers (in the domain of the model) has the property expressed by P, that b = c if, and only if b and c refer to the same object,…

[17] Although there is no perfect distinction between pure syntactical and semantical activity, there are some general agreements. Syntactic activity is usually "the construction of proofs from premises or axioms according to formal rules of inference or rewriting rules […] while demonstrating that a certain set of axioms is consistent by showing that it has a model is giving a semantic argument. On the syntactic side are wellformedness rules, derivations, proofs, and other notions definable in terms of the forms of expressions. On the semantic side are notions like truth and reference, properties which expressions may have relative to one model or interpretation and fail to have with respect to another." (Partee, p.94)

is a semantical argument, while deducing it from the four axioms of statement logic is a syntactical argument. Now, because one can deduce only tautological statements from the four axioms of statement logic, and tautologies are true no matter the truth-value of their constituents, we can say that propositional logic (eventually also first-order logic)[18] is true in *every* model. Extending the logical[19] axioms with non-logical[20] axioms usually means we introduce some non-logical vocabulary on which we can specify function F and thus various meanings. A non-tautological statement that is formally *deduced* from a set of axioms is true or false *relative* to a model.[21] So, not every model complies with a particular formal theory.[22] "Finding a model for a [formal] theory requires finding some abstract or concrete structured domain and an interpretation for all of the primitive expressions of the theory in that domain such that on that interpretation, all of the statements in the theory [i.e. statements which are formally deduced from the axioms] come out *true* for that model on that interpretation." (Partee, p.200)

One can ask oneself if there really is a one-on-one correspondence between syntax and semantics such as we have sketched here. In other words; will a formally deduced statement also be *true* in every model that complies with its axioms? In 1929, Gödel, two

[18] Same reasoning we can follow for the first-order logic. (Note that first-order logic is sometimes called predicate logic.)

[19] For (logical) axioms of proposition logic, see Partee, p.220, and for (logical) axioms of first-order logic, see Partee, p.225-226

[20] Note that non-logical axioms does *not* mean that the axioms are not logical (i.e. strange or whatever). It just means that they are not tautological or not trivial (i.e. not deducible from the four axioms in case of proposition logic).

[21] Cf. "If a theory has an axiomatic characterization, something is a model for that theory iff [i.e. if, and only if] it is a model for the axioms." (Partee, p.200)

[22] For a fully worked out example of an axiom system, and its interpretation, one can refer to Partee, '8.5.4 An elementary formal system'. For an example of different interpretations (good and bad) of axiomatized systems such as Peano's axioms, one can refer to Partee, '8.5.7 Models for Peano's axioms'.

years before proving the two incompleteness theorems, proved that, for fist-order logic[23], every formally deducible statement is also semantically deducible (i.e. true in *every* model of the system), and *vice versa* (i.e. they are equivalent). This will prove important for the future. It allows me to present a first-order logic model, and forget about its exact formalization, although one can assume there exists a formalization for it (and all other isomorphic models). So, in other words, this allows concentrating on the semantics (i.e. metalanguage) and forgetting about syntax (i.e. object language). [24]

Now we are able to understand every aspect of Convention T. Tarski chooses as his first object language the 'calculus of classes'. In metalanguage, one has for every expression of the object language a name and a translation (or interpretation if the language is purely formal). I shall not go into detail here as doing so would extend this paper too much.[25] What is important, is that "to every sentence of the language of the calculus of classes there corresponds in the metalanguage not only a name of this sentence of the structural-descriptive kind, but also a sentence having the same meaning [which one can define exactly]." (Tarski 1933, p.187) So, to relate it to the Convention T, the 's' stands for the structural description of some sentence of the object language, and p stands for the translation of it into metalanguage (which is a meaningful expression for us). For example, corresponding to the sentence 'ΠyIy,x' (in the object language) is the name[26] '$(\forall x)(x \in X)$', and the sentence 'for any element x we have that x is an element of X'.

---

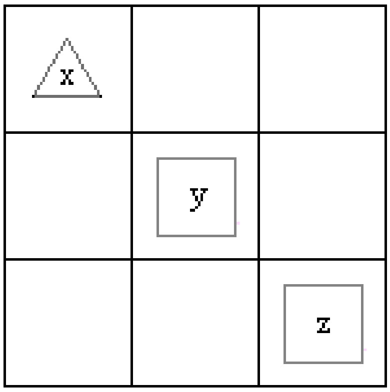
[23] 'first-order' means that quantifiers bind individual variables but not variables ranging over predicates of individuals. So first-order logic is what is known as predicate logic.

[24] Of course, this does not always have to the case. We can use our metalanguage in turn and consider it the object language of some meta-metalanguage. Nonetheless, the most 'basic' distinction one can draw between object- and metalanguage is the one where the object language has no meaning at all (i.e. is pure syntax). And as we consider only formal languages as our object languages in this chapter, the used distinction is further supported.

[25] For details see Tarski 1933, p.168-185. He takes there as formal (object-)language the 'calculus of classes' and defines in these pages the interpretation of a constant, variable, axiom, consequence and theorem in the metalanguage.

[26] Whether it's a name of s, or a structural description of s, it doesn't matter to Tarski.

So the Convention T sentences can be seen as "belonging to the metalanguage and explain in a precise way, in accordance with linguistic usage, the meaning of phrases of the form 'x is a true sentence' which occur in them." (Tarski 1933, p.187)



Let's take for the rest of our excursion in this chapter as our simplified world the model (image) on the right. This will be our model that will give meaning for the formal languages we will design. (I will call this particular model 'model M' or just 'M' in the future.) Its purpose is a representation of a possible situation. Note, that in order for the model to give meaning, the structure of the formal language must correspond with the structure of the model. So, our model must have a domain, that about which a language speaks and an interpretation for each predicate. As its domain *D* we take the objects x, y and z. Further, two shapes and nine chessboard places represent some quality of the objects. We say that M is a model for some formal theory (i.e. formal language) when all the axioms are true in M.

All following languages which I will compose will have axioms that are true in M (i.e. model M can function as an interpretation for that language). I will treat them in meta-language as Convention T requires no object language (but only that there exists a defined one-on-one correspondence between all sentences s in object language and all names 's' and sentences p in metalanguage, which as we already have insured on page 13, does exist). I'll start with a finite language L_1 composed of three (and not one more) sentences (axioms), being a, b, and c, and extend it, along language L_2 (statement logic), to a language L_3 (first-order logic) composed of infinite sentences formed by quantifiers and truth-functional operators, and define truth in each of them. Besides, axioms that compose the proposition and first-order logic are, as already noted, true in all models. But our model has more than those axioms. 'x is a triangle' could be an axiom in our model, but in some other, it would not be necessarily. So our model can be assigned more than those logical axioms. They are called postulates or non-logical axioms and their aim is to capture what is special about a particular structure.

Let's start with language L_1. It is composed of sentences a, b and c. Correlate 'a' to sentence 'z is a square', 'b' to sentence 'y is between x and z' and 'c' to sentence 'x is a triangle'. Note that 'a' is the structural description of the sentence a (i.e. axiom a) in the object language L_1, 'z is a square' is a's meaning in the metalanguage. Note that we are talking here about L_1 on the semantic level, so is obvious that all three of these sentences are true in our model. (We can verify this by looking at M.) According to the material adequacy condition, one could simply define truth for this simple language L_1 by a logical conjunction of all the T-sentences[27]. Simplified, the result would be equivalent to the following:

$$\begin{aligned}\text{For all senteces s of language } L_1 \ [\text{'s' is true} \equiv\ & (\text{'s'} = \text{'a' and z is a square}) \\ & \text{or } (\text{'s'} = \text{'b' and y is between x and z}) \\ & \text{or } (\text{'s'} = \text{'c' and x is a triangle})]\end{aligned}$$

Now, suppose the language L_2 has the three sentences from L_1 and two truth-functional operators with their structural descriptors being respectively '$\neg$' and '$\wedge$', ant their metalanguage meanings being respectively 'not…' and '…and…'. The object language in this case would consist in the logical axioms of statement logic and the three non-logical axioms a, b and c. With these operators and our three basic sentences, one could form an infinite[28] number of sentences, which is a property of natural language. One would have to use a recursive definition to define truth here. Doing so would yield:

$$\begin{aligned}\text{For all senteces s of language } L_2 \ [\text{'s' is true} \equiv\ & (\text{'s'} = \text{'a' and z is a square}) \\ & \text{or } (\text{'s'} = \text{'b' and y is between x and z}) \\ & \text{or } (\text{'s'} = \text{'c' and x is a triangle})] \\ & \text{or } (\text{'s'} = \text{'}\neg\text{ p' and it is not true that p}) \\ & \text{or } (\text{'s'} = \text{'p} \wedge \text{q' and p is true and q is true})]\end{aligned}$$

Now, the third step in complicating the language involves extending L_2 with quantifiers. This implies treating the nouns in 'a', 'b' and 'c' as variables. (This implies a, b and c

---

[27] "[…] complete definition [of truth] would be a "logical conjunction" or "logical product" of all of them [T-sentences] (Kirkham, p.145 citing Tarski 1944, p.16 and Tarski 1933, p.187)

[28] Sentences such as 'a $\wedge$ a' or 'a $\wedge$ a $\wedge$ a' or etc. count as genuine distinct sentences.

will become relation identifiers with a number of placeholders.) Let us define them in the following way:

a(z) means ‘z is a square’ so a’s name in meta-language becomes ‘… is a square’, which I will indicate with ‘Square(…)’,

b(y,x,z) means ‘y is between x and z’ so b’s name in meta-language becomes ‘… lies between … and …’, which I will indicate with ‘IsBetween (…,…,…)’,

c(x) means ‘x is a triangle’ so c’s name in meta-language becomes ‘… is a triangle’ which I will indicate with ‘Triangle(…)’,

where ‘…’ are placeholders for variables. Having variables means we can bind them with quantifiers or replace them with objects (i.e. x, y or z in our model).

The big task now is to define truth in this language. We cannot treat the noncompound[29] quantified sentences here as we treated the three basic sentences in L_2, because according to the latter treatment, sentence such as $\forall(i)[b(i) \wedge c(i)]$ would not be construable (only something like $\forall(i)b(i) \wedge \forall(j)c(j)$ would be construable, which is not the same), although they would exist in L_3. So, as we cannot define truth this way for any possible truth-bearing sentence (which in turn means our adequacy criterion is not fulfilled) means we have to look at another way for defining truth in L_3. We *can* treat the noncompound open[30] and quantified sentences as we treated the three basic sentences in L_2 (i.e. defining truth recursively for each of them) and fulfill the material adequacy condition, but in doing so we create another problem: “since the open sentences […] have no truth value, *we cannot recursively define the truth of such sentences in terms of the truth value of its parts*.” (Kirkham, p.152) What we could do, is introduce “a more general concept which is applicable to any sentential function [i.e. open or closed sentence], can be recursively defined, and, when applied to sentences, leads us directly to the concept of truth [i.e. is

---

[29] Sentences such as ‘p’ and ‘q’ are noncompound. Sentences such as ‘$p \vee p$’ are compound (i.e. composed with truth-functional operators).

[30] An open sentence is a sentence which has at least one variable that is not bound by a quantifier, and which in turn has no truth-value. Closed sentence on the other hand is any sentence that has truth-value.

also not in contradiction with Convention T].” (Tarski, p.189) So the definition of truth would look like this:

For all senteces s of language L_3 ['s' is true $\equiv$'s' has 'some property']

“These requirements [see previous citation] are met by the notion of the satisfaction of a given sentential function by given objects.” (Tarski, p.189) So the definition of truth in L_3 becomes:

For all senteces s of language L_3 ['s' is true $\equiv$'s' is satisfied by some sequence S of objects]

In the proposed definition we see four concepts that demand explanation. Those are ‘satisfied’, ‘sequences’ (i.e. infinite sequences) and ‘objects’. The meaning of ‘objects’ should be clear by now. It refers to the objects x, y and z of M’s domain in our case. The other two require a more extensive explanation.

An infinite sequence S is a one-to-many[31] relation whose domain is the set of all objects of M and whose counter domain is the set of all natural numbers excluding zero.[32] Intuitively, we could describe it as consisting of an infinite amount of references S_x to objects in M, where x denotes a natural number.[33] Two infinite sequences S and T are equivalent if, and only if for all natural numbers x holds that $T_x = S_x$.

---

[31] “If, for every object y belonging to the counter domain of a two-termed relation R, there is only one object x such that xRy, then the relation R is called *one-many*.” (Tarski, p.171)

[32] In more technical terms, this relation forms a set S of 2nd-tuplets <x,y>, which is a subset of $D \times \mathbb{N}$. If the counter domain consists of natural numbers ranging from 1 to n, as is the case for a finite sequence, then S consists of n 2nd-tuplets {<x,1> , <x,2> , … , <x,n>} where x denotes some object of domain *D*, dependant on sequence S. When we write S_1, it refers to x from tuplet <x,1> in set S. We can use S_1 as a variable, where 1 refers to the first tuplet <x,1> of S, and S is some infinite sequence (functioning as a variable here) such that x depends on S. So if T and S are two infinite sequences denoting subsets of $D \times \mathbb{N}$, then $T_1 = S_1$ if, and only if, object denoted by the first tuplet of T is the same as the object denoted by the first tuplet of S (i.e. if, and only if, T_1 and S_1 denote the same object)

[33] Note that this doesn’t imply we must have an infinite amount of objects in *D*. Various S_x, with x a natural number and S any sequence, can denote the same object.

The meaning of satisfaction I will only try to make clear by means of some examples. As we have four[34] kinds of sentences in first-order logic, we have to explain satisfaction for each of them.

Intuitively[35], we say that S_1 of S satisfies the predicate Square(x) if, and only if S_1 is a square and S_1 has the name x.[36] In the same way a infinite sequence S satisfies the predicate Square(i_1) (i.e. an open sentence with variable i_1) if, and only if S_1 (i.e. the object denoted by S_1) is a square. Thus, if S_1 refers to z, then the open sentence is satisfied by sequence S. "Since satisfaction is a relation between expressions and parts of the world, it counts as a semantic concept," (Kirkham, p.153) and this relation we verify on the semantic level. A universally quantified sentence is satisfied by sequence S if, and only if it is satisfied by all sequences differing from S in at most the specified place. A sentence such as $\forall(i_2)$ [Square(i_2)] is satisfied by sequence S if, and only if every possible sequence that differs[37] from S in at most the second term satisfies the open sentence 'i_2 is a square' (i.e. if all objects in our world are square). In our model M, this sentence is clearly not satisfied by S as we can construct sequence T such that T_2 refers to x which is not square. The satisfaction of an existentially quantified sentence we can understand by noting that the existential quantifier can be written with the universal quantifier and negation (cf. Partee, p.148). For example, a sentence such as $\exists(i_2)$ [Square(i_2)] is equivalent with $\neg\forall(i_2)\neg$[Square(i_2)]. Intuitively, if $\neg$[Square(i_2)] is not

---

[34] As already noted, those are open, universally and existentially quantified, and nonquantified sentences.

[35] I will not go into details in this paragraph. For an elaboration on common use of satisfaction, it's definition for the language of calculus of classes, and one example of it's use, see Tarski, p.189-196.

[36] Notice here that we are back in L_2. Satisfaction and truth are identical concepts for that language.

[37] Note that we *could* also say in this *particular* case "A sentence such as $\forall(i_2)$ [Square(i_2)] is satisfied by sequence S if, and only if every possible sequence ~~that differs from S in at most the second term~~ satisfies the open sentence 'i_2 is a square'." Reason why 'that differs from S in at most the second term' matters, is that we cannot say the above sentence for a open sentence such as $\forall(i_2)$ [IsBetween(i_1,i_2,i_3)], as it would imply too much (i.e. everything is in between of everything). That's why one could say that 'that differs from S in at most the second term' captures the sophistication of the universal quantifier and will be important for the recursive definition of it.

satisfied by all sequences which differ at most the second term, then [Square(i_2)] is by them all, so $\exists(i_2)$ [Square(i_2)] is satisfied by all such sequences.

Now we know everything we need to construct a recursive definition of satisfaction in our first-order language L_3 consisting of six predicates, two truth-functional operators and two quantifiers[38]:

For all senteces φ of language L_3 ['φ' is satisfied by some infinite sequence S $\equiv$
('φ' = 'x-izes(i_m)' and the m-th term in S refers to object with name x)
or ('φ' = 'y-izes(i_m)' and the m-th term in S refers to object with name y)
or ('φ' = 'z-izes(i_m)' and the m-th term in S refers to object with name z)
or ('φ' = 'Triangle(i_m)' and the m-th term in S refers to object which is a triangle)
or ('φ' = 'Square(i_m)' and the m-th term in S refers to object which is a square)
or ('φ' = 'IsBetween(i_m, i_n, i_o)' and the m-th is between the n-th and o-th refered object in S)
or ('φ' = '$\neg\phi$' and S does not satisfy ϕ)
or ('φ' = '$\psi \wedge \phi$' and S satisfies ψ and S satisfies ϕ)
or ('φ' = '$\forall(i_m)\phi$', and every sequance that differs from S in at most the m-th place satisfies ϕ)
or ('φ' = '$\exists(i_m)\phi$' = '$\neg\forall(i_m)\neg\phi$']

Now, one may ask why all this fuss about satisfaction. Working in metalanguage means we already know the meaning of a sentence (because of interpretation and the metalanguage is English), so we might equally well look at the world and convince ourselves of the truth of some sentence. We might equally well prove that a sentence has no counter example with Beth Tableaux (i.e. that it is true = semantic argument) (cf. Partee, p.165-170) or we might even try to deduce it formally from the axioms (= syntactic argument). These are all good arguments for verifying whether some sentence is true or false. But one can hereby make two remarks. Firstly, Tarski wanted to define truth in a formally correct way. As we have already seen, defining truth recursively in first-

---

[38] ϕ and ψ are (open) sentences, m, n, o range over integers, and the names of the objects are taken to be properties of objects, denoted by predicates x-izes, y-izes and z-izes, as W.V.O. Quine suggested for a language with names. Tarski himself "did not try to define satisfaction or truth for any language with names." (Kirkham, p.160)

order logic is not feasible. Property 'satisfaction' allowed him to do so. Secondly, in first-order logic, it is true that every sentence that is true is also formally deducible. But higher-order logic, or calculus of classes that Tarski first considers, are not complete systems. It means we could have two sentences φ and $\neg\varphi$ which are both not provable, although one of the two is true. In other words, the set of provable sentences forms a subset of true sentences. Tarski rejected this apparent equivalence of provability and truth for the following reason: "no definition of true sentence which is in agreement with the ordinary usage of language should have any consequences which contradict the principle of the excluded middle." (Tarski, p.186)

How truth leads to meaning

In Tarski's whole elaboration on truth in formal languages, there's one condition at work, presupposed by Tarski, but not made explicit in his definition of Convention-T. "Material adequacy condition would not be credible as a criterion of adequacy for a theory of truth unless it is specified that the sentence of which truth is being predicated be the *same* sentence that asserts that the state of affairs p obtains." (Kirkham, p.164) This very fact Tarski made clear by saying that 'p' should be replaced by the corresponding translation of the sentence s into the metalanguage, and 's' replaced by the name of the sentence, after he defined the Convention T. So Tarski presupposed some correlation between s and p in his Convention "'s' is true ≡ p'".

One can track this correlation back to his original semantic definition of truth on which he based his theory. "Amongst the manifold efforts which the construction of a correct definition of truth for the sentences of colloquial language has called forth, perhaps the most natural is the search for a *semantical definition*. By this I mean a definition which we can express in the following words:

(I) *a true sentence is one which says that the state of affairs is so and so, and the state of affairs is so and so*."[39] (Tarski 1933, p.155)

One could put this in a more conveniently arranged way:

---

[39] He mentions that this definition I not new, and compares it with the well-known words of Aristotle.

(II) ‘s’ is true ≡ (s means p) and p

and compare it to Convention T:

(III) ‘s’ is true ≡ p.

Tarski considers such a form (II), but he rejects it because it contains a semantic term (i.e. ‘means’) in the *definiens*.[40] According to Kirkham, his model-theoretic and physicalist ambitions direct him to do that.

Now, let us see what (II) implies. When we say the proposition ‘x is a triangle’ is true we imply that x is indeed a triangle (i.e. that some state of affairs is obtained). But just saying that ‘table is red’ implies the same thing. So, aren’t we asserting anything more when we assert truth of some sentence than just the compliance with some state of affairs? I think we do. By adding the word true to the sentence we *also* state that we *understand* the sentence. In Davidson’s words: “what Convention T […] reveals is that the truth of an utterance depends on just two things: what the words as spoken mean, and how the world is arranged.” (Davidson 1986, p.309) So Tarski’s extensional analysis of ‘s’ is true isn’t complete: it should be: ‘s’ is true ≡ p $\wedge$ s means p. If we take ‘s means p’ as granted (if for example our metalanguage contains the object language), it follows that ‘s’ is true ≡ p, which is Convention T. This shows that Tarski takes it for granted that ‘s means p’; but it might not be the case.

An easy experiment to see that one indeed needs two satisfied conditions for the truth of ‘s’, is this: consider someone saying ‘x is a triangle’. One immediately can verify that this sentence s (‘x is a triangle’ in our case) is indeed true by looking at our simplified world. So ‘s’ it true ≡ p, p is the case, so s is true. Now consider someone saying ‘x je trokut’ (i.e our new sentence s). To verify that this sentence is true, a monolingual speaker of English

---

[40] “In this construction [of a definition of truth] I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts.” (Tarski 1933, p.152-153)

needs two satisfied conditions to be met before he can accept that sentence s is indeed true. Firstly, the state p must be the case. Secondly, he also must know that s means p.[41]

As I already have mentioned, it would be a mistake to say that Tarski wasn't aware of this issue. None the less, he chose for a simpler approach: the Convention T. Now, Davidson concludes from this that Tarski took the semantic primitive of meaning as granted, while he takes the concept of truth as granted. The following citation, which Davidson wrote, presumably in 2001 for the introduction of his book, is pretty much to the point. "One thing that only gradually dawned on me was that while Tarski intended to analyze the concept of truth by appealing (in Convention T) to the concept of meaning (in the guise of sameness of meaning, or translation), I have the reverse in mind. I considered truth to be the central primitive concept, and hoped, by detailing truth's structure, to get at meaning." The question that rises is; how does knowing the concept of truth constitute meaning?

To show this, first let us start with the sentence a of L_1. To know the meaning of sentence a, as already shown, all one needs to know are truth-conditions of the sentence a. 'a' is true ≡ x is a triangle. Take another example: 'x je trokut' is true ≡ x is a triangle. Note that one does not need to know what 'je' and what 'trokut' mean in order to know the meaning of the sentence. So says Davidson: the truth-conditions constitute meaning.

Now, as for a more complicated case, imagine yourself being a baby-block y in our simplified Tarski-world. You don't know the language L_2 that your parents speak. Note that L_2 implies that your parents speak a language with an infinite amount of sentences. How are you ever going to understand it? The answer is when you know how the truth-conditions depend upon the use of the two truth-functional operators (i.e. when you know the truth table of the truth-functional operators), and when you know the truth-conditions for the atomic parts (i.e. our three propositions). Recall that truth is a semantic primitive in Davidson's theory of meaning. So knowing the truth-conditions of sentence a is the same as knowing under which circumstances the sentence a is true. Now, knowing under

---

[41] Note the difference between the two examples: in the first example metalanguage contains the object language; in the second example the metalanguage contains the translation of the object language.

which circumstances the sentence a is true, is knowing the meaning of it. Knowing the truth conditions of truth-functional operators is knowing what they mean. And, by knowing the definition of truth for a language, one can, for every sentence s in L_2, check what the truth-conditions are, and thus, know the meaning of s.

Example: we know what 'a' means if we know when 'a' is true. We know what 'x je trokut' means if we know when 'x je trokut' is true. If we know the truth-conditions of both sentences, and if they are the same, we can say that the meaning is also the same. This is not implausible. Think for example on the difference of the Morningstar and Eveningstar that Frege made. Their intension (i.e. meaning) is different, but their extension is not. Although truth is defined extensionally, it doesn't imply that their meaning is the same, as their truth conditions are not. The extensional truth conditions for Morningstar could be 'planet Venus' and 'celestial body visible just before sunrise', while the extensional truth conditions for Eveningstar could be 'planet Venus' and 'celestial body visible just after sunset on the western hemisphere'.[42] As their truth conditions are not the same, neither is their meaning.

Now we can go on to L_3. Here one needs to know for every predicate, under which circumstances it is true (i.e. one would need (non-logical) axioms corresponding with each predicate). And because you can construct any sentence from these predicates, truth-function operators and quantifiers, and because you have an unambiguous truth definition for L_3, you can extract the truth conditions (which constitute meaning) for any sentence.

Now, let us glance through the three constraints that Davidson has put on his theory, and see how our sketch of his theory of meaning complies with them. As for the first constraint, we have seen how an infinite amount of meaningful sentences is constructible from a finite number of entities which are learnable.

As for the third constraint, "the fact that meaning is to be cashed in terms of truth conditions makes it possible for a Davidsonian theory of meaning to be tested empirically." (Kirkham, p.228) For formal languages, the empirical test is verification. In

[42] The truth conditions used here are probably not complete, but are purely meant as an example.

case of natural languages, this would lead us to the issues of translation which I will not discuss here.

His second constraint, meaning for all utterances, is guaranteed by the truth conditions and the material adequacy condition. For formal languages the theory of meaning is the same as the theory of truth, apart from their interpretation of the *definiens*. "What appears to the right of the biconditional […] plays its role in determining the meaning of s not by pretending synonymy but by adding one more brush-stroke to the picture which, taken as a whole, tells what there is to know of the meaning of s." (Davidson, p.26) We can compare this to Tarski's view on truth of a particular sentence s. For a particular sentence s, the *definiens* in Convention T would contain one or more entities comprising the truth of s. This is because the truth in L "would be a 'logical conjunction' or 'logical product' of all of them [T-sentences]."[43] (Kirkham, p.145) So the truth of a particular sentence would be a part of truth of L. Note that this theory presupposes at least one semantic primitive, namely truth in Davidson's case and meaning or translation[44] in Tarski's case. Note also that for statement or first-order logic languages, a meaningful sentence is just a syntactically well formed sentence. Although there are many resemblances, there are also a few subtle differences. Tarski explicitly said that one can define truth (in his way) only for formal languages. Davidson on the other hand, maintains that his theory of meaning can be applied to natural languages also, so he has to show how the broad spectrum of word classes affects the meaning (i.e. truth conditions). He wants to do this the axiomatic way. Using a formal language, such as L_3, this would imply giving the satisfaction conditions for each predicate or name in that language.[45] For example, one could define the meaning of '… je trokut' in the following way:

---

[43] Kirkham is referring here to Tarski 1944, p.16 and Tarski 1933, p.187

[44] If the object language is pure syntax, then our meta-language will enclose it, and give it an interpretation (i.e. meaning). If we take this meta-language as our object language, then our meta-metalanguage will either enclose it, or have a full translation of it. (See also footnote 24.)

[45] I am using Kirkham's speculative interpretation of Davidson here, as Davidson himself, as Kirkham notes, did not give any concrete examples of such axioms. (Kirkham, p.227)

For all integers m and sequences S, x_m je trokut is satisfied by sequence S $\equiv$
the m-th object that S refers to is a triangle.

So, the meaning of 'x_m je trokut' would then be explicated by the proof of the truth of it.

Davidson's program: meaning in natural languages

As we have seen for formal languages, defining meaning of sentences by truth-conditions seems possible, because we have an unambiguous and explicit definition of truth. Problem is that we don't have any explicit truth definition for natural language. Tarski said that it is impossible to construct a truth definition for natural language, and if this is indeed the fact, then Davidson program is already doomed to fail. None the less, Davidson persists it is not, and that all problems eventually can be solved. So it brings us to the problems of trying to define truth for natural language. There is no place here for an exhaustive (or even extensive) account to on all such problems. My goal, as stated in the introduction, is only to give a sense of the problem.

Treating natural language as first-order logic language is not feasible, because, for example, natural language contains an infinite amount of predicates. For example, predicate 'the first', 'the second', the third', and so on.[46] Having an infinite amount of predicates means having an infinite definition of satisfaction[47] (and hence truth), and that is unwanted in Tarski's words or unlearnable in Davidson's. Another problem would be the sentences that predict something.

How about axioms? How should they look like? As a rulebook, a good starting point would be to use school grammar to divide words in classes, and define specific axioms by

---

[46] You might be able to solve this recursively. But there are similar others waiting. Take color for example. We perceive a particular color which is correlated to some wavelength(s). The electromagnetic spectrum where all our perceivable colors belong, is somewhere between 350nm and 750nm. There are an infinite amount of wavelengths between those two. A human can not distinct more than 2^{24} of them. How and by which standard is one going to specify those colors, knowing that not every 'green' for me is also 'green' for you? Although you 'might' solve this (which I think is highly unlikely), still there will be others waiting…

[47] Notice that in the definition of satisfaction (p.20) each predicate has a separate clause.

classes. Note that we need a limited set of rules, as insisted by Davidson's view of languages. As an example, take the following sentence e = 'John is a good man'. Understating it would imply knowing the truth conditions. It isn't problematic, as the truth conditions for it are simply to check whether there is any 'John' in the set of 'good men' where 'is' denotes the 'is an element of'-relation. We could constitute this set by taking the set of all people, and extract from it all men. We could then use adjective 'good' to extract from it all good men. Now the sentence f = 'John is a good actor', treated the same way, would create an undesirable result. Repeating our procedure, this would imply taking the set of all people, and extracting from it all actors. We would then use adjective good to extract from it all good actors. But then we have a set of all good actors in the sense that they are good (morally), and this is not the set we want[48], *although* our school grammar doesn't make any distinction between good in 'e' and good in 'f', as they are both adjectives. In other words, school grammar doesn't say we should treat those two same words differently. So we should extend school grammar in some way, but the question is in what way.

I already mentioned that we would need one semantic indefinable primitive in a theory of meaning (i.e. truth). Others might be primitive, but still axiomatizable (i.e. definable with satisfaction for instance). But take as example someone saying 'I am tired'. This sentence is true if it is uttered by a person that is indeed tired, and at the right moment when he *is* tired. So one would also need to presuppose time and consciousness (if that's not vague enough) as semantic indefinable primitives, and they both should have a separate clause in the Convention T. So the Convention T would become something like: 's is true at t (time) and spoken by i (person) ≡ …'. And what when the person i is laying?

Should we even treat the language logically? During 1940s, 50s and 60s much has been done on enriching the formal languages of symbolic logic to "express more and more of the concepts that play central roles in natural language ─ for example, modal concepts […], temporal concepts […], indexical expressions […] and propositional attitude verbs […]." (Soames, p.294) Now, what kind of logic shall we take? The newest?

[48] We want the set of all good actors which are better than average actors.

So, is there any progress? Davidson says there is. “He cites his own work on sentences with doxastic operators and sentences containing quoted expressions. He also cites the work of other philosophers on mass terms, 'ought' sentences, and the way that proper names refer.” (Kirkham, p.238) But, as Kirkham notes, there can be progress only if two steps are fulfilled. Analyzing the dubious entities *and* integrating them in into a solid theory of meaning (such as first-order logic for example). “Until we can integrate the results of a semantic research project into a Davidsonian theory of meaning for some language, we have no good reason for thinking that the results of the project really represent progress, as distinct from a blind alley […]” (Kirkham, p.238) Kirkham even gives an example of analysis of tensed verbs, one done by Quine, other by A.N. Prior, but which are incompatible. The question is “which of these represents progress from the standpoint of Davidson's program, if either? […] [So it seems that] citing supposed instances of progress is more an act of faith than argument. ” (Kirkham, p.239) On the contrary, Kirkham would probably find the following citation on linguistic applications in higher-order logic a way to make progress: “Linguistic applications of Skolem-functions have been primarily in the semantics of question-answering. A multiple question like "Which student got which grade?" is interpreted as a set of statements which constitute true answers to it, which must give a specific grade for each student. The answer to the wh-quantifier ‘which grade’ depends on the answer to ‘which student’.” (Partee, p.233)

Now, if you still aren’t convinced that such a theory of meaning is impossible, then you might have the Davidson disease. Anyhow, I am convinced, and although I do find his point about meaning interesting and very plausible, I think he should either drop the ‘natural language demand’ or drop his constraint of ‘being learnable’. The first constraint, as I tried to show, is not feasible. But one, I think, might still hold it, if one drops the ‘being learnable’ constraint. It is almost obvious to me that we do not know the truth conditions of every sentence we utter. We might suppose that there is some definition of truth, and that it might be known by some transcendent being, computer, nature or whatever your metaphysical convictions may call it. But *we* don’t know it. And a theory of meaning like the one Davidson proposes is one that just tries to explicate this impossibility.

What concerns the Davidson's program and Kirkham's critique on it that the second most important step (i.e. integrating linguistic results in into a solid theory of meaning) isn't fulfilled, I can only make the following remark. Why not build up on logic? It is at least a solid foundation wherein much study has been done that in turn can teach us something about language itself.[49] Although the structure of formal languages is far simpler than that of natural languages, we can extend them step by step, and look at the consequences. We started with declarative sentences. We then found a way to joint them together and form compound sentences, incorporating connectives such as 'not', 'and', 'or', etc. [50] This led us to statement logic. We then found a way how to extend it to first-order logic, hereby incorporating determiners such as 'some', 'all', 'none', etc. and 'properties of individuals'[51]. The next step could be higher-order logic as this logic allows 'properties of properties of … of individuals' to be incorporated. Of course, there are connectives such as 'because', 'while', 'after', etc. or determiners such as 'many' and 'few' which seem to be excluded. But at least, one has a strong foundation this way which can be extended, and it gives us also the means by which we can judge whether or not we make progress at all.

---

[49] To name a few: the completeness of proposition- and first-order logic, incompleteness of higher-order logic, the decidability of proposition logic, the enumerability of first-order logic and innumerability of higher-order logic. One can ask oneself whether one can attribute any of these properties to languages, and in what sense.

[50] Of course, the implementation might put some restriction on our use of natural languages. Our common use of implication might not, in all cases, correspond with logical implication. Question that arises is whether we should adjust our natural language to comply with logic, or search for different logical structures that can better comply with our common use of implication. To me, this seems a more profitable analysis, than the one Kirkham criticized as stagnant (see p.28)

[51] Predicates in logic do not need to correspond with grammatical predicates. For example, P(John, Mary) where P stands for 'likes', is as good as G(Mary) where G stands for 'John likes', where the latter is not a predicate in grammatical sense.

Epilogue

As for the conclusion, I would like go over the things we have seen. Firstly we have seen in what context and how Tarski defined truth. My main concern was to explicate the main components of his theory, and to give the reader a sense of what truth means in formal languages. Secondly, I tried to show that the notion of meaning was all but new in the formal logical tradition, and how Davidson tried to introduce it to a broader group of linguistic philosophers by putting some of basic conceptions of semantics in form of a program for constructing theories of meaning for natural languages.

References

Primary literature:

Davidson, Donald, 2001, ‘Truth and Meaning (1967)’ in *Inquiries into truth and interpretation.* Oxford, Clarendon, p.17-43

Kirkham, Richard, 1992, *Theories of truth: a critical introduction*. Cambridge, MIT press. p.401

Tarski, Alfred, 1933, ‘The concept of Truth in Formalized Languages’, in Tarski, 1983, *Logic, Semantics, Metamathematics*. 2nd ed. Translated by J. H. Woodger, Indianapolis, Hackett, p.152-209

Secondary literature:

Davidson, Donald, 2001, ‘Introduction’ and ‘Theories of Meaning and Learnable Languages (1965)’ in *Inquiries into truth and interpretation.* Oxford, Clarendon, p.i-xxiii and p.3-17

Horsten, Leon, 2007, ‘De structuur van wetenschappelijke theorieën’ in *Wetenschapsfilosofie*. Assen, Van Gorcum, p.21-51

Partee, Barbara Hall, 1990, ‘PART B: LOGIC AND FORMAL SYSTEMS’ in *Mathematical methods in linguistics.* Dordrecht, Kluwer, p.87-248

Soames, Scott, 2003, ‘Introduction’ and ‘Donald Davidson on truth and meaning’ in *Philosophical analysis in the twentieth century; The age of meaning*. Princeton (N.J.), Princeton university press, p.i-xxii and p.288-332

Tarski, Alfred, 1944, ‘The Semantic Conception of Truth and the Foundations of Semantics’, in *Philosophy and Phenomenological Research*, Vol. 4, No. 3, (Mar., 1944), p.341-376