The Concept of Truth in Formalized Languages
by Alfred Tarski

1 The Concept of True Sentence in Everyday or Colloquial Language

In this section, Tarski demonstrates some of the difficulties in trying to define a semantic notion of truth in natural languages. Ideally, one would like a definition expressing the following:

> a true sentence is one which says that the state of affairs is so and so, and that the state of affairs indeed is so and so ([Tar35] 155)

In attempting to construct such a definition, however, it is difficult to avoid inconsistency. The liar's paradox seems to be an unavoidable problem, as shown through many examples (pages 157-162), leading Tarski believes it is impossible to obtain a consistent definition of semantic truth in natural languages.

> the attempt to construct a correct semantical definition of the expression 'true sentence' meets with very real difficulties ([Tar35] 162).

What then about a *structural* definition of truth? i.e.

> a true sentence is a sentence which possesses such and such structural properties or which can be obtained from such and such structurally described expressions by means of such and such structural transformations ([Tar35] 163).

Tarski provides the following examples of such a structural definition.

1. every expression consisting of four parts of which the first is the word 'if', the third is the word 'then', and the second and fourth are the same sentence, is a true sentence.

2. if a true sentence consists of four parts, of which the first is the word 'if', the second a true sentence, the third the word 'then', then the fourth part is a true sentence.

Such an approach to defining truth, however, would require many (if not an infinite) number of rules. And as Tarski notes, natural languages are not "finished", but rather constantly changing (with the addition of new words and expressions, the removal of existing ones, and chaning notions for existing ones). Also, the very structure of a natural language can change, implying that any structural definition would also have to be dynamic. Thus,

> The attempt to set up a structural definition of the term 'true sentence' – applicable to colloquial language is confronted with insurperable difficulties ([Tar35] 164).

One of the most important features of natural languages is that they are *universal* in the sense that anything that can be spoken of can be spoken of in a natural language (in contrast with formal languages which are quite limited in what they can speak of). To maintain this universality when developing a semantic notion of truth for a given natural language, one must allow any expression in that language mentioning 'true sentence', 'name', or 'denote'. But as Tarski notes, this seems to *force* any such language to be inconsistent.

Tarski concludes that a semantic or even structural notion of truth is completely impossible in natural languages and abandons the attempt all-together, concerning himself only with formalized languages for the remainder of the paper.

2 Formalized Languages, especially the Language of the Calculus of Classes

Properties of all formal languages, as described in [Tar35] page 166...

1. A list or description is given in structural terms for all of the symbols with thich the expressions of the languages are formed.
2. Among all possible expression which can be formed with these signs, those called *sentences* are distinguished by means of purely structural properties.
3. A list, or structural description, is given of the sentences called *axioms* or *primitive statements*.
4. In special rules called *rules of inference*, certain operations of a structural kind are embodied which permit the transformation of sentences into other sentences, called *consequences*. In particular, the consequences of the axioms are called *provable* or *asserted sentences*.

Unlike natural languages, formal languages are not universal, specifcially in that they do no posses (directly) terms denoting symbols or expressions of the same (or other) language. Thus, we must carefully distinguish between the languages used...

- when speaking *in* a given language (called *the language* or *the object language*), and
- when speaking *about* a given language (called *the metalanguage*).

Similarly, we must carefully distinguish between the theory that is the *object* of investigation (called *the theory*) and the theory in which the investigation is being *carried out* (called *the metatheory*). It is possible to give a *method* by which a correct definition of truth can be constructed for *any formal language* ([Tar35] 168).[1] Describing such a method is difficult, and Tarski chooses instead to construct a definition of truth for a specifc formal language, namely, the language of the calculus of classes.

Tarski then begins a meticulous construction of both the language and the metalanguage. Oddly (and at least I think so), Tarski chooses to use *letters* for the operations in the object language (in Polish notation no less) and *symbols* for the metalanguage (in normal infix notation).

Symbol	Meaning
N	negation
A	disjunction
Π	universal quantification
I	inclusion

The real focus of this section is the development of the *metalanguage* for the calculus of classes and the *metatheory*, called the *metacalculus of classes*. The tedious construction is not noted here, but one key aspect of the metalanguage is the following.

> The fact that the metalanguage contains both an individual name and a translation of every expression (and in particular of every sentence) of the language studied will play a decisive part in the construction of the definition of truth... ([Tar35] 172).

Some Key Definitions...

Let S be the class of all well-formed sentences in the language, and $Cn(X)$ be the class of all *consequences* of the class of sentences X.

Definition 17 (Provable Sentence / Theorem). *x is a* provable (accepted) sentence *or a* theorem *– in symbols $x \in Pr$ – if and only if x is a consequence of the set of all axioms.*

Definition 18 (Deductive System). *X is a* deductive system *if and only if*

$$Cn(X) \subseteq X \subseteq S$$

[1] Is this really what he means? If so, has anyone done this?

Definition 19 (Consistency). *X is a* consistent *class of sentences if and only if $X \subseteq S$ and if, for every sentence x, either $x \notin Cn(X)$ or $\neg x \notin Cn(X)$.*[2]

Definition 20 (Completeness). *X is a* complete *class of sentences if and only if $X \subseteq S$ and if, for every sentence x, either $x \in Cn(X)$ or $\neg x \in Cn(X)$.*

Definition 21 (Equivalence). *The sentences x and y are* equivalent *with respect to the class X of sentences if and only if $x \in X$, $y \in S$, $X \subseteq S$ and both $\neg x \vee y \in Cn(X)$ and $\neg y \vee x \in Cn(X)$ (which is the same as saying $x \rightarrow y \in Cn(X)$ and $y \rightarrow x \in Cn(X)$).*

3 The Concept of True Sentence in the Language of the Calculus of Classes

In this section, Tarski constructs a semantic definition of 'true sentence' for the calculus of classes.

4 The Concept of True Sentence in the Languages of Finite Order

still working.

5 The Concept of True Sentence in the Languages of Infinite Order

still working.

6 Summary

still working.

7 Postscript

still working.

References

[Tar35] Alfred Tarski, *'The concept of truth in formalized languages'*, Logic, Semantics and Metamathematics, Oxford University Press, Oxford (1956), 1935, JH Woodger (trans.); First published as 'Der Wahrheitsbegriff in Den Formaliserten Sprachen', *Studia Philosophica* I (1935).

[2] I've never seen consistency phrased this way... just curious.