



LUND UNIVERSITY

Tarski's one and only concept of truth

Smid, Jeroen

Published in:
Synthese

DOI:
[10.1007/s11229-014-0450-1](https://doi.org/10.1007/s11229-014-0450-1)

2014

[Link to publication](#)

Citation for published version (APA):

Smid, J. (2014). Tarski's one and only concept of truth. *Synthese*, 191(14), 3393-3406.
<https://doi.org/10.1007/s11229-014-0450-1>

Total number of authors:
1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Tarski's One and Only Concept of Truth

Jeroen Smid

Final version in *Synthese* (2014) 191:3393–3406

DOI: 10.1007/s11229-014-0450-1.

Link: <http://link.springer.com/article/10.1007/s11229-014-0450-1>

In a recent article, Marian David (2008) distinguishes between two interpretations of Tarski's work on truth. The standard interpretation has it that Tarski gave us a definition of truth in- L within the meta-language; the non-standard interpretation, that Tarski did not give us a definition of true sentence in L , but rather a definition of truth, and Tarski does so for L within the meta-language. The difference is crucial: for on the standard view, there are different concepts of truth, while in the alternative interpretation there is just one concept. In this paper we will have a brief look at the distinction between these two interpretations and at the arguments David gives for each view. We will evaluate one of David's arguments for the alternative view by looking at Tarski's 'On the Concept of Truth in Formalized Languages', and his use of the term 'extension' therein, which, we shall find, yields no conclusive evidence for either position. Then we will look at how Tarski treats 'satisfaction', an essential concept for his definition of 'true sentence'. It will be argued that, in light of how Tarski talks about 'satisfaction' in §4 of 'On the Concept of Truth in Formalized Languages' and his claims in the Postscript, the alternative view is more likely than the standard one.

Keywords: Tarski; Truth; Satisfaction; True-in- L

1. Introduction

Tarski's *On the Concept of Truth in Formalized Languages* (1935; hereafter, CTF) made a major contribution to twentieth-century philosophy. Despite its clarity, two fundamentally different interpretations of it are possible. On the standard interpretation, Tarski did *not* have *one* concept of truth, but rather various different concepts of 'true sentence in- L ', such as 'true sentence-in-German', 'true sentence-in-calculus-of-classes' etc. The non-standard, or alternative interpretation has it that there is *one* concept of truth, even though the extension of this (one) concept depends on a specific language. Marian David (2008) was the first, to my knowledge, to point out these different interpretations, yet does not indicate which one he thinks is correct (2008, 152-153). Here we will present the arguments in favour of both positions, and consider whether a case can be made in favour of the alternative interpretation – basing ourselves on a close-reading of CTF.

First, however, let us consider some quotes from other authors that make it evident that what David calls the 'standard interpretation' is indeed how some have interpreted Tarski's work. The standard interpretation has it that Tarski actually defines a concept expressed by 'true sentence *in* L '¹ rather than the concept expressed by 'true sentence'. It is important to keep in mind that L in 'true sentence in L ' is a schematic letter (not a variable) taking the place of a name or a description of a language. Hence, according to the standard interpretation, Tarski gave a definition of the concept expressed by 'true-in-the-language-of-the-calculus-of-

¹ Better: according to the standard interpretation Tarski provides the method to define *a* truth concept for a specified language, since L is not a variable, but a name for a language.

classes' and, by using the same methods, one could define *similar* concepts such as 'true-in-mereology', etc. So, the expression '*the* concept true sentence in *L*' is a bit misleading, since – according to the standard interpretation – there is no one such concept, but there are numerous truth concepts, which are all language-specific.² In order to keep this distinction clear, we will write 'true sentence-in-*L*' when we talk about the standard interpretation.

A first example of an adherent of the standard interpretation is Field:

[A Tarskian] truth definition works for a single language only, and so if it 'explains the meaning of' the word 'true' as applied to that language, then for *any* two languages L_1 and L_2 , the word 'true' means something different when applied to utterances of L_1 than it means when applied to utterances of L_2 ! I make this point not in criticism of [a Tarskian truth definition] but in criticism of the idea that the significance of [a Tarskian truth definition] can be explained by saying that it 'gives the meaning of' the word 'true' (Field 1972, 356).

From this quote it is evident that Field takes Tarski to have defined 'true sentence-*in-L*' rather than 'true sentence'. Sher (1999) quotes a portion of the above quote from Field, and, as can be seen from her comment directly following, agrees with him: "The idea of truth, however, is essentially the same for all languages; hence Tarski's theory fails to capture the intended idea" (Sher 1999, 158).

Similarly for Moreno:

[A]ccording to Tarski the concept of truth and the remaining semantic concepts are relative to particular languages (...). [I]t is understandable that Tarski did not define – and *he did not intend* to define either – the general concept of truth, (...) since the language in which such a definition would have to be formulated would contain its own truth predicate. (Moreno 2001, 140 – italics original)

Another example³ of an adherent of the standard interpretation is Kirkham, who writes:

We can, however, make explicit in the definition of truth the fact that the definition applies to only one language by explicitly naming, in the definition, the language, [...], for which truth is defined. [...] So there is not one theory of truth. Indeed, we really do not even have two theories of truth here. What we

² One should keep in mind that the difference between the standard and the alternative interpretation does not depend on whether Tarski held truth to be *domain-relative*. After giving his definition of 'true sentence' for the calculus of classes (CTF, 195), Tarski gives three definitions for truth relative to a domain (viz. 'true sentence for an individual domain *a*', 'true sentence for an individual domain with *k* elements', and 'true sentence for every individual domain' (CTF 200-201)). So, Tarski did provide domain-relative truth definitions, but these are different from his definition of 'true sentence'. The dichotomy between the alternative and the standard interpretation concerns solely the question whether the latter definition is language relative or not.

³ Thanks to an anonymous referee for drawing my attention to the following three examples of adherents of the standard interpretation.

have is one theory of ‘true-in- L_1 ’ and one theory of ‘true-in- L_2 ’. In the end, Tarski has defined not truth but truth-in-this-language and truth-in-that-language. (Kirkham 1992, 162)

Also, Burgess and Burgess (2011) write:

The latter [i.e. taking ‘true’ to be elliptical for “true in L ”] is Tarski’s way. For him quotation marks are used to designate orthographic types, and what he is concerned with under the label “truth” is truth-in-English, or truth-in- L for some other language L . (Burgess and Burgess 2011, 18)

A final example⁴ is provided by Künne who writes “what Tarski wants to define are predicates of the type ‘ s is a true sentence of language L ’(‘ s is true in L ’ for short).” (2003, 181) The use of ‘predicates’ (plural) is illustrative: according to the standard interpretation, there are different concepts of truth.⁵

The above quotes⁶ provide evidence that some interpret Tarski’s CTF as giving us the tools to define various concepts of true sentence, such as the concept expressed by ‘true sentence-in-predicate-logic’, the concept expressed by ‘true sentence-in-German’, etc. This is what David calls the interpretation that takes Tarski to define ‘true sentence-in- L ’⁷, and will be called ‘standard’ – following David (although it might be a bit of a misnomer).

An adherent of the *alternative* interpretation is Patterson, who thinks Tarski had one – absolute – concept of truth in mind, rather than numerous language-specific concepts. His reasons for thinking so lie in the distinction between the extension of the concept (which does depend on the language in question) and the content (which is absolute), together with an analogy with the concept of ‘sentence’. (Patterson 2012, 67-72) These points will also play an important role in this paper.⁸

⁴ Another *possible* example is Quine who writes: “Attribution of truth in particular to ‘Snow is white’, for example, is every bit as clear to us as attribution of whiteness to snow. In Tarski’s technical construction, moreover, we have an explicit general routine for defining truth-in- L for individual languages L which conform to a certain standard pattern and are well specified in point of vocabulary. We have indeed no similar definition of ‘true-in- L ’ for variable ‘ L ’; but what we do have suffices to endow ‘true-in- L ’, even for variable ‘ L ’, with a high enough degree of intelligibility so that we are not likely to be averse to using the idiom.” (Quine 1953, 138) However, one could also hold that since Quine claims that there is a “general routine” that he might hold that Tarski had *one concept* of truth in mind which needs to be defined *for a particular language* (i.e. that Quine is an adherent of the alternative interpretation).

⁵ If, however, Künne is of the opinion that the different predicates all express a single concept of truth, Künne should be seen as an adherent of the alternative interpretation.

⁶ Another example: “[T]arski expressly aimed to define truth (or, rather, “true-in- L ”) without assuming any semantic notions.” (Raatikainen 2008, 112)

⁷ Better: according to the standard interpretation Tarski provides the method to define *a* truth concept for a specified language, since L is not a variable, but a name for a language.

⁸ Other possible adherents of the alternative interpretation are Feferman (2008, 83-84) and Simons (2009), who both hold that Tarski holds that truth is an *absolute* notion. But it should be noted – as a referee pointed out to me – that Feferman and Simons do not seem to address the dichotomy between the standard and the alternative interpretation, but rather a dichotomy between on the one hand the

What conclusions follow or could be seen to follow on the basis of either interpretation will not be under discussion. Here we shall be concerned only to answer the question of whether Tarski defined various concepts expressed by ‘true sentence-in-*L*’ (i.e. the standard interpretation, in which *L* is a name for a language, and there is no one concept of ‘true sentence’) or the concept expressed by ‘true sentence’ (alternative interpretation, on which there is *one* concept of ‘true sentence’, and it is defined *for* various languages). We will try to answer this question solely by looking at what Tarski writes in CTF, not taking into account historical or testimonial evidence.⁹

Neither will we be occupied with tracing the (historical) roots of the dispute between the standard and alternative view, which, according to Simons, is due to the fact that Tarski’s CTF is now usually read through the lens of his latter model theoretical work. (Simons 2009, 4). Related to this, Feferman holds that it depends on emphasis: if one emphasizes Tarski *qua* (meta-) mathematician, the standard view becomes likely; if one emphasizes Tarski *qua* philosopher, one would have to adhere to the alternative view. (Feferman 2008, 86)

Now that we have seen examples of both interpretations, let us look at David’s arguments for either position. He explains that, throughout CTF, Tarski uses ‘true sentence’ instead of ‘true sentence *in L*’, and ‘*the* concept of truth’ instead of ‘the concepts of truth’ (David 2008, 150). This is in line with the alternative interpretation. Moreover, the title of Tarski’s paper suggests that there is just *one* concept applicable to formalized *languages*. (Note the plural: in the standard view, there is not *one* concept of truth for different *languages*.) (Ibid.)

However, the alternative view has to deny the truth of the principle: *if different extensions, then different concepts* – which is, admittedly, a very likely principle, and one which the standard view embraces and uses as an argument in favour of its interpretation. (Ibid.)

David explains that the above principle is not completely general: “where *context sensitivity* comes into play the transition appears to fail.” (Ibid.) The alternative view could very well claim that the language under consideration (i.e. the language for which the concept expressed by ‘true sentence’ is defined) determines the context such that the extension of the term ‘true sentence’ is different in different languages, though the concept remains the same in all of them. (Ibid.) An example, given by David, is ‘today’: the extension of this concept is context-sensitive, although the meaning of the concept is fixed. According to the alternative interpretation, the context that determines the extension of the concept expressed by ‘true sentence’ is

alternative interpretation – in which ‘true sentence’ expresses one *absolute* concept and a *third* interpretation, on the other hand, according to which ‘true sentence’ is one *relative* concept. We will not take this third interpretation into account, although the arguments in this paper could possibly also be put forward against it.

⁹ Although there is some such evidence: Jan Wolenski has reported that Jan Tarski, Alfred’s son, told him “his father considered the absoluteness of truth as truth’s important feature.” (Murawski & Wolenski 2008, 33 note 21) Furthermore, Carnap writes in his intellectual autobiography: “When Tarski told me for the first time he had constructed a definition of truth, I assumed that he had in mind a syntactical definition of logical truth or provability. I was surprised when he said that he meant truth in the customary sense, including contingent factual truth.” (Carnap 1963, 60) Together with the assumption that ‘truth in the customary sense’ is absolute, this latter quote suggests that Tarski was, at least himself, of the opinion that there was only one concept of truth.

language: language would make the extensions of ‘true sentence’ differ, although the concept referred to would always be the same, i.e. it would be one (Ibid.).

By looking at quotes from CTF we will first consider whether or not Tarski held the principle ‘if different extension, then different concepts’. For ease of reference, we will give the principle a name and an explicit characterization:

Extension Principle: If the extension of the concept $x \neq$ the extension of the concept y , then $x \neq y$

Our question, then, shall be whether Tarski thought that the *Extension Principle* (hereafter: EP) applies to the concept expressed by ‘true sentence’, or whether he thought that the concept expressed by ‘true sentence’ is context relative such that it violates EP. The former case would mean that there is not *one* concept expressed by ‘true sentence’, but rather various concepts such as ‘true sentence-in-*the-calculus-of-classes*’, and ‘true sentence-in-*Polish*’, etc.

2. Extended analysis of Extension

The word ‘extension’ occurs twenty-seven times in CTF. Of these occurrences, only the ones most relevant for us here will be considered. Criteria for relevance are due to EP. Most quotes will be given additional remarks to help make clear the context. The quotes are given in page-order. (All italics are original and exclamation marks [!] are added to draw attention to certain words.)

(T1) The extension [!] of the concept to be defined depends [!] in an essential way on the particular language under consideration. The same expression can, in one language, be a true statement, in another a false one or a meaningless expression. There will be no question at all here of giving a single general definition of the term. (CTF, 153)

This quote follows immediately after Tarski has declared his intention to consider only the “*classical* conception of truth (‘true – corresponding to reality’) in contrast with, for example, the *utilitarian* conception (‘true – in a certain respect useful’)” (CTF, 153). The above quote does not tell us whether Tarski thinks that a concept of the form ‘true sentence’ obeys EP. It does tell us that the extension of ‘true sentence’ is not always the same. (T1) is interesting because, as David notes, on the standard view “Tarski is speaking rather misleadingly [...]: there is no such thing as *the* concept of truth to be defined: instead there are different concepts with different extensions” (2008, 151). On the alternative view, such talk *is* legitimate.

(T2) It might appear [...] that ‘true sentence’ with respect to the language of a formalized deductive science means nothing other than ‘provable theorem’ [...] [T]his view [however] must be rejected for the following reason: no definition of true sentence which is in agreement with the ordinary usage of a language should have any consequences which contradict the principle of excluded middle. This principle, however, is not valid in the domain of provable sentences. [...] The extension of the two concepts is thus not identical. (CTF, 186)

Note that (T2) appeals to EP: a concept expressed by ‘true sentence’ has a different extension than a concept expressed by ‘provable sentence’, and so is not identical.

This is not, however, conclusive evidence in favour of the standard interpretation, since the defender of the alternative view could hold that *within* a language EP does of course hold¹⁰. The defender of the alternative view could hold that EP fails across multiple languages, which is in accordance with Tarski's claim that "[t]he same expression can in one language, be a true statement, in another a false one or a meaningless expression". (CTF 153)

Most of the other quotes from Tarski's text that contain 'extension' do not give conclusive evidence in favour of or against EP precisely because the move just sketched is available to the defender of the alternative view. There is however one quote that seems to hint that Tarski does adhere to EP with respect to 'true sentence':

(T3) [a] As I have already emphasized in the Introduction, the concept of truth essentially depends, as regards both extension and content, upon the language to which it is applied. [b] We can only meaningfully say of an expression that it is true or not if we treat this expression as part of a concrete language. [c] As soon as the discussion is about a large number of languages the expression 'true sentence' ceases to be unambiguous. [d] If we are to avoid this ambiguity we must replace it by the relative term 'a true sentence with respect to a given language'. (CTF, 263)

Let's take a closer look at (T3), the sentences of which I have enumerated for ease of reference. This quote exposes clearly the way in which the two different interpretations diverge, and what their divergence means for EP. (David also discusses this passage at length (2008, 151-153).) Note, first, that Tarski starts by using *concept* (the concept of truth - [a]), then talks about *expression* (the expression 'true sentence' - [c]) and ends up with *term* (the term 'a true sentence with respect to a given language' - [d]). Yet it makes sense to read the word 'concept' in all three of these instances; the meaning of (T3) seems unaltered and still open for both the alternative and standard interpretation. (Though not in the case of [b], for there the two appearances of 'expression' cannot be read as 'concept'.) This move is *not* relevant for my argument, but only a heuristic move undertaken to make the difference between the alternative and standard interpretations as clear as possible.

It is useful to consider the two possible interpretations of each sentence of (T3), and to do so while substituting the word 'concept' for 'expression' in [c] and 'term' in [d]. We will do this one by one, where '*Sta*' refers to the interpretation of the sentence that is in accordance with the standard reading and '*Alt*' to the interpretation that is in accordance with the alternative one; and '[a]' refers to sentence (T3)[a], '[b]' to sentence (T3)[b], etc.

[a]-*Sta*: there are various concepts expressed by 'true sentence' such as 'true sentence-in-calculus of classes', 'true sentence-in-Polish', etc., because the extension and content of 'true sentence' depends on the language.

[a]-*Alt*: the extension and the content of the concept expressed by 'true sentence', but not the concept itself, depend on language.

[b]-*Sta*: talk about the truth of a sentence is language-relative.

[b]-*Alt*: talk about the truth of a sentence is language-relative.

¹⁰ As long as there are no context-sensitive expressions in the language, otherwise EP would of course fail in those cases.

- [c]-*Sta*: when many languages are taken into consideration, there is *not one* concept of true sentence applicable to all of those languages at once.
- [c]-*Alt*: when many languages are taken into consideration, the concept of true sentence cannot be *defined* for all of those languages at once.
- [d]-*Sta*: the concept of true sentence with respect to a given language has to be defined instead.
- [d]-*Alt*: the concept of true sentence with respect to a given language has to be defined instead.

The readings of sentences [b] and [d] are the same in the alternative as in the standard reading. It is important to note, however, that the [d] pair is ambiguous and hence provides a clear example of the reason why there are two interpretations of CTF in the first place. As David notes, when commenting on (T3)[d], the use of the term ‘relative’ by Tarski “might suggest the standard ‘true in *L*’ interpretation.” (David 2008, 152). However, as David rightly points out:

[The] continuation of the passage shows that Tarski is thinking here of constructing a single metalanguage common to the object-languages under consideration. He seems to be saying that, with such a metalanguage in hand and provided the object-languages are well-behaved ones, we should be able to define a genuine relation term, ‘*x* is true in language *y*’, albeit one whose range of application, i.e. the range of the variable ‘*y*’, will be restricted to the object-languages under consideration (cf. CTF 263-4). This does not really fit well the ‘true in *L*’ interpretation on which ‘true in *L*’ is not a relational term but merely a stand-in for various one-place predicates. (David 2008, 152)

Given that Tarski uses the expression ‘true sentence with respect to a language’ quite often, the two different interpretations might not be so surprising. The clearest examples of the different interpretations are provided by the [a] and [c] pairs, since here the question presents itself directly: is there just one concept expressed by ‘true sentence’, or not?

Another question is: what does Tarski mean by ‘ambiguous’? David notes that “[i]f [Tarski] means that different occurrences of the term expresses different concepts, then the remark point towards the standard interpretation. If he merely means that different occurrences of the term have different extensions (or intensions), then the remark is compatible with the non-standard interpretation” (2008, 152). Looking at Tarski’s use of the term ‘ambiguous’ in CTF, one finds that no conclusive evidence for either interpretation can be given.

In light of this section we may conclude, on the basis of (T2), that Tarski upheld the validity of EP at least *within* a language. We also saw, in (T1), that Tarski sometimes speaks as if there is *one* concept of truth, favouring the alternative interpretation. In the last quote, (T3), we discovered that both interpretations can make perfect sense of some claims made by Tarski, which might lead to the unfortunate conclusion that we may never know the answer to our question.

3. Satisfaction

Yet there are good reasons to support the alternative view. For when one does a close reading of §4 of CTF, and takes the Postscript into consideration, the idea that Tarski had *one* concept of truth in mind strongly suggests itself. It makes sense to do

this, since in §4 some difficulties arise (and are dealt with) related to the definition of ‘satisfaction’.

Satisfaction is the key notion in Tarski’s definition of ‘true sentence’. In §3, Tarski states, after defining ‘satisfaction’:

(T4) The concept just defined [i.e. satisfaction] is of the greatest importance for investigations into the semantics of language. With its help the meaning of a whole series of concepts in this field can easily be defined, e.g. the concepts of denotation, definability, and truth (CTF, 194).

In §4 of CTF, Tarski begins by describing the general method for constructing a definition of ‘true sentence’ (CTF, 209). For our purposes, it shall be interesting to look at how Tarski does this. Tarski explains that, in order to define the concept expressed ‘true sentence’ (the standard interpretation would say “to define ‘true sentence-in- L ’¹¹”), one has first to construct a metalanguage that meets specific requirements (CTF, 210). The requirements are such that the metalanguage is capable both of *naming* every expression of the language, and also of *translating* every expression of the language in question (CTF, 210-211). Tarski then states that important expressions of the language in question need to be distinguished, such as those falling under the concept expressed by ‘sentential function’ and, especially, those falling under the concept expressed by ‘sentence’. He explains how these concepts need to be defined (CTF, 212). Tarski still talks very generally here, which can also be seen in the following quote:

(T5) The expressions of the language investigated consist of *constants* and *variables*. Among the constants [...] we find, as a rule, certain signs belonging to the sentential calculus and the predicate calculus [...] In addition to these [constants], we sometimes [!] find other signs which are connected with the individual [!] peculiarities of the language and denote concrete individuals, classes, or relations (CTF, 212).

Tarski does make clear that some language might be slightly different (‘individual peculiarities of the language’), but has a very general picture in mind. He continues by explaining such notions as ‘primitive sentential function’, ‘(sentence forming) functor of the given primitive sentential function’, ‘arguments’ and ‘fundamental operations on the expressions’ (CTF, 212-13). Throughout, the procedure is taken to be applicable to every formal language. After defining sentences as “sentential functions without free variables” (CTF, 214), Tarski explains how ‘axiom’, ‘consequence’ and ‘theorem’ can be defined (Ibid.). What is interesting here is that Tarski seems to have *one* concept of axiom in mind, although the definition of this concept depends, again, completely on the science (or language) under investigation. One may thus argue that the concept of axiom does not obey EP: the extension of this concept differs depending on the language to which the concept is applied, though the concept remains always the same. By the standard interpretation, however, it seems that one must claim that Tarski actually gave a procedure to define the concept falling under the schema expressed by ‘axiom in- L ’, rather than *one* concept of ‘axiom’. Given the generality of Tarski’s presentation of his procedure in §4, there seems to be no obvious reason to draw a distinction here between a concept of axiom and that of

¹¹ To repeat, ‘ L ’ is a name or description of a language, such that (according to the standard interpretation) there is a concept ‘true sentence-in-calculus-of-classes’ and a – different! – concept ‘true sentence-in-topology’, and so one for various languages.

true sentence with respect to language. (Cf. Patterson 2012, 67) I regard this as a drawback of the standard view, since no one seems to want to suggest that Tarski thought of the concept expressed by ‘axiom’, ‘consequence’, and the like, are actually various concepts clustered in the phrase ‘axiom in-*L*’ and ‘consequence in-*L*’. Here, however, we shall focus on ‘satisfaction’, which is also, arguably, not seen by Tarski as being various *different* concepts.

Directly after defining ‘theorem’, Tarski writes:

(T6) After this preliminary work we turn now to our principal task – the construction of a correct definition of *true sentence*. As we saw in §3, the method of construction available to us presupposes first a definition of another concept of a more general kind which is of fundamental importance for investigations in the semantics of language. I mean the notion of the *satisfaction of a sentential function by a sequence of objects* (CTF, 214).

Again, Tarski does not mention or seem to imply that we are actually concerned with different concepts expressed by ‘satisfaction of a sentential function by a sequence of objects in-calculus-of-classes’, ‘satisfaction of a sentential function by a sequence of objects in-mereology’, etc. He rather seems to think that there is just one concept expressed by ‘satisfaction of a sentential function by a sequence of objects’, which also seems to be the case when he explains how to give a (recursive) definition of satisfaction:

(T7) [i] For this purpose it suffices [...] to establish two facts: (1) which sequences satisfy the fundamental functions, and (2) how the concept of satisfaction behaves under the application of any of the fundamental operations [...] [ii] As soon as we have succeeded in making precise the sense of this concept of satisfaction, the definition of truth presents no further difficulty: the true sentences may be defined as those sentences which are satisfied by an arbitrary sequence of objects. [iii] In carrying out the plan just sketched in connexion with various concrete languages we nevertheless meet with obstacles of a fundamental kind; in fact, just at the point where we try finally to formulate the correct definition of the concept of satisfaction (CTF, 214-5).

This quote may be taken as evidence for the claim that Tarski thinks there is *one* concept of satisfaction. If this is true, then Tarski does not define various concepts such as ‘satisfies in-calculus-of-classes’, ‘satisfies in-Russian’, etc., but rather, simply, ‘satisfies’, and does so *for* a language. The second part of (T7)[i] suggests that there is only *one* concept of satisfaction. Moreover, in (T7)[iii] Tarski links the method he has described to ‘various concrete languages’, and it is precisely here that problems arise. Let us consider what exactly the problem is with defining satisfaction for various languages.

The problem lies in the fact that the terms of a language can belong to different semantical categories. Tarski accepts the so-called “theory of semantical categories” (CTF, 215), and characterizes ‘belonging to the same semantical category’ as a reflexive, symmetrical and transitive relation holding between two expressions, so long as “(1) there is a sentential function which contains one of these expressions, and [...] (2) no sentential function which contains one of these expressions ceases to be a sentential function if this expression is replaced in it by the other” (CTF, 216). Every category can receive a number that is called the ‘order of the category’. This is done by convention:

- (T8) (1) the 1st order is assigned only to names of individuals and to the variables representing them; (2) among expressions of the $n+1$ th order, where n is any natural number, we include the functors of all those primitive functions all of whose arguments are of at most the n th order, where at least one of them must be of exactly the n th order (CTF, 218).

With these notions, Tarski is able to classify the formal languages into four kinds:

- (T9) (1) languages in which all the variables belong to one and the same semantical category; (2) languages in which the number of categories in which the variables are included is greater than 1 but finite; (3) languages in which the variables belong to infinitely many different categories but the order of these variables does not exceed a previously given natural number n ; and finally (4) languages which contain variables of arbitrarily high order. We shall call the languages of the first three kinds *languages of finite order*, in contrast to languages of the fourth kind, the *languages of infinite order* (CTF, 220).

The classification of the languages on the basis of the order of their semantical categories makes it possible for Tarski to keep to a rather general approach in §4 of CTF. All formal languages belong to a certain kind, and each kind receives its own discussion with respect to how ‘satisfaction’ is defined in its context. By taking into account every kind, one by one (the fourth kind in §5), Tarski is able to maintain rather well the generality of his procedure that he started at §4. Tarski reminds us that the calculus of classes (§3) is an example of a language of the first kind; there are no difficulties in that kind (CTF, 221). Difficulties arise only for the other three kinds (CTF, 222). There are, however, methods that make it possible to define ‘satisfaction’ (and, so, ‘true sentence’) for languages of the second and third kinds. These methods are two: the method of many-rowed sequences, and the method of semantical unification (CTF, 227ff.). Their technical aspects need not detain us here. It is interesting, however, to see *why* Tarski starts using these methods. The reason lies in the semantical ambiguity of the concept of satisfaction: different sentential functions belonging to different semantical categories can enter into relation with this concept, and there is thus in these cases, strictly speaking, not just one relation of satisfaction, but many (CTF, 225). Tarski first reminds us of what we saw in §3:

- (T10) In order to avoid this ambiguity, in dealing with the calculus of classes [in §3] we had recourse to an artifice which is used by logicians and mathematicians in similar situations [...] [W]e tried to operate with the semantically uniform, if somewhat artificial, concept of satisfaction of a function by a sequence of objects. It happened that this concept is sufficiently more general than the previous one to include it – intuitively speaking – as a special case (CTF, 225-26).

The ‘artificial’ concept of satisfaction made it possible to have one concept of satisfaction rather than many. This was necessary, since the domain of the original relation of satisfaction was ambiguous, i.e. it had different domains (CTF, 224). But the construction of such an artificial concept of satisfaction cannot be accomplished directly for languages of the second kind, because within any of these languages there are variables of at least two semantical categories. Yet not all is lost:

(T11) Nevertheless the methods used in §3 can be applied to the language now being investigated, although with certain modifications. In this case also it is possible to find an interpretation of the concept of satisfaction in which this notion loses its semantical ambiguity and at the same time becomes so general that it includes all special cases of the original concept. (CTF, 227)

Tarski then presents the two methods, then proceeds to explain how each applies to a language of the second kind. Languages of the third kind can use only the method of unification (CTF, 231).¹² After providing a brief sketch (CTF, 231-35) of how this method applies to the logic of many-termed relations (*an* example of a language of the third kind), Tarski summarizes his results:

(T12) *In this way the various methods at our disposal enable us to define the concept of satisfaction and with it to construct a correct definition of truth for any language of finite order.* (CTF, 235; italics in original)

This quote seems to suggest that Tarski thinks there is *one* concept of satisfaction, and that by means of this one concept a *single* definition of ‘true sentence’ can be constructed. The claim that Tarski seems to think that there is one concept (or, that it is possible to conceive of satisfaction as *one* concept applicable to different languages) seems to be justified by the fact that Tarski, in (T11), talks about *one*, unambiguous, concept; yet one that is nevertheless so general as to include ‘all the special cases of the original concept’. (CTF, 227) Tarski resorts to the method of many-rowed sequences and the method of semantical unification precisely because he wants to have *one* concept rather than many.

Another reason to claim that Tarski thinks there is *one* concept of satisfaction might be the definite article ‘the’ just preceding the word ‘concept’ in (T12); this rather than ‘a’ or ‘some’. That the concept of satisfaction is necessary for the definition of truth is evident from (T4), (T6), (CTF, 195) (where Tarski first defines ‘true sentence’ and employs the notion of satisfaction) and (CTF, 214) (where Tarski repeats his claim of (T4)).

Now, if all this is true, then a good case can be made in favour of the alternative interpretation: for the claim, that is, that there is *one* general concept of satisfaction, applicable to many languages, and that this *one* concept is necessary to define the concept expressed by ‘true sentence’. The only other concept, next to satisfaction, necessary to define ‘true sentence’ is ‘sentence’, and one could argue – in defence of the standard interpretation – that this not *one* concept, but rather various concepts that fall under the schema ‘sentence in-*L*’. I would like to remind the reader, however, of the fact that in §4 Tarski explains how to define ‘sentence’ for *any* language. So, again, there seems to be just *one* concept, applicable to many languages (rather than many concepts falling under ‘sentence in-*L*’). No other concepts are necessary to define ‘true sentence’, and nowhere in the definition of ‘true sentence’ does the language *for which* the concept is defined seem to be essential. This suggests that Tarski did not define various concepts such as ‘true sentence-in-calculus-of-classes’, ‘true sentence-in-German,’ etc. but rather, simply, *one* concept expressed by ‘true sentence’.

¹² More precisely: languages of the third kind first have to use the method of unification to ‘lower it down’ (so to speak) to a language of the second kind, then both the method of unification and the method of many-rowed sequences can be applied again. (CTF, 231-235)

And there is more. In the next section we will briefly consider Tarski's Postscript, which provides additional evidence in favour of the idea that Tarski thought that there was *one* concept of satisfaction and therefore *one* concept of true sentence.

4. No semantical categories

The negative results obtained in §5 of CTF – i.e. that for languages of the fourth kind no definition of true sentence can be given, because 'satisfaction' cannot be defined for languages of the fourth kind (CTF, 245-46) – rely on the premise that there are semantical categories. In the Postscript of CTF, added in the German and English translations of the Polish original, Tarski writes that he no longer holds the theory of semantical categories. (CTF, 268) By abandoning this theory, the negative results of §5 disappear. Thus Tarski claims that:

(T13) *For every [i.e., finite or infinite] formal language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language – but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation (CTF, 273).*

The specifics that make it possible, according to Tarski, to define 'true sentence' for languages of infinite order (contrary to his earlier claim), need not detain us here. Even Tarski himself merely hints at how to do this. What *is* interesting to note, however, is that Tarski speaks very generally after claiming that, for languages of infinite order, a metalanguage of an higher order can be constructed (CTF, 271-72); just as generally as he did at the beginning of §4:

(T14) But now we are in a position to define the concept of truth for any language of finite or infinite order[...] It is perhaps interesting to emphasize that the construction of the definition is then to a certain degree simplified. We can adhere strictly to *the method* outlined in §3 without applying the artifices which were compelled to use in §4 in the study of languages of the 2nd and 3rd kind (CTF, 272, emphasis added).

Tarski reminds us that he no longer needs the methods of many-rowed sequences or of semantical unification ('the artifices'), but that instead he can directly use the method of §3. This method was described in §4, in all generality, up to the point where satisfaction had to be defined. Unfortunately, Tarski does not explain how 'satisfaction' has to be defined when the theory of semantical categories is abandoned; the word 'satisfaction' is even absent from the Postscript. Yet (T14) does give us reason to think that there is, for Tarski, just *one* concept of true sentence, which can be defined *for* any (finite or infinite) formal language. Only because Tarski held the theory of semantical categories was there a reason for him to consider different (kinds of) formal languages with respect to his method for constructing a definition of truth. Neither the concept expressed by 'true sentence' itself, nor that of satisfaction, had to be defined differently for different (kinds of) languages, because they would actually be – as the standard interpretation claims – different language-specific concepts; only adherence to the theory of semantical categories limited the possibilities of his method. Abandoning this theory made it possible for Tarski to achieve the degree of generality he wanted at the beginning of CTF, when he stated:

[The present article's] task is to construct – with reference to a given language – a *materially adequate and formally correct definition of the term 'true sentence'*. (CTF, 152)

5. Conclusion

The alternative interpretation denies that EP holds for the concept expressed by 'true sentence'. In CTF, there is no evidence in favour of or against EP with respect to this concept. But looking at what Tarski does in §4 for the concept of satisfaction, and taking the Postscript of CTF into account, does, however, suggest that Tarski thought he had given a procedure to define the concept of true sentence as a *single* concept, rather than a procedure to define various different concepts such as 'true sentence-in-calculus-of-classes', 'true sentence-in-French', etc. This suggests that the alternative interpretation is correct.

If the alternative interpretation is correct, then some arguments against Tarski's truth definition need to be revised. To me, there seems to be a modus ponens here.

Acknowledgments

Many thanks to Arianna Betti, Iris Loeb, and three anonymous reviewers for commenting on previous versions of this paper. Also, many thanks to Jonathan Sozek for proofreading my manuscript. Whatever mistakes remain are entirely my own.

References

- Burgess, A.G. & Burgess, J.P., (2011). *Truth*. Princeton: Princeton University Press.
- Carnap, R., (1963). Intellectual Autobiography, in *The Philosophy of Rudolf Carnap* (P.A. Schilpp, ed.) *The Library of Living Philosophers*, vol. 11 Open Court (La Salle).
- David, M., (2008). Tarski's Convention T and the Concept of Truth, in Patterson (ed.) *New Essays on Tarski and Philosophy*, Oxford: Oxford University Press: 133-156.
- Feferman, S., (2008). Tarski's Conceptual Analysis of Semantical Notions, in Patterson (ed.) *New Essays on Tarski and Philosophy*, Oxford: Oxford University Press: 72-93.
- Field, H., (1972). Tarski's Theory of Truth, *The Journal of Philosophy* 69(13): 347-375.
- Kirkham, R.L., (1995). *Theories of Truth – A Critical Introduction*. Cambridge (MA): The MIT Press.
- Künne, W., (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Moreno, L.F., (2001). Tarskian Truth and the Correspondence Theory, *Synthese* 126: 123-147.
- Murawski, R. & Wolenski, J., (2008). Tarski and his Polish Predecessors on Truth, in Patterson (ed.) *New Essays on Tarski and Philosophy*, Oxford: Oxford University Press: 21-43.
- Patterson, D., (2012). *Alfred Tarski: Philosophy of Language and Logic*, Palgrave Macmillan.
- Quine, W.V.O., (1953). *From a Logical Point of View*. Cambridge (MA): Harvard University Press.
- Raatikainen, P., (2008). Truth, Correspondence, Models, and Tarski, in Philström et al. (eds.) *Approaching Truth: Essays in Honour of Ilkka Niiniluoto*, London: College Press: 107- 119.
- Sher, G., (1999). What is Tarski's Theory of Truth?, *Topoi* 18: 149-166.

- Simons, P., (2009). Twardowski on Truth, *The Baltic International Yearbook of Cognition, Logic and Communication*, 4: 1-14.
- Tarski, A., (1935)[1956]. The Concept of Truth in Formalized Languages, *Logic, Semantics, Metamathematics; Papers from 1923 to 1938*. Oxford University Press: 152-278.