

Truth in Tarski's Philosophy: Logical Foundations and Semantic Challenges

تعريف الصدق في فلسفة تارسكي: الأسس المنطقية والإشكالات الدلالية

Naim Bouzidi ¹

¹Yahya Fares University - Medea (YFUM), psda.: (Algeria)

bouzidi.naim@univ-medea.dz

Received: 24/09/2025

Accepted: 14/01/2026

Published:23/01/2026

Abstract:

This research aims to analyze Tarski's theory of truth within formal languages, where he sought to provide a precise definition that overcomes the ambiguities associated with the term "truth" in natural languages. Tarski relied on semantic analysis linking truth to logic and mathematics, using the principle of bivalence (true/false) to resolve philosophical contradictions like the "liar paradox."

The study concluded that Tarski's theory effectively contributes to providing a precise definition of truth in scientific and formal contexts, despite criticisms from philosophers like Donald Davidson, who questioned the usefulness of concepts such as "correspondence" and "states of affairs." The research also demonstrated

Keywords: Tarski, truth, formal languages, semantic analysis, logic.

ملخص:

يهدف هذا البحث إلى تحليل نظرية تارسكي في تعريف الصدق ضمن اللغات الصورية، حيث سعى لتقديم تعريف دقيق يتجاوز الغموض المرتبط بكلمة "الصدق" في اللغات الطبيعية. اعتمد تارسكي على التحليل الدلالي الذي يربط الصدق بالمنطق والرياضيات، مستخدماً مبدأ ثنائية القيمة (الصدق والكذب) لحل التناقضات الفلسفية مثل "مفارقة الكاذب".

توصل البحث إلى أن نظرية تارسكي تساهم بشكل فعال في توفير تعريف دقيق للصدق في السياقات العلمية والصورية، رغم الانتقادات التي وجهها بعض الفلاسفة مثل دونالد ديفيدسون، الذين شككوا في فعالية مفاهيم مثل "المطابقة" و"حالات الأشياء". وأظهر البحث أيضاً كيف يمكن توسيع تطبيقات هذه النظرية في المنطق والفلسفة.

كلمات مفتاحية: تارسكي، الصدق، اللغات الصورية، التحليل الدلالي، المنطق.

Corresponding author: naim Bouzidi

1. Introduction

The quest for a precise and formal definition of truth has been central to 20th-century analytical philosophy. Among the most influential contributions in this domain is Alfred Tarski's work, which is marked by its rigorous and systematic approach to truth, particularly in the context of formal languages. While Tarski adhered to the Aristotelian tradition, his goal was to formulate a definition of truth that would transcend the ambiguities inherent in

the everyday use of the term "true" in natural languages. Rather than focusing on the meaning of "truth" in ordinary language, Tarski opted for a semantic analysis that links truth to a logical and mathematical framework, providing a rigorous definition applicable to scientific and formal contexts (TARSKI, 2021) (Tarski, 1944) (TARSKI, 1969).

The Aristotelian philosophy defines truth as a correspondence between a proposition and the state of affairs it describes: a proposition is true if it corresponds to reality. This classical notion of truth inspired Tarski, yet he recognized its limitations, especially when applied to logical propositions or more complex contexts. Tarski's aim was not to reject the Aristotelian view but to reinterpret it within the formal sciences, offering a more precise definition to resolve the ambiguities of natural language (Blackburn & Simmons, 1999) (Kirkham, 1992).

Tarski's project differs from prior approaches in that it not only defines truth from a linguistic perspective but anchors it within a logical and mathematical structure that overcomes the challenges posed by the varied uses of the term in different contexts. However, this approach has faced criticism, notably from philosophers like Donald Davidson, who questioned the utility of concepts such as "correspondence" or "state of affairs" for understanding truth, arguing that they unnecessarily complicate the discourse without contributing essential elements to its definition (Blackburn & Simmons, 1999).

This article explores Tarski's semantic theory of truth, examining its theoretical foundations, implications, and criticisms. We will particularly focus on how Tarski redefined the classical notion of truth to meet the demands of the formal sciences and how his approach has paved the way for new reflections on the nature of language and truth.

2. Application of the term "True" to Propositions

In this section, I will discuss the entities to which the term "true" can be applied according to Tarski. The concept of "veridicality"—that which is attributed a truth value—must first be explained. In everyday life and ordinary language, the term "true" is not clearly defined. For example, in English, "true" is used in expressions like "true friend," which bears little relation to the concept of truth itself. Similarly, logicians and philosophers do not offer a clear answer to what precisely can be termed "true." The candidates for this designation include beliefs, propositions, judgments, assertions, theories, mentions, ideas, acts of thought, statements, types of sentences, unspecified utterances, and speech acts.

Since Tarski sought to establish a theoretical semantics within logic, one might suggest that "true" should be applied to propositions. However, Tarski rejected this possibility, as he deemed the term "proposition" too ambiguous and its meaning fluctuating. The term "proposition" has long been a subject of endless disputes among philosophers and logicians, and it appears never to have been explained clearly and unambiguously. For various reasons, it seems more appropriate to apply the term "true" to statements, and we shall follow this practice.

At the time Tarski wrote his major works on truth, the term "proposition" was not clearly defined, and his rejection of the term is well justified. Peter Milne summarizes Tarski's view on this matter as follows: "[For Tarski, truth is attributed] to statements, not to

their content, nor to statements combined with their content" (Milne, 1997). Thus, for Tarski, truth pertains to statements, not to their content or meaning.

Tarski's conception of truth as applying to statements, rather than their content or meaning, raises the question of what Tarski understands by "statements"—whether he considers them as types, tokens, or unspecified utterances. Unfortunately, the answer to this question is not unequivocal. Tarski (1944) illustrates the concept of "statement" in a way that can be understood as "classes of statements of a similar form," which aligns closely with the traditional conception of statement types as defined by Kirkham. However, Kirkham notes that Tarski initially supported the idea of statement types playing the role of "truth bearers," but later, towards the end of his life, he shifted to using statement tokens as truth bearers in his theory. This shift occurred in the late 1960s. Given that the work at hand is based on Tarski's writings prior to this shift, we will interpret "statement" as referring to statement types. Nonetheless, Tarski did not oppose the idea that his conception of truth could be adapted to allow other entities to play the role of truth bearers (Tarski, 1944).

2.1. Formal Correctness Criterion

Tarski imposed two requirements on his definition of truth. I now turn to the first of these: formal correctness. Tarski himself did not provide detailed explanations for this condition. As Milne observes, Tarski's second requirement, material adequacy, has received more attention in the literature. The formal correctness condition—sometimes described as methodological correctness—is less clearly specified than material adequacy. In addition to imposing restrictions on what can properly belong to the metalanguage, it demands that the definition of truth meet the standard criteria typically required of definitions.

Unfortunately, secondary literature does not provide much clarity on this point. Both Tarski and other philosophers and logicians tend to emphasize the requirement of material adequacy, which I will address in the next section. However, Tarski himself wrote about formal correctness as follows: "We must specify the words or concepts we wish to use to define truth and submit the definition to the formal rules it must adhere to. In other words, we must describe the formal structure of the language in which the definition will be formulated" (Tarski, 1944).

Tarski's goal in formulating his definition was not to abandon standard logical laws and principles. One decisive principle is the principle of bivalence—the requirement that every statement must have exactly one of two truth values: true or false. The formal correctness criterion aims to prevent falsehoods or paradoxes from arising from Tarski's definition of truth. This is closely related to the second requirement, material adequacy, which I will discuss in the next chapter. Tarski was particularly concerned with the possibility of generating the liar paradox.

A version of the paradox can be formulated as follows:

(LP) The sentence (LP) is false.

For (LP), ordinary logical rules must apply. With other statements, this would pose no difficulty, but in the case of (LP), a paradox arises. It can be schematized as follows:

1. (LP) is a sentence.
2. Every sentence is either true or false, but not both.

3. Therefore, (LP) is either true or false, but not both.
4. If (LP) is true, then what it asserts is true; but (LP) asserts its own falsity. Hence, (LP) is false.
5. If (LP) is false, then what it asserts is not true; but it asserts its own falsity. Hence, (LP) is true.
6. Therefore, (LP) is both true and false—by (4) and (5).
7. But (6) contradicts (2) and (3), leading to the paradox.

The paradox lies in the simultaneous truth and falsity of (LP). One might attempt to avoid the liar paradox by rejecting or modifying one of the premises above. However, Tarski would disagree, as the given schema does not account for all the premises relevant to the emergence of the paradox. This solution has, however, been challenged by modified versions of the paradox, such as the revenge liar. Tarski's work has also been critiqued through related paradoxes, such as those of Jourdain and Buridan, which are specific variants of the liar paradox (Milne, 1997).

2.2. The Criterion of Material Adequacy

It is necessary to revisit Tarski's approach to the problem of truth. Tarski's primary goal was not to analyze the term "true" as it is used in ordinary language, but rather to construct a definition of truth that would align with the intuitions associated with the classical Aristotelian concept of truth (Tarski, 1944). As (SOAMES, 1999) clarifies, "For Tarski and many others, our ordinary concept of truth is defective precisely because its lack of boundaries leads to paradoxes. The analysis Tarski attempts to provide is one that removes this supposed flaw while preserving the central and useful elements of the concept." (SOAMES, 1999)

Tarski did not seek to specify how the term "true" is commonly understood but instead aimed to clarify how it should be understood in logical, mathematical, and scientific theories. Such analyses are often referred to as explications (COFFA, 1987). Tarski's approach can be expressed in Carnap's definition of explication: "To say that a pre-scientific concept must be transformed into an exact concept means introducing a precise concept that corresponds to the given concept." According to Tarski, the term "true" as used in everyday language cannot serve as the explication's subject matter (explicandum) because its vagueness would likely lead to an explanation that deviates from what Tarski had in mind. Instead, Tarski's explication refers to the classical theory of correspondence and its associated intuitions, which clarify the problem of truth (Tarski, 1944).

However, Tarski's alignment with the correspondence theory of truth has been questioned by some philosophers, who argue that his conception of truth cannot be strictly classified as a correspondence theory (SOAMES, 1999). Tarski's objective was not to develop a new theory of truth but to provide an adequate explanation of the correspondence theory, most notably in its classical Aristotelian form. Aristotle's definition holds that a statement is true when what it asserts corresponds to the state of affairs in the world: "Snow is white" is true if and only if snow is white. For Tarski, such equivalences must follow directly from the definition of truth, as they represent intuitively true statements (Tarski, 1944). Without such equivalences, a definition of truth would not qualify as a

comprehensive theory of truth, as there would be true statements that could not be derived from it.

3. The T-schema and T-scheme

To determine which sentences must follow from the definition of truth, it is necessary to generalize statement (5) from the previous chapter. This is done by noting that in the equivalence (5), the expression "snow is white" appears twice: once in quotation marks (on the left) and once without (on the right). The appearance of the name of the expression on the left is crucial. Tarski remarked that, "If we want to say a sentence is true, we must use the name of the sentence, not the sentence itself" (Tarski, 1944). This distinction between a sentence and its name is fundamental to understanding Tarski's formalization of truth.

This generalization leads to the following schematic form:

(T) X is true if and only if p.

This is the T-schema, and each equivalence obtained by substitution within this schema is referred to as a T-equivalence or a T-sentence. The T-schema plays a crucial role in Tarski's conception of truth, which can be more fully understood through the so-called Convention T:

Convention T: A formally correct definition of the symbol "Tr" in the metalanguage is called an adequate definition of truth if it has the following consequences:

- (α) All sentences derived from " $x \in \text{Tr}$ if and only if p," by replacing "x" with the descriptive name of any sentence in the object language and replacing "p" with the translation of that sentence into the metalanguage.
- (β) The sentences "for all x, if $x \in \text{Tr}$, then $x \in S$ " (in other words, $\text{Tr} \subseteq S$).

The critical role of consequence (α) is that an adequate definition of truth must lead to all T-equivalences when substituting sentences from the object language into the T-schema. However, it is essential to note that the T-schema itself cannot be a theory of truth because it merely presents a pattern of sentences rather than a complete theory (Tarski, 1944).

2.2 Extension of the T-schema

In the previous chapter, I presented Tarski's T-schema as used in his work. However, this version is somewhat "reduced" because it includes only the predicate "true" or "Tr" (in the more formal article *On the Concept of Truth in Formalized Languages*). This simplification can be confusing, as Tarski explicitly stated that the term "true" must always be relativized to a language. The ideal way to make this relativization explicit would be to incorporate it directly into the T-schema, which Tarski himself did not do. Fortunately, commentators on Tarski have typically addressed this issue, as noted by Boisvert (BOISVERT, 1999): "If the truth predicate is applied to sentences, it must necessarily be relativized to a language, as the same string of symbols can have different meanings in different languages."

In the literature, the use of a truth predicate explicitly relativized to the language it applies to is common. An example of a T-schema that accounts for language dependency is found in Soames's formulation:

T-schema: X is T in L if and only if p.

Here, X represents the name of the sentence, p its translation into the metalanguage, T the truth predicate, and L the language with respect to which the truth predicate is applicable. It is important to note that T and L are typically presented together as a unit. Various notations for the relativized truth predicate appear in the literature, such as "true-in-L" or "Tr L."

Since Tarski was concerned primarily with constructing a definition of truth for formalized languages, relativizing the truth predicate to the language may seem sufficient. However, many authors have also considered applying Tarski's conception of truth to natural languages, where relativization to the language alone does not seem adequate. This is why the truth predicate has also been linked to additional parameters (Blackburn & Simmons, 1999).

Boisvert, in his critique of Harrison, uses an extended version of the T-schema to apply Tarski's theory of truth to examples involving natural language sentences. His extended version of the T-schema is as follows:

TN. S is true {u,t} in L if and only if p.

In this schema, S represents the name of the sentence, p its translation into the metalanguage, and L the object language for which the truth theory is defined. The most striking modification compared to the previous extended version of the T-schema is the inclusion of the parameters u and t in the truth predicate "true {u,t}". The set u represents speakers, and the set t represents times. In this way, Boisvert relativizes the truth predicate not only to the language but also to the speaker who uttered the sentence and the time it was uttered.

Soames (1999), in turn, proposes a conception in which truth is relativized to circumstances, understood either as partial states of the world or possible worlds. In this case, the truth predicate takes the form "true in C." At this point, however, I deviate too far from Tarski's original conception, so I will turn to the next topic.

4.1. Two Types of Sentence Names

In the previous chapter, I mentioned that the T-schema involves both the sentence itself and its name. This section is dedicated to the creation of sentence names. In his work *On the Concept of Truth in Formalized Languages*, Tarski presents two ways of constructing sentence names, which can then be inserted into the T-schema.

The first type of sentence name is the quotation name. A quotation name is created by enclosing the sentence in quotation marks. Thus, the name consists simply of the sentence itself, surrounded by the quotation marks. Tarski explains regarding quotation names:

"As for quotation names, they can be treated as individual words of the language, i.e., as syntactically simple expressions. The quotation marks and the expression they contain fulfill the same function as the letters and letter complexes in a word. [...] Each quotation name is thus a constant individual name of a certain expression (the expression enclosed in the quotation marks) and, in fact, a name of the same kind as a proper name of a person." (HITCHCOCK & STROIŃSKA, 2024, p. 270)

A quotation name uniquely refers to the expression it contains. For example, the sentence "snow is white" is a quotation name. Even though the internal structure of such a

name is complex, it is treated as a single word/name. Tarski compares this type of name to a proper name, such as "Albert Einstein," which has a complex internal structure but still refers to a unique individual.

In practice, the use of quotation names in the T-schema would look like this:

(5) The sentence "snow is white" is true if and only if snow is white.

The second type of sentence name is the structural descriptive name. The creation of these names is more complex, as they do not simply quote the expression but describe it in a particular way. To construct a structural descriptive name, one must use names for each letter of the language to which the sentence belongs, as well as for all other symbols used in that language (such as punctuation marks, mathematical and logical symbols, etc.). These names are not quotation names.

The construction of a structural descriptive name can be compared to assembling a puzzle: each symbol in the sentence is replaced with a name that represents that symbol. The arrangement of these names in the structural descriptive name must correspond to the arrangement of the symbols they designate in the sentence. For example, if the sentence "snow is white" were transformed into a structural descriptive name, it might look something like this:

(6) The sentence composed of the words "the," "snow," "is," and "white," where "the" is represented by the name of the symbol "t," "snow" by a sequence of symbols "s," "n," "o," etc., and so on, is true if and only if snow is white.

Both types of sentence names, quotation names and structural descriptive names, are interchangeable in the T-schema. Indeed, a quotation name can be replaced with a structural descriptive name and vice versa, depending on the context and how one wishes to formalize the sentence in the metalanguage.

4.2 Semantically Closed Languages

Before turning to Tarski's position on natural languages, it is necessary to establish some theoretical foundations by examining the class of languages that Tarski regards as incoherent. Tarski introduces the problem of natural language and the definability of the predicate "true" only indirectly, by way of his discussion of so-called semantically closed languages. Following his method, I shall begin with them.

A semantically closed language, in Tarski's sense, is one that satisfies the first two of the following three conditions:

(a) The language contains, in addition to its ordinary expressions, names of its own expressions (or an equivalent device enabling reference to its sentences), together with semantic predicates applying to those sentences.

(b) Within the language, it is possible to formulate all the sentences needed to specify the correct use of those semantic predicates.

(c) The usual logical laws hold within the language.

Tarski showed, in the context of his discussion of formal correctness, that whenever a language satisfies (a) and (b) together with (c), the Liar Paradox can be formulated within it. Consider, for instance, the paradox-generating sentence (LP):

(LP) "The sentence (LP) is false."

By virtue of condition (a), sentences in such a language can refer to themselves and can employ semantic predicates in this act of self-reference. Sentence (LP) clearly satisfies this condition, since it contains its own name and thus refers to itself. But the explicit use of its name is not necessary: the paradox would equally arise from the formulation, “This sentence is false.” In this version as well, the sentence refers to itself and applies the semantic predicate “false” to itself, thereby fulfilling the semantic requirement of condition (a).

Condition (b) requires that (LP) be treated as a legitimate instance of the use of the predicate “false.” Since the ordinary employment of semantic predicates such as “is false” is permitted in the language, (LP) satisfies this condition as well. Finally, the paradox only emerges if condition (c) is also in force, since the paradoxical structure depends on the validity of the usual logical laws—most notably, the law of the excluded middle. Without this logical framework, no contradiction would arise.

Before turning to Tarski’s own diagnosis, it is worth noting a specific feature of the Liar. The version cited above depends on the presence of the predicate “false” ($\neg\text{Tr}$). If, instead, the predicate “true” (Tr) were employed, the paradox would not arise in the same form.

According to Tarski, it is not acceptable to reject condition (c) in order to avoid the paradox, for that would amount to abandoning the fundamental laws of logic. His proposal, rather, is to deny conditions (a) and (b). A coherent language, in his view, must exclude devices that allow sentences to refer to themselves. By this restriction, Tarski eliminates all self-referential sentences—a decision for which he has been criticized. Likewise, he denies that a language can contain semantic predicates applicable to its own expressions. This prohibition follows directly from his celebrated undefinability theorem, according to which truth for a given language cannot be defined solely within that same language.

4.3 The Status of Natural Languages

In the preceding sections of this essay, we have noted that Tarski’s primary objective was to provide a rigorous definition of truth for languages with a precisely specified structure, and we outlined the rules to which such languages must conform. We have also examined semantically closed languages, for which no such definition of truth is possible. The question that now arises is how natural language is to be situated within this framework.

As established above, every semantically closed language must satisfy conditions (a)–(c). Tarski regards as incoherent any language that meets these conditions. Conditions (a) and (b) are especially relevant to the character of natural language, since one of its distinctive features is universality:

If something can be meaningfully said, it can be said in natural language.

To preserve this universality, natural language must include not only its own expressions, but also names for these expressions, together with sentences containing such names and semantic expressions such as “true sentence,” “name,” or “denotes.” Thus, in serving as the medium for all meaningful discourse, natural language necessarily acquires

the characteristics of a semantically closed system. From this, one might expect Tarski to declare natural language incoherent. Surprisingly, however, he writes:

“Our everyday language is certainly not a language with a precisely specified structure. [...] For this reason, the problem of the coherence of this language has no exact meaning. At best, one might express the opinion that a language with a precisely specified structure, resembling ordinary language as closely as possible, would be incoherent.” (TARSKI, 1956, p. 153)

Despite this caution, it is generally assumed that Tarski did in fact regard natural languages as incoherent. Yet the textual evidence suggests a more nuanced picture. As commentators have observed, one of the hallmarks of Tarski’s style was his extraordinary restraint in drawing explicit philosophical conclusions from his technical work. This methodological reserve has given rise to disagreements about the extent to which he intended his results to apply more broadly. The question of the incoherence of natural language is no exception.

Some philosophers maintain that Tarski indeed considered natural language incoherent and have gone on to dispute this conclusion. Hans Herzberger, for example, explicitly criticizes Tarski’s position in (HERZBERGER, 1967). Others, however, contend that it is unwarranted to attribute to Tarski the view that natural languages are incoherent. This interpretative stance is exemplified in the exchange between A. B. Levison (LEVISON, 1965) and Vito Sinisi (SINISI, 1967).

5. Object-Language and the Hierarchy of Metalanguages

In order to construct a satisfactory definition of truth for a given language, it is essential to introduce a fundamental distinction: that between object-language and metalanguage. The object-language is the language under investigation, for which we seek to provide a definition of truth; it is the system whose expressions and sentences are the subject of inquiry. Since we must speak about a given language, it is evident that we require another language in which to formulate statements concerning it. This second language is the metalanguage — the language in which we make assertions about the object-language and in which the definition of truth for the object-language is to be framed.

Although this distinction plays a crucial role in Tarski’s conception, he was not its originator. It derives from the tradition of the Lvov–Warsaw School, with Stanisław Leśniewski as its principal advocate. Leśniewski introduced the distinction to expose certain shortcomings in *Principia Mathematica*. As Peter Milne has observed, Tarski subsequently adopted it in his own philosophical work.

Tarski describes the metalanguage in the following way:

“The metalanguage contains both individual names and translations of all the expressions (in particular, of all the sentences) of the language under study.” (TARSKI, 1956, p. 173)

Because the metalanguage must speak about the object-language, it must be sufficiently rich to allow reference to the sentences of that language. This is achieved by incorporating into the metalanguage the resources required to construct names for every expression of the object-language. Yet this provision alone is insufficient. The metalanguage

must also contain translations of all the expressions of the object-language. These translations are indispensable for the formulation of Tarski's (T)-equivalences. Each instance of schema (T) has, on its left-hand side, the name of the sentence under study expressed in the metalanguage, and on its right-hand side, the appropriate translation of that sentence, likewise within the metalanguage.

If an expression of the object-language were to lack both a name and a corresponding translation in the metalanguage, then it would be impossible to formulate statements about it within the metalanguage, thereby preventing any systematic and complete study of the object-language.

The resources of the metalanguage are therefore not limited to names and translations of object-language expressions. This requirement is illustrated by the schema (T):

(T) X is true if and only if p

In this schema, the metalanguage must contain (i) translations of the sentences of the object-language, corresponding to p, and (ii) names for those sentences, corresponding to X. Furthermore, the metalanguage must also include logically general terms, such as the connectives and quantifiers of propositional and predicate logic—for instance, the expression “if and only if.”

Another indispensable component of the metalanguage is the presence of semantic terms that refer to expressions of the object-language. Tarski insists that such terms be introduced solely through explicit definitions. While the metalanguage may contain primitive or undefined terms, these are permissible only if they are already present in the object-language itself and must therefore be included in the metalanguage. An additional exception applies to terms that refer to the syntactic form of object-language expressions, insofar as these are used in constructing names for those very expressions.

Tarski's central point is that the semantic vocabulary employed to speak about the object-language must always be introduced by explicit definition. Any violation of this requirement risks reviving semantic paradoxes, most notably the liar paradox.

It is essential to note that the notions of object-language and metalanguage are relative. A given language, when studied within a higher-level language, functions as an object-language with respect to that higher-level language. However, if we wish to study the higher-level language itself, this requires the introduction of another metalanguage, which incorporates translations of the lower-level language as a proper part. In this context, the higher-level language becomes an object-language relative to the newly introduced metalanguage. If this new metalanguage itself becomes the subject of study, it would, in turn, function as an object-language with respect to yet another metalanguage, and so forth. The process can, in principle, continue indefinitely. Thus, we are not confronted with a single object-language and a single metalanguage, but rather with an open-ended hierarchy of metalanguages.

Nevertheless, this hierarchy is not itself Tarski's central concern. He does not posit the necessity of constructing an infinite sequence of languages, nor does he claim that the hierarchy must be finite. What matters for his purposes is the following principle: whenever we make statements about the sentences of a given object-language, we must do so in a

metalanguage. If such statements were instead formulated directly within the object-language itself, then that language would fail to meet Tarski's requirements for a precisely specified structure, for it would become a semantically closed language.

Before presenting Tarski's definition of truth, it is important to recall a principle of decisive significance: Tarski's theorem of the undefinability of truth. This result, derived from Theorem I in *On the Concept of Truth in Formalized Languages*, can be summarized as follows:

Under certain general conditions, it is impossible to construct a correct definition of truth using only the resources of the language under investigation.

In other words, it is not possible to provide an adequate definition of truth for an object-language if we confine ourselves to the categories available within that very language. For this reason, an adequate definition of "true" must be formulated in the metalanguage, which must itself contain semantic concepts applicable to the expressions (primarily sentences) of the object-language.

6. Construction of the Definition of Truth

The central objective of this essay is to examine Tarski's rigorous definition of truth, which is constructed on the semantic notion of satisfaction. At first glance, the reliance on a semantic concept may seem problematic, since the definition must be framed in a metalanguage free of primitive semantic terms. Tarski's strategy, however, is to define satisfaction in purely non-semantic terms, a move that secures the internal coherence of his system. As Milne points out, such a reduction purges semantic concepts of paradox, grounds them in clear notions, and brings them into alignment with the demands of physicalism. (Milne, 1997, p. §2)

Satisfaction, in Tarski's account, is best understood as a relation between linguistic expressions and objects. Sentential functions—open sentences containing free variables—are not truth-evaluable until their variables are replaced by names of objects, at which point they become proper sentences with truth-values. Because there are infinitely many such functions, Tarski employs the method of recursive definition: elementary cases are defined directly, and systematic rules are then given for generating more complex ones. In this framework, a sentence is true if and only if it is satisfied by all infinite sequences of objects. This formulation provides both rigor and generality, offering a precise way to ground the notion of truth.

The adequacy of Tarski's proposal, however, has been subject to extensive debate. (Kirkham, 1992) outlines five criteria for evaluating solutions to semantic paradoxes: (1) specificity in identifying the source of paradox, which Tarski locates in semantically closed languages—a claim contested by (MARTIN, 1967); (2) the absence of ad hoc postulates, challenged by critics of the hierarchical structure of metalanguages, particularly Kripke (1975); (3) avoidance of excessive restriction, since Tarski excludes all self-referential sentences, a move seen by Martin as overly severe; (4) completeness, given that it remains unclear whether the hierarchy addresses all forms of self-reference; and (5) conformity to intuition, where critics such as Harrison (1998) and Kripke argue that Tarski's system

departs from the ordinary concept of truth by multiplying distinct truth predicates across languages.

Taken together, these points illustrate both the strength and the limits of Tarski's approach. His recursive account of satisfaction and his metalanguage hierarchy establish a powerful method for defining truth within formal systems, yet they do so at the price of excluding natural language and introducing theoretical structures that many philosophers find counterintuitive. Nevertheless, his work remains a decisive turning point in the modern study of truth, combining logical precision with philosophical depth, and continuing to shape debates about the very possibility of a coherent definition of truth.

7. Conclusion

In this paper, we have examined Tarski's conception of truth through both its logical foundations and its semantic implications. By analyzing his formal definition, we highlighted how his approach aimed to resolve the paradoxes arising from natural languages and to establish truth as a mathematically rigorous notion. The central idea — that truth can only be consistently defined within a formally structured object language, relative to a meta-language — remains one of the most influential contributions to twentieth-century logic and philosophy of language. Furthermore, we have shown that the semantic stakes of Tarski's theory extend beyond technical formalism, for they touch upon broader debates concerning the adequacy of formal methods to capture intuitive notions of meaning and reference. Ultimately, Tarski's work demonstrates that truth, far from being an ordinary linguistic predicate, requires a precise logical framework in order to avoid inconsistency. This underscores the enduring relevance of his project, both for contemporary logic and for ongoing discussions in the philosophy of science and language.

Bibliographie

- Ayer, A. (1971). *Language, Truth and Logic*. Harmondsworth: Penguin Books.
- Blackburn, S., & Simmons, K. (1999). *Truth*. Oxford : Oxford University Press.
- BOISVERT, D. (1999). The Trouble with Harrison's 'The Trouble with Tarski'. *The Philosophical Quarterly*, 376-383.
- COFFA, A. (1987). Carnap, Tarski and the search for truth. *Noûs*, 547-572.
- HERZBERGER, H. (1967). The truth-conditional consistency of natural languages. *The Journal of Philosophy*, 29-35.
- HITCHCOCK, D., & STROIŃSKA, M. (2024). *Alfred Tarski on Scientific Semantics*. Springer Nature Switzerland.
- Kirkham, R. (1992). *Theories of Truth*. Bradford Books.
- KRIPKE, S. (1975). Outline of a theory of truth. *The journal of philosophy*, 690-716.
- LEVISON, A. (1965). Logic, language, and consistency in Tarski's theory of truth. *Philosophy and Phenomenological Research*, 384-392.
- MARTIN, R. (1967). Toward a solution to the liar paradox. *The Philosophical Review*, 279-311.
- Milne, P. (1997). Tarski on truth and its definition.. *Filosofia*. pp. 198-210. In C. Timothy, P. Kolft, & V. Svoboda, *Logica '96: Proceedings of the 10th International Symposium*.
- SINISI, V. (1967). Tarski on the inconsistency of colloquial language. *Philosophy and Phenomenological Research*, 537-541.
- SOAMES, S. (1999). *Understanding truth*. Oxford: Oxford University Press.
- Tarski, A. (1944). The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, 341-376.
- TARSKI, A. (1956). The concept of truth in formalized languages. Dans A. TARSKI, *Logic, Semantics, Metamathematics*. Oxford, New York:: Oxford University Press.
- TARSKI, A. (1969). Truth and proof. *Scientific American*, 63-77.
- TARSKI, A. (2021). The semantic conception of truth. Dans M. Lynch, *The Nature of Truth Classic and Contemporary Perspectives* (pp. 331-363). The MIT Press.