VIII

THE CONCEPT OF TRUTH IN FORMALIZED LANGUAGES†

Introduction

The present article is almost wholly devoted to a single problem—*the definition of truth.* Its task is to construct—with reference to a given language—*a materially adequate and formally correct definition of the term 'true sentence'.* This problem, which belongs to the classical questions of philosophy, raises considerable difficulties. For although the meaning of the term 'true sentence' in colloquial language seems to be quite clear and intelligible, all attempts to define this meaning more precisely have hitherto been fruitless, and many investigations in which this term has been used and which started with apparently evident premisses have often led to paradoxes and antinomies (for which, however, a more or less satisfactory solution has been found). The concept of truth shares in this respect the fate of other analogous concepts in the domain of the semantics of language.

The question how a certain concept is to be defined is correctly formulated only if a list is given of the terms by means of which the required definition is to be constructed. If the definition is to fulfil its proper task, the sense of the terms in

† Bibliographical Note. The results presented in this paper date for the most part from 1929. I discussed them, in particular, in two lectures given under the title 'On the concept of truth in reference to formalized deductive sciences' at the Logic Section of the Philosophical Society in Warsaw (October 8, 1930) and at the Polish Philosophical Society in Lwów (December 5, 1930). A short report of these lectures is in Tarski [73]. The paper was presented (by J. Łukasiewicz) to the Warsaw Scientific Society on March 21, 1931. For reasons beyond my control, publication was delayed by two years. In the meantime the original text was supplemented by some substantial additions (see p. 247, footnote 1). Also, a summary of the chief results of the paper was published in Tarski [76].

The Polish original of the paper appeared finally in print as Tarski [76*a*]. It was subsequently translated in several languages, first in German (Tarski [76*b*]), and later, in addition to the present English translation, in Italian (Tarski [84*b*]) and French (Tarski [84*c*]). Each of these translations has been provided with a postscript in which some views stated in the Polish original have undergone a rather essential revision and modification.

In two later articles, Tarski [82] and Tarski [84*d*], I have attempted to outline the main ideas and achievements of this paper in a non-technical way. In the first of these articles I have also expressed my views regarding some objections which have been raised to the investigation presented here.

this list must admit of no doubt. The question thus naturally arises: What terms are we to use in constructing the definition of truth? In the course of these investigations I shall not neglect to clarify this question. In this construction I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts.

A thorough analysis of the meaning current in everyday life of the term 'true' is not intended here. Every reader possesses in greater or less degree an intuitive knowledge of the concept of truth and he can find detailed discussions on it in works on the theory of knowledge. I would only mention that throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called *classical* conception of truth ('true—corresponding with reality') in contrast, for example, with the *utilitarian* conception ('true—in a certain respect useful').[1]

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. The problem which interests us will be split into a series of separate problems each relating to a single language.

In § 1 colloquial language is the object of our discussion. The final conclusion is totally negative. In that language it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic.

In the further course of this discussion I shall consider exclusively the scientifically constructed languages known at the present day, i.e. the formalized languages of the deductive sciences. Their characteristics will be described at the beginning of § 2. It will be found that, from the standpoint of the present problem, these languages fall into two groups, the division being based on the greater or less stock of grammatical forms in a particular language. In connexion with the 'poorer' languages the problem of the definition of truth has a positive solution: there is a uniform method for the construction of the required

[1] Cf. Kotarbiński, T. (37), p. 126 (in writing the present article I have repeatedly consulted this book and in many points adhered to the terminology there suggested).

definition in the case of each of these languages. In §§ 2 and 3 I shall carry out this construction for a concrete language in full and in this way facilitate the general description of the above method which is sketched in § 4. In connexion with the 'richer' languages, however, the solution of our problem will be negative, as will follow from the considerations of § 5. For the languages of this group we shall never be able to construct a correct definition of the notion of truth.† Nevertheless, everything points to the possibility even in these cases—in contrast to the language of everyday life—of introducing a consistent and correct use of this concept by considering it as a primitive notion of a special science, namely of the theory of truth, and its fundamental properties are made precise through axiomatization.

The investigation of formalized languages naturally demands a knowledge of the principles of modern formal logic. For the construction of the definition of truth certain purely mathematical concepts and methods are necessary, although in a modest degree. I should be glad if this work were to convince the reader that these methods are now necessary tools even for the investigation of some purely philosophical problems.

§ 1. The Concept of True Sentence in Everyday or Colloquial Language

For the purpose of introducing the reader to our subject, a consideration—if only a fleeting one—of the problem of defining truth in colloquial language seems desirable. I wish especially to emphasize the various difficulties which the attempts to solve this problem have encountered.[1]

[1] The considerations which I shall put forward in this connexion are, for the most part, not the result of my own studies. Views are expressed in them which have been developed by S. Leśniewski in his lectures at the University of Warsaw (from the year 1919/20 onwards), in scientific discussions and in

† Regarding this statement compare the Postscript.

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:

(1) *a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.*†

From the point of view of formal correctness, clarity, and freedom from ambiguity of the expressions occurring in it, the above formulation obviously leaves much to be desired. Nevertheless its intuitive meaning and general intention seem to be quite clear and intelligible. To make this intention more definite, and to give it a correct form, is precisely the task of a semantical definition.

As a starting-point certain sentences of a special kind present themselves which could serve as partial definitions of the truth of a sentence or more correctly as explanations of various concrete turns of speech of the type 'x is a true sentence'. The general scheme of this kind of sentence can be depicted in the following way:

(2) *x is a true sentence if and only if p.*

In order to obtain concrete definitions we substitute in the

private conversations; this applies, in particular, to almost everything which I shall say about expressions in quotation marks and the semantical antinomies. It remains perhaps to add that this fact does not in the least involve Leśniewski in the responsibility for the sketchy and perhaps not quite precise form in which the following remarks are presented.

† Very similar formulations are found in Kotarbiński, T. (37), pp. 127 and 136, where they are treated as commentaries which explain approximately the classical view of truth.

Of course these formulations are not essentially new; compare, for example, the well-known words of Aristotle: 'To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.' (Aristotle, *Metaphysica*, Γ, 7, 27; *Works*, vol. 8, English translation by W. D. Ross, Oxford, 1908.)

place of the symbol 'p' in this scheme any sentence, and in the place of 'x' any individual name of this sentence.

Given an individual name of a sentence, we can construct an explanation of type (2) for it, provided only that we are able to write down the sentence denoted by this name. The most important and common names for which the above condition is satisfied are the so-called *quotation-mark names*. We denote by this term every name of a sentence (or of any other, even meaningless, expression) which consists of quotation marks, left- and right-hand, and the expression which lies between them, and which (expression) is the object denoted by the name in question. As an example of such a name of a sentence the name "'it is snowing'" will serve. In this case the corresponding explanation of type (2) is as follows:

(3) *'it is snowing' is a true sentence if and only if it is snowing.*[1]

Another category of names of sentences for which we can construct analogous explanations is provided by the so-called *structural-descriptive names*. We shall apply this term to names which describe the words which compose the expression denoted

[1] Statements (sentences) are always treated here as a particular kind of expression, and thus as linguistic entities. Nevertheless, when the terms 'expression', 'statement', etc., are interpreted as names of concrete series of printed signs, various formulations which occur in this work do not appear to be quite correct, and give the appearance of a widespread error which consists in identifying expressions of like shape. This applies especially to the sentence (3), since with the above interpretation quotation-mark names must be regarded as general (and not individual) names, which denote not only the series of signs in the quotation marks but also every series of signs of like shape. In order to avoid both objections of this kind and also the introduction of superfluous complications into the discussion, which would be connected among other things with the necessity of using the concept of likeness of shape, it is convenient to stipulate that terms like 'word', 'expression', 'sentence', etc., do not denote concrete series of signs but the whole class of such series which are of like shape with the series given; only in this sense shall we regard quotation-mark names as individual names of expressions. Cf. Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 661–6 and—for other interpretations of the term 'sentence'—Kotarbiński, T. (37), pp. 123–5.

I take this opportunity of mentioning that I use the words 'name' and 'denote' (like the words 'object', 'class', 'relation') not in *one*, but in many distinct senses, because I apply them both to objects in the narrower sense (i.e. to individuals) and also to all kinds of classes and relations, etc. From the standpoint of the theory of types expounded in Whitehead, A. N., and Russell, B. A. W. (90) (vol. 1, pp. 139–68) these expressions are to be regarded as systematically ambiguous.

by the name, as well as the signs of which each single word is composed and the order in which these signs and words follow one another. Such names can be formulated without the help of quotation marks. For this purpose we must have, in the language we are using (in this case colloquial language), individual names of some sort, but not quotation-mark names, for all letters and all other signs of which the words and expressions of the language are composed. For example we could use 'A', 'E', 'Ef', 'Jay', 'Pe' as names of the letters 'a', 'e', 'f', 'j', 'p'. It is clear that we can correlate a structural-descriptive name with every quotation-mark name, one which is free from quotation marks and possesses the same extension (i.e. denotes the same expression) and vice versa. For example, corresponding to the name "'snow'" we have the name 'a word which consists of the four letters: Es, En, O, Double-U (in that order)'. It is thus evident that we can construct partial definitions of the type (2) for structural-descriptive names of sentences. This is illustrated by the following example:

(4) *an expression consisting of three words, of which the first is composed of the two letters I and Te* (*in that order*), *the second of the two letters I and Es* (*in that order*), *and the third of the seven letters Es, En, O, Double-U, I, En, and Ge* (*in that order*), *is a true sentence if and only if it is snowing.*

Sentences which are analogous to (3) and (4) seem to be clear and completely in accordance with the meaning of the word 'true' which was expressed in the formulation (1). In regard to the clarity of their content and the correctness of their form they arouse, in general, no doubt (assuming of course that no such doubts concern the sentences which we substitute for the symbol 'p' in (2)).

But a certain reservation is nonetheless necessary here. Situations are known in which assertions of just this type, in combination with certain other not less intuitively clear premisses, lead to obvious contradictions, for example the *antinomy of the liar*. We shall give an extremely simple formulation of this antinomy which is due to J. Łukasiewicz.

For the sake of greater perspicuity we shall use the symbol '*c*' as a typographical abbreviation of the expression '*the sentence printed on this page, line 5 from the top*'. Consider now the following sentence:

c is not a true sentence.

Having regard to the meaning of the symbol '*c*', we can establish empirically:

(α) '*c is not a true sentence*' *is identical with c.*

For the quotation-mark name of the sentence *c* (or for any other of its names) we set up an explanation of type (2):

(β) '*c is not a true sentence*' *is a true sentence if and only if c is not a true sentence.*

The premisses (α) and (β) together at once give a contradiction:

c is a true sentence if and only if c is not a true sentence.

The source of this contradiction is easily revealed: in order to construct the assertion (β) we have substituted for the symbol '*p*' in the scheme (2) an expression which itself contains the term 'true sentence' (whence the assertion so obtained—in contrast to (3) or (4)—can no longer serve as a partial definition of truth). Nevertheless no rational ground can be given why such substitutions should be forbidden in principle.

I shall restrict myself here to the formulation of the above antinomy and will postpone drawing the necessary consequences of this fact till later. Leaving this difficulty aside I shall next try to construct a definition of true sentence by generalizing explanations of type (3). At first sight this task may seem quite easy—especially for anyone who has to some extent mastered the technique of modern mathematical logic. It might be thought that all we need do is to substitute in (3) any sentential variable (i.e. a symbol for which any sentence can be substituted) in place of the expression 'it is snowing' which occurs there twice, and then to assert that the resulting formula holds for every value of the variable; we would thus