

PANU RAATIKAINEN

TRUTH, CORRESPONDENCE, MODELS, AND TARSKI

In the early 20th century, scepticism was common among philosophers about the very meaningfulness of the notion of truth – and of the related notions of denotation, definition etc. (i.e., what Tarski called semantical concepts). Awareness was growing of the various logical paradoxes and anomalies arising from these concepts. In addition, more philosophical reasons were being given for this aversion.¹ The atmosphere changed dramatically with Alfred Tarski's path-breaking contribution. What Tarski did was to show that, assuming that the syntax of the object language is specified exactly enough, and that the metatheory has a certain amount of set theoretic power,² one can explicitly define truth in the object language. And what can be explicitly defined can be eliminated. It follows that the defined concept cannot give rise to any inconsistencies (that is, paradoxes). This gave new respectability to the concept of truth and related notions. Nevertheless, philosophers' judgements on the nature and philosophical relevance of Tarski's work have varied. It is my aim here to review and evaluate some threads in this debate.

1. EARLY TARSKI AND MODEL THEORY

It has been common (see e.g. Vaught 1974, 1986) to trace the key notion of model theory, the *satisfiability-in-a-structure*, or *truth-in-a-model*, back to Tarski's seminal paper "The concept of truth in formalized languages" (Tarski 1935, henceforth CTFL), and more generally, to associate Tarski's contribution to the theory of truth with model theory. Wilfrid Hodges (1986), however, reports "a disconcerting experience" while reading CTFL to see what Tarski says about the notion of truth-in-a-structure: "The notion was simply not there." According to Hodges, it appears in Tarski's writings only in the 1950s. In agreement with Hodges, Peter Milne (1999) adds that even then Tarski was reluctant to use the term "truth" in the model-theoretic context (See also Feferman 2004).

Ilkka Niiniluoto has repeatedly expressed his disagreement, and suggested that Tarski's early account can be seen as a special case of the model-theoretic approach (Niiniluoto 1994, 1999a, 1999b, 2005). Niiniluoto refers to Tarski's remark, in CTFL, which mentions "the Göttingen school grouped around Hilbert" and recognizes the relative notion of a "correct or true sentence in an individual domain a " (Tarski 1935, p. 199). Niiniluoto also proposes that in his paper on logical consequence from the same period (Tarski 1936), Tarski clearly presupposes the general concept of *truth-in-a-model* when he writes: "The sentence X follows *logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X " (Tarski, 1936, 417).

Solomon Feferman (2004) goes even further. He suggests that the notion of *truth-in-a-structure* is present implicitly already in Tarski's 1931 paper on definability, since Tarski's explication of the concept of *definability-in-a-structure* makes use of satisfaction. Feferman points out that in a footnote to the introduction to this paper Tarski says of the metamathematical definition that "an analogous method can be successfully applied to define other concepts in the field of metamathematics, e.g., that of *true sentence* or of a *universally valid sentential function*" (Tarski 1931, p. 111, fn1). Universal validity can, Feferman adds, only mean valid in every interpretation, and for that the notion of *satisfaction-in-a-structure* is necessary. Feferman also presents an impressive body of evidence that Tarski, just like the early model-theorists who preceded him, had been working comfortably with the informal notion of model at least since 1924.

A somewhat related debate has now been going on in the literature on Tarski's account of logical consequence. Namely, John Etchemendy (1988, 1990) has criticized Tarski for advancing a fixed-domain conception of logical consequence (i.e., the idea that all models share a single domain), which creates all sorts of problems. Gila Sher (1991, 1996), Mario Gómez-Torrente (1996, 1998, 1999), and Greg Ray (1996) have all suggested more a charitable interpretation of Tarski and have proposed that Tarski just could not have intended such an implausible conception. Scott Soames (1999) refers to these replies approvingly.

Timothy Bays (2001), in turn, argues that Tarski definitely adhered to a fixed-domain conception, but also that it does not cause any of the problems which both Etchemendy and his critics assume it causes. Bays' arguments are quite persuasive. Recently Paolo Mancosu (2006) has offered new arguments to show that Tarski indeed upheld a fixed-domain

conception of model in his 1936 logical consequence paper and that he was still propounding that view in 1940. In particular, he provides new evidence from an unpublished lecture by Tarski from 1940 which shows very clearly that Tarski even then held a fixed-domain conception.

To recap, evidently Tarski was, to some extent, thinking in model-theoretical terms and had some kind of notion of *truth-in-a-model* from early on. However, his early (pre-1950s) view was not quite the full-blooded model-theoretical view with variable domains, for it now seems clear that Tarski held onto the fixed-domain conception for quite some time. At best, Tarski may not have always succeeded, in his mathematical work, to be completely faithful to this official view of his. And contrary to appearances, even in his logical consequence paper (Tarski 1936), Tarski did not yet have a wholly *general* notion of *truth-in-a-model* (as e.g. Niiniluoto (1994, 2003) seems to suggest).

Nevertheless, one may grant that whatever is really original in Tarski's formal definition of satisfaction was already there in the 1930s, even if, at the time, the satisfying sequences were picked up from a single comprehensive universe (although, one may think that even there, Tarski was really "belaboring the obvious" and that the definitions of satisfaction and truth are "practically forced on us"; see Feferman 2004). The relativization of this notion to arbitrary domains was, once one gave up the philosophical obstacles to it, certainly a routine move.

2. THE CONCEPT OF TRUTH AND TRUTH-IN-A-MODEL

But what is, more exactly, the relation of Tarski's work on truth to model theory? In modern model theory, the standard approach is now the following: Given a language L and a structure \mathcal{W} with a domain D , one fixes an interpretation function I which maps the non-logical symbols of L to elements of D (that is, the function I maps individual constants to elements of D , predicates to subsets of D , etc.). Consequently, an L -structure \mathcal{W} is often defined as a pair (D, I) , consisting of the domain D and the interpretation function I . In such a model-theoretic setting, a language L is completely a uninterpreted and syntactic formal language. Niiniluoto adds that "an interpreted language could be defined as the pair (L, I) " (Niiniluoto 2005, p. 64).

Tarski's approach to truth differs from such a model-theoretic approach in several important respects. We have already discussed Tarski's early commitment to a single and fixed comprehensive universe. This is, of course, quite different from modern model theory, where one is free to choose any arbitrary set as the domain. But there are also other differences.

In model theory, languages are uninterpreted, and when a model is switched to another, one varies the interpretation, but the language remains the same. In his writings on the concept of truth, Tarski, on the other hand, repeatedly insisted that the 'formalized' languages whose truth is under consideration were, and had to be, always already interpreted languages:

It remains perhaps to add that we are not interested here in 'formal' languages and sciences in one special sense of the word 'formal', namely sciences to the signs and expressions of which no meaning is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful. We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider. (Tarski 1935, 166–167)

I should like to emphasize that, when using the term 'formalized languages', I do not refer exclusively to linguistic systems that are formulated entirely in symbols, and I do not have in mind anything essentially opposed to natural languages. On the contrary, the only formalized languages that seem to be of real interest are those which are fragments of natural languages (fragments provided with complete vocabularies and precise syntactical rules) or those which can at least be adequately translated into natural languages. (Tarski 1969, 68)

Furthermore, this was not just an accidental philosophical opinion from Tarski's side, but it is an essential part of Tarski's whole approach to truth that the meanings of the object language must be fixed. Only that way can a truth definition (applied to sentences) make any sense at all:

[W]e must always relate the notion of truth, like that of a sentence, to a specific language; for it is obvious that the same expression which is a true sentence in one language can be false or meaningless in another. (Tarski 1944, 342)

We shall also have to specify the language whose sentences we are concerned with; this is necessary if only for the reason that a string of sounds or signs, which is a true or a false sentence but at any rate meaningful sentence in one language, may be a meaningless expression in another. (Tarski 1969, 64)

[T]he concept of truth essentially depends, as regards both extension and content, upon the language to which it is applied. We can only meaningfully say of an expression that it is true or not if we treat this expression as a part of a concrete language. As soon as the discussion concerns more than one language the expression 'true sentence' ceases to be unambiguous. If we are to avoid this ambiguity we must replace it by the relative term 'a true sentence with respect to the given language'. (Tarski 1935, 263)

Therefore, it is necessary in Tarski's setting to focus on an interpreted language with constant meanings. If one changes the interpretation of the symbols of the object language, the language changes to a different language, and a former truth definition is not a truth definition for this latter language.

Can this difference be overcome by following Niiniluoto's above-mentioned suggestion that an interpreted language is defined as the pair (L, I) ? Now although this idea works perfectly in the ordinary model-theoretic context, I do not think that it is an acceptable line to take in a Tarskian approach to truth. Namely, Tarski expressly aimed to define truth (or, rather, "true-in- L ") without assuming any semantic notions: "In this construction [of the definition of truth] I shall not make use of any semantical concept if I am not able to previously reduce it to other concepts" (Tarski 1935, 153). The interpretation function I , however, establishes a link between the language and a domain of extra-linguistic objects, and hence is a semantical concept in Tarski's sense (see also below). Hence, it would be problematic to presuppose it in the definition of truth. Although Tarski assumed that the object language must be an interpreted language, its interpretation cannot be specified by leaning on the model-theoretic interpretation function.

But the question then arises how Tarski can, and indeed can he, specify the object language as an interpreted language with meanings, without begging the question. Rudolf Carnap, in his logical semantics, assumed that the interpretation of the object language is fixed with the help of truth conditions, which in turn appeal to the definition of truth. Whether this makes Carnap's approach viciously circular or not, it is important to note that this is not Tarski's approach. Tarski here explicitly points out the difference between his own approach and that of Carnap (see Tarski 1944, 373, note 24). For Tarski, the interpreted object language is instead specified simply through its metalinguistic translation (see e.g. Tarski 1935, 170-71; cf. Fernández Moreno 1992, 1997; Milne 1997, Raatikainen 2003, Feferman 2004).

However, Tarski's approach still assumes the notion of *meaning*, in the disguise of translation or the sameness of meaning. Does this mean that, at the end of the day, Tarski fails to achieve his aim, that is, to define truth without assuming any *semantical* concepts? It has been repeatedly suggested that this is indeed the case (see Davidson 1990, 1996, Field 1972,

Soames 1984). But it is not necessarily so. In order to find out, we need to take a closer look on what Tarski meant by ‘semantical’. Tarski’s paradigm examples of semantical concepts were satisfaction, denotation, truth and definability (see Tarski 1935, 164, 193-4; 1936, 401). He explained his understanding of ‘semantical concept’ as follows:

A characteristic feature of the semantical concepts is that they give expression to certain relations between the expressions of language and the objects about which these expressions speak, or that by means of such relations they characterize certain classes of expressions or other objects. (1935, p. 252)

Now the model-theoretic interpretation function I discussed above is definitely a semantic notion in this sense. Hence, it would be against Tarski’s explicit commitments to assume it in defining truth. This cannot be the way in which the object language is interpreted. But how about translation? I submit that it is possible to view translation, in this context, as a purely syntactic mapping between two languages, without assuming any relations between either language and the external objects. Translation, so viewed, is not a semantical concept in Tarski’s sense. Hence, it is admissible for Tarski to presuppose it in this approach (see also Milne 1997).

But let us take a closer look at the details. The interpretation, or translation, of the object language in the metalanguage is specified, in Tarski, through primitive denotation. Let us then recall how exactly Tarski specifies primitive denotation in the object language. For names, this is done by a simple list-like explicit definition such as:

Denotes_{OL}(x , y) \leftrightarrow

$$\begin{aligned} &[(x = \text{‘Frankreich’} \ \& \ y = \text{France}) \vee \\ &(x = \text{‘Deutschland’} \ \& \ y = \text{Germany}) \vee \\ &\quad \quad \quad : \\ &\quad \quad \quad : \\ &(x = \text{‘Köln’} \ \& \ y = \text{Cologne})]. \end{aligned}$$

An analogous definition can be given for the denotation (or, application) of predicates. Such an enumerative characterization of primitive denotation may be philosophically disappointing (cf. Field 1972), but at least it frees Tarski from any charge of begging the question. The interpretation of the object language is fixed through fixing primitive denotation, which in

turn can be done by explicit definitions. And what can be explicitly defined can be eliminated. Certainly Tarski's reliance on such notions is unlikely to be problematic.

3. IS TARSKI'S TRUTH DEFINITION A CORRESPONDENCE THEORY?

The question of whether Tarski's account is a version of the correspondence theory of truth or not has resulted much debate among philosophers. Karl Popper famously declared that Tarski had "rehabilitated the correspondence theory of absolute or objective truth" and "vindicated the free use of the intuitive idea of truth as correspondence to the facts" (Popper 1963, 223). Also, Ilkka Niiniluoto (1994, 1999a, 1999b, 2003), Gila Sher (1998) and Luis Fernández Moreno (2001) have argued that Tarski's definition of truth is a correspondence theory.

Susan Haack, among others, disagrees: "Tarski did not regard himself as giving a version of the correspondence theory" (Haack 1978, 114). According to Haack, Tarski's notion of satisfaction at best "bears some analogy to correspondence theories" (*ibid.*). However, she adds, "Tarski's definition of truth makes no appeal to specific sequences of objects, for true sentences are satisfied by all sequences, and false sentences by none" (Haack 1978, 113). A. C. Grayling (1998, 156) largely repeats Haack's criticism. Earlier, Donald Davidson (1969, 1983) took Tarski's account as a variant of correspondence theory, but later changed his mind for reasons similar to those of Haack: "[T]here is nothing interesting or instructive to which true sentences might correspond" (Davidson 1990; see also Davidson 1996).

I think that the objection of Haack and others, which leans on the fact that truth amounts to satisfiability by *all* sequences, is less conclusive than it may appear to be. To begin with, one can define truth for atomic sentences without the notion of satisfaction. In their case, it is particular individuals and their properties and relations which make a sentence true (cf. Niiniluoto 1999a, 2004). Further, I think that even in the case of quantified sentences the situation is not as desperate as Haack and others suggest. To be sure, a sentence is true if and only if it is satisfied by *every* sequence of objects. However, this is more a consequence of a technical trick Tarski used in his definition of satisfaction in order to handle quantification.

Namely, let us consider, for example, the existentially quantified sentence $(\exists x_1) P(x_1)$, and let us assume that the intended interpretation of $P(x)$ is, say: "x is a Ph.D. student of Leśniewski". According to Tarski's definition of satisfaction, a sequence σ satisfies $(\exists x_1) P(x_1)$ if and only

if *some* sequence σ' , which agrees with σ except possibly at the variable x_1 , satisfies $P(x_1)$. The only sequences σ' that will do (i.e., satisfy $P(x_1)$) are ones which have Tarski as their first member.³ Moreover, the rest of the sequence is irrelevant and could be omitted (Tarski assumed, for simplicity, that all such satisfying sequences are infinite; but it is well known that one could manage with just finite sequences; in such case, the finite sequence with Tarski as its first and only member would be the only relevant sequence σ'). Hence, it is quite plausible to consider Tarski, and nothing else, as the “truth-maker” of the existentially quantified sentence $(\exists x_1) P(x_1)$, even if the sentence is, according to the technical definition of satisfaction, satisfied by every sequence of objects. Given an arbitrary sequence σ , we are, so to say, allowed to switch its relevant member (here, the first member) to a relevant object (here, only Tarski is suitable) and produce a sequence σ' which does the real work.

But the question of whether Tarski’s account is a full-blown substantial correspondence theory of truth, in contradistinction to deflationist views on truth, is different and more complicated. Part of the difficulty is, of course, that it is not altogether clear what exactly is the essence of deflationism. For example, Stephen Leeds (1978), Paul Horwich (1982) and Scott Soames (1984) have all suggested that Tarski’s truth definition amounts in fact to a deflationary theory of truth (cf. Davidson 1990). As we have already seen, Fernández Moreno, Niiniluoto and Sher, for example, disagree and argue that it can be instead seen as a correspondence theory.

It is useful to distinguish, in this context, between weak and strong correspondence theories (Woleński & Simons 1989), or, correspondence-as-congruence and correspondence-as-correlation (see Kirkham 1992, 119). According to the *weak correspondence theories* or *correspondence-as-correlation* views, every truth-bearer is correlated to a state of affairs, and if that state of affairs to which a given truth bearer is correlated actually obtains, the truth bearer is true; otherwise it is false. The *strong correspondence theories*, or *correspondence-as-congruence* views, require further that there is a structural isomorphism between truth bearers and the facts to which they correspond, if true; a truth bearer mirrors or pictures the state of affairs to which it is correlated. Nothing of the sort is assumed by the former, weaker idea of correspondence. According to it, a truth bearer as a whole is correlated to a state of affairs as a whole. Weak correspondence involves only the idea that truth depends on how things are in the world.

Jan Woleński and Peter Simons (1989) submit that Tarski's theory is a correspondence theory only in the weak (or correlation) sense. Sher (1998), on the other hand, argues that it is a correspondence theory even in the strong sense (or this is at least how Patterson (2003) interprets her). Niiniluoto (1999a, 2004) in turn argues that in the case of atomic sentences, Tarski's theory is a strong correspondence theory, but with compound and quantified sentences, only a weak correspondence theory.

But what are the grounds for thinking that Tarski's truth definition really is a version of correspondence theory? Popper (1960, p. 224) seems to think that T-sentences state correspondences between sentences and worldly facts. Similarly, Niiniluoto writes that a T-sentence "states something about the relation between the language L and the world", and hence, "Tarski's semantic definition of truth is not merely disquotational" (Niiniluoto 1994, 63). Also Sher (1998) makes analogous claims. So did Davidson at one point (Davidson 1983; but see Davidson 1969, 1990).

But, as Douglas Patterson (2003) points out, if it is assumed that T-sentences as such state correspondences between the sentences they mention and something extra-linguistic, then even deflationary and disquotational theories are correspondence theories, at least in the weak sense. This, however, is far too weak a notion of correspondence to be of any interest if we wish to understand what is at issue between contemporary deflationists and their correspondence theoretic opponents. Only strong correspondence theories will be interesting from this perspective, Patterson concludes. One must agree with Patterson's main point. However, contrary to what he may seem to suggest, having a genuine and substantive correspondence theory does not necessarily require a general strong or congruence view of correspondence.⁴ The two distinctions substantial/deflationary and weak/strong (correspondence) do not coincide. Patterson also points out that T-sentences simply are not of the right form to state a relation at all, and so cannot state a correspondence relation. A T-sentence is a biconditional and does not predicate a relation between the sentence it mentions and some other objects.

Michael Devitt (2001) in turn argues that although Tarski seemed to view himself as a correspondence theorist about truth, the theory he actually presented is deflationary. Namely, he first reminds us that – as especially Hartry Field (1972) has emphasized – Tarski's truth

definition rests on a list-like definition of primitive denotation (see also above). “But such list-like definitions are in no way explanatory, but are essentially deflationary and so could not yield anything substantial about reference.” Consequently, Devitt maintains, Tarski’s truth definition itself does not show us anything substantial about truth: “Tarski’s definition tells us a lot about ‘true-in-L’. It tells us nothing about truth-in-L because it is implicitly committed to the view that there is nothing to tell.” I think we must accept Devitt’s conclusion.

However, a fix is now available. Devitt too adds that his conclusion, that Tarski’s definition tells us nothing about truth, concerns only Tarski’s definition as it stands. However, if we revised it by dropping its list-like definitions, then we could see it as yielding an explanation of truth in terms of reference, as Field points out. If this were then supplemented by a substantial theory of reference,⁵ we would have a genuine correspondence theory of truth, Devitt concludes. Patterson too agrees that this theory is indeed a real correspondence theory. There seems to be no question that such a modified Tarskian theory of truth is a robust and substantial correspondence theory.

This move has, however, its price. One must then relax Tarski’s initial requirement that no semantical concepts are presupposed. But this just is the price one necessarily has to pay if one wants to turn Tarski’s definition into a substantial account of truth. However, unlike the more general semantical notions, primitive denotation is a very elementary notion and does not lead to any paradoxes.⁶ Hence, it is a rather harmless concession.

NOTES

1. For more of the historical background, see e.g. Niiniluoto 1999b, Sluga 1999, Woleński and Simons 1989.

2. But one should exaggerate the amount of set theory needed. Relatively little suffices. For example, if the object language is the language of first-order arithmetic (i.e., that of PA), the relatively weak and predicative subsystem of second-order arithmetic ACA is sufficient. Tarski certainly thought that this much set theory is quite unproblematic, especially when compared to the semantical notions in question.

3. Tarski was Leśniewski's one and only Ph.D. student.
4. For example, a broadly Tarskian theory supplemented with a substantial theory of reference along the lines that Field has suggested (mentioned at the end of this paper) is agreed by all parties to be a substantial theory of truth; however, there is no reason to think that it has to be a strong correspondence theory (correspondence-as-congruence).
5. However, I don't think that it is likely that we will ever have a strictly physicalistic theory of reference, as Field demands – but that is a wholly different and independent issue.
6. The situation is, of course, very different with the general notion of denotation, which easily leads, e.g., to Berry's paradox.

REFERENCES

- Bays, T.: 2001, 'On Tarski on Models', *The Journal of Symbolic Logic* 66, 1701-1726.
- Davidson, D.: 1969, 'True to the facts', reprinted in D. Davidson, *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 37-54.
- Davidson, D.: 1986, 'A Coherence Theory of Truth and Knowledge', *Kant oder Hegel*. Edited by D. Henrich. Stuttgart: Klett-Cotta, Stuttgart, 1983. 423-438. Reprinted in: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Edited by E. LePore. Oxford: Blackwell, 1986.
- Davidson, D.: 1990, 'The Structure and Content of Truth', *Journal of Philosophy* 87, 279-328.
- Davidson, D.: 1996, 'The Folly of Trying to Define Truth', *Journal of Philosophy* 93, pp.263-78.
- Devitt, M.: "The Metaphysics of Truth." In *The Nature of Truth*, Michael Lynch, ed. Cambridge, MA: MIT Press (2001), pp. 579-611.
- Etchemendy, J.: 1988, 'Tarski on Truth and Logical Consequence', *Journal of Symbolic Logic* 53, 51-79.
- Etchemendy, J.: 1990, *The Concept of Logical Consequence*, Harvard, Cambridge.
- Feferman, S.: 2004, 'Tarski's conceptual analysis of semantical notions', in A. Benmakhlof (ed.), *Sémantique et épistémologie*, Editions Le Fenec, Casablanca [distrib. J. Vrin, Paris], pp. 79-108.
- Fernandez Moreno, L.: 1992, 'Putnam, Tarski, Carnap und die Wahrheit', *Gräzer philosophische Studien* 43, 33-44.
- Fernandez Moreno, L.: 1997, 'Truth in Pure Semantics: A Reply to Putnam', *Sorites* 8, 15-23.
- Fernandez Moreno, L.: 2001, 'Tarskian Truth and the Correspondence Theory', *Synthese* 126, 123-147.
- Field, H.: 1972, 'Tarski's Theory of Truth', *Journal of Philosophy*, 69, 347-75; page
- Gomez-Torrente, M.: 1996, 'Tarski on logical consequence', *Notre Dame Journal of Formal Logic* 37, 125-151.

- Gomez-Torrente, M.: 1998, 'On a fallacy attributed to Tarski', *History and Philosophy of Logic* 19, 227-234.
- Gomez-Torrente, M.: 1999, 'Logical truth and Tarskian logical truth', *Synthese* 117, 375-408.
- Haack, S.: 1978, *Philosophy of Logics*, Cambridge University Press, Cambridge.
- Hodges, W.: 1985/6, 'Truth in a Structure', *Proceedings of the Aristotelian Society*, 86, 135-151.
- Leeds, S.: 1978, 'Theories of reference and truth', *Erkenntnis* 13, 111-27
- Milne, P.: 1997, 'Tarski on Truth and Its Definition', in Childers, Kolár and Svoboda (eds.), *Logica '96: Proceedings of the 10th International Symposium*, Filosofia, Prague, pp. 189-210.
- Milne, P.: 1999, 'Tarski, Truth and Model Theory', *Proceedings of the Aristotelian Society* **XCIX** (1998-1999), 141-167.
- Mancosu, P.: (2006), 'Tarski on models and logical consequence', in J. Gray and J. Ferreiros (eds.), *The Architecture of Modern Mathematics*, Oxford University Press. Forthcoming.
- Niiniluoto, I.: 1994, 'Defending Tarski against his Critics', in B. Twardowski and J. Woleński (eds.), *Sixty Years of Tarski's Definition of Truth*, Philed, Warsaw, pp. 48-68.
- Niiniluoto, I.: 1999a, 'Tarskian truth as correspondence – replies to some objections', in J. Peregrin (ed.), *Truth and its Nature (if any)*, Kluwer, Dordrecht, 91-104.
- Niiniluoto, I.: 1999b, 'Theories of truth: Vienna, Berlin, and Warsaw', in J. Woleński and E. Köhler (eds.), *Alfred Tarski and the Vienna Circle*, Kluwer, Dordrecht, 17-26.
- Niiniluoto, I.: 2004, 'Tarski's definition and truth-makers', *Annals of Pure and Applied Logic* 126, 57-76.
- Patterson, D.: 2003, 'What is a correspondence theory of truth?', *Synthese* 137, 421-444.
- Popper, K. R.: 1963, *Conjectures and Refutations*, Routledge, London.
- Putnam, H.: 1985-86, 'A comparison of something with something else', *New Literary History* 17, 61-79.
Reprinted in H. Putnam: *Words and Life*. Cambridge: Harvard University Press, (1994).
- Raatikainen, P.: 2003, 'More on Putnam and Tarski', *Synthese* 135, 37-47.
- Ray, G.: 1996, 'Logical consequence: a defence of Tarski', *The Journal of Philosophical Logic* 25, 617-677.
- Sher, G.: 1991, *The Bounds of Logic*, MIT Press, Cambridge.
- Sher, G.: 1996, 'Did Tarski commit Tarski's fallacy?', *Journal of Symbolic Logic* 61, 653-686.
- Sher, G.: 1998, 'On the Possibility of a Substantive Theory of Truth', *Synthese* 117, 133 – 172.
- Sluga, H.: 1999, 'Truth before Tarski', in J. Woleński and E. Köhler (eds.), *Alfred Tarski and the Vienna Circle*, Kluwer, Dordrecht, 27-41.
- Soames, S.: 1984, 'What is a Theory of Truth?', *Journal of Philosophy* 81, 411-29.
- Soames, S.: 1999, *Understanding Truth*, Oxford University Press, New York.
- Tarski, A.: 1931, 'Sur les ensembles définissables de nombres réels, I.', *Fundamenta Mathematicae* 17, 210-239;
English translation in Tarski, 1983, 110-142.
- Tarski, A.: 1933, *Pojęcie prawdy w językach nauk dedukcyjnych* (The concept of truth in the languages of deductive sciences), *Prace Towarzystwa Naukowego Warszawskiego, wydział III, no. 34*.
- Tarski, A.: 1935, 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* (Lemberg), 1, 261-405. English translation in Tarski 1983, 152-278.

- Tarski, A.: 1936, 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophie Scientifique 7*, Actualités Scientifiques et Industrielles, Herman, Paris, pp. 1-11. English translation in Tarski 1983, 409-420.
- Tarski, A.: 1940, 'On the Completeness and Categoricity of Deductive Systems', unpublished typescript, Alfred Tarski Papers, Carton 15, Bancroft Library, U.C. Berkeley.
- Tarski, A.: 1944, 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* **4**, 341-376.
- Tarski, A.: 1969, 'Truth and Proof', *Scientific American* **220**, 63-77.
- Tarski, A.: 1983, *Logic, Semantics, Metamathematics*, Oxford University Press, Oxford. Second edition. (First edition, 1956).
- Vaught, R. L.: 1986, 'Tarski's work in model theory', *Journal of Symbolic Logic* **51**, 869-882.
- Vaught, R. L.: 1974, 'Model theory before 1945', in L. Henkin, J. Addison, C. C. Chang, W. Craig, D. Scott, R. Vaught (eds.), *Proceedings of the Tarski Symposium*, American Mathematical Society (Providence), pp. 153-172.
- Woleński, J. and Simons, P.: 1989, 'De Veritate: Austro-Polish Contributions to the Theory of Truth from Brentano to Tarski', in K. Szaniawski (ed.), *The Vienna Circle and the Lvov-Warsaw School*, Dordrecht, Kluwer, pp. 391-442.