Could Tajtelbaum Question What Tarski Could Not?

Jan Wiślicki*

Received: 29 October 2019 / Accepted: 4 May 2020

Abstract: The paper discusses Tarski’s approach to quotation. It starts from showing that it is vulnerable to semantic inconsistencies connected with what is known as Reach’s puzzle, formulated in 1938 by a Czech logician Karel Reach. This fact gives rise to serious problems concerning the relation between the metalanguage and an object language. Moreover, the paper touches upon a historic aspect, pointing out that the problem at hand is discussed in the only paper signed up as Al. Tajtelbaum, i.e. Alfred Tarski’s original name. It argues that the puzzle reveals the importance of reopening the discussion on the understanding and limitations of deriving the metalanguage from an object language.

Keywords: Enquotation; metalanguage; quotation; Reach’s puzzle; Tajtelbaum; Tarski.

1. Introduction

The present squib delivers arguments for reopening the discussion on Tarski’s approach to metalanguage. There are two crucial issues underlying the present discussion. First, Tarski’s view on quotation is more often than not taken as unproblematic, also for Tarski himself; I show that these
assumptions are not justified. Second, reconsidering this problem gives rise to serious questions concerning the present understanding of metalanguage.

The paper is organised as follows. In Section 2 I present the problem arising for the functional approach to quotation (Tarski 1933/1983). In Section 3 I show that not only the crucial problem had been formulated in the late thirties (Reach 1938) and became known as Reach’s puzzle, but also that it was acknowledged in Tarski’s seminal paper (1933/1983). In Section 4 I discuss a note signed by Tarski’s original name, Al. Tajtelbaum (Tajtelbaum 1957), that has remained nearly unnoticed for the last 60 years. I contrast this paper with Reach’s puzzle and the related discussion, showing that Tarski might have been less certain about the nature of quotation than it is standardly assumed. In Section 5 I show why these facts are problematic for Tarskian semantics and what is the general lesson following from them. Section 6 summarizes the discussion and suggests a path for future work.

2. Quotation-function: a problem

Let us first have a look at the problem underlying the whole discussion. The puzzle arises for the functional operation of enquotation yielding quotational names. Tarski (1933/1983) defines quotation-function as follows:

The expression “‘p’” [...] must be regarded as a function, the argument of which is a sentential variable and the values of which are constant quotation-mark names of sentences. We shall call such functions quotation-functions. The quotation marks then become independent words belonging to the domain of semantics, approximating in their meaning to the word ‘name,’ and from the syntactical point of view they play the part of functors. (Tarski 1933/1983, 161)

This fragment shows that Tarski, at least in his seminal work, did not simply subscribe to the atomic treatment of quotational expressions. A toy formalisation of the above definition results in a trivial formula of the form:

\[(1) \quad Q(x) = 'x'\]
where \( x \) ranges over expressions of the object language (here sentences), \( Q \) stands for the operation of enquotation (most standardly marked by quotes), and ‘\( x' \) for its value being a quotational name of \( x \).

Assuming, as standardly, that a function is the meaning of the corresponding functor, I take the formula in (1) to be equivalent to the following interpretation:

(2) The meaning of ‘\( Q \)’ maps the meaning of ‘\( x \)’ onto the meaning of ‘\( \text{"}x\text{"} \).’

Note that, though not unproblematic, the equivalence of (1) and (2) is coherent with Tarski’s view expressed elsewhere:

We may, admittedly, replace this formula \([3 = 2 + 1]\) by a sentence which expresses the same idea but is about symbols, namely, by a sentence which asserts that the symbols “3” and “2 + 1” designate the same number. (Tarski 1994, 55)

So, in order to obtain the quotational name ‘\( \text{dog} \)’ picking out the noun \( \text{dog} \), the computational mechanism makes the substitution \([\text{dog}/x]\) yielding (3):

(3) The meaning of ‘\( Q \)’ maps the meaning of ‘\( \text{dog} \)’ onto the meaning of ‘\( \text{"} \text{dog} \text{"} \).’

Though at first sight innocent, (2) is in fact quite problematic. To see this, let us make the following substitution \([\text{the first word of this paper}/x]\), perfectly fine according to (1):

(4) The meaning of ‘\( Q \)’ maps the meaning of ‘\( \text{the first word of this paper} \)’ onto the meaning of ‘\( \text{the first word of this paper} \).’

A quick look shows that (4) is untenable; the line of reasoning, assuming the referential approach to meaning enriched by the minimum context securing the computation of indexicals, is very simple:

(5) i. The meaning of the expression \( \text{the first word of this paper} \) is the word \( \text{could} \).

ii. Thus, the expression \( \text{the meaning of ‘the first word of this paper’} \) can be substituted \( \text{salva veritate} \) by the expression \( \text{the word ‘could’} \).
iii. The result of the substitution is:

The meaning of ‘Q’ maps the word ‘could’ onto the meaning of ‘the first word of this paper’.

iv. This means that the quotational name of the expression the word ‘could’ is the quotational expression ‘the first word of this paper.’

However, the quotational name of the expression the word ‘could’ is ‘the word ‘could’,’ not ‘the first word of this paper.’ Thus iv. is contrary to the expected effect. It follows, then, that the definition of enquoteation as formulated in (1) is untenable.

So, the straightforward Tarskian implementation of quotation-function gives rise to serious semantic problems. In the next section I show that the problem is not new, going back at least to Tarski (1933/1983).

3. Quotation-function and its domain

The puzzle underlying the effect laid out in (5) was first tackled by a Czech logician Karel Reach (Reach 1938). Reach’s line of reasoning goes as follows. Given an object and its name, it is impossible to express the name relation holding between the two. The reason is that whenever one formulates a sentence meant to express this relation, one uses not the name but the quotational name of this name. Because any name (including quotation) and its bearer are two distinct entities, the name relation cannot be informatively expressed.

1 Without going into a detailed ontological discussion, I take an abstract word to be the meaning of its quotational name, and the particular token of this word to be what satisfies the meaning at hand. Note also that the problem laid out in (5) cannot be resolved by making further assumptions connected with the ontology of language. For instance, one might, as suggested by an anonymous reviewer, assume that the meaning of the particular expression is not a word but a token. Still, this would not resolve the problem. The obtained effect would be that the quotational name of the expression the token ‘could’ is the quotational expression ‘the first word of this paper,’ contrary to expectations. Although the first word of this paper may seem to be an appropriate way to refer to the token could, it is not its quotational name.
To illustrate, suppose that someone asks about another person’s name. The answer is something like (6):

(6) His name is *Alfred*.

Clearly, when uttering (6) the speaker is not using the name, contrary to (7):

(7) Alfred is reading a book.

Rather, what he is using is the quotational name of his name. What obscures this fact is that it is the special property of quotational names that their bearers—the units of the object language—can be immediately grasped from the form of the quotational name. Thus, the word *Alfred*—the bearer of the quotational name ‘*Alfred*’—can be immediately grasped on the basis of the form of its quotational name ‘*Alfred*’ (cf. Read 1997, Yourgrau 1982).

While Reach does not explicitly address the problem of quotation, his observations share one important point with the discussion in Section 2. What is crucial for Reach’s analysis is that there is a problem with setting up the relation between a name and its bearer. Put more formally, no function standing for the name relation can be defined by letting it take expressions of an object language as arguments (Anscombe 1965; Geach 1980; Mendelsohn 2005; Gaskin, Hill 2013; a.o.).

Indeed, it is exactly this fact, i.e. letting the domain of the quotational name relation be a set of expressions of an object language as in (1), that lies at the heart of the problem presented in (5). The problem arises because *Q* is assumed to be extensional and thus to take as arguments expressions that undergo substitution *salva veritate* for expressions of a different form but with an equivalent meaning. And what is especially interesting from the historic point of view is that Tarski was aware of that; immediately after the passage quoted is Section 2 he continues:

But then new complications arise. The sense of the quotation-function and of the quotation marks themselves is not sufficiently clear. In any case such functors are not extensional [...] a deeper analysis shows it to be impossible to give any precise meaning of such functors. (Tarski 1933/1983, 161)
Moreover, in Footnote 2 on the same page he suggests that name-forming functors should be considered as something distinct from sentence-forming functors. However, no further comments on these problems are provided, despite the fact that they are by no means marginal for Tarski’s concept of metalanguage. But what had just been touched upon by Tarski was discussed in a more detailed way by Tajtelbaum. In the next section I take a closer look at Tajtelbaum’s paper and try to find a more far-reaching explanation of the effect at hand.

4. Tajtelbaum’s doubts

What is now known as Reach’s puzzle would have probably passed unnoticed had Anscombe (1956, 1965) not raised it in the context of her discussion on Wittgenstein’s Tractatus. Anscombe (1956) posed the problem taken from Reach’s observations in the form of competition. The winning response appeared as Tajtelbaum (1957)—a paper signed by Tarski’s original name.\(^2\) The offered solution, however, is by no means coherent with the functional approach presented in Tarski (1933/1983). The core of Tajtelbaum’s approach rests upon the following idea:

we have the (tacit) convention that a name and its name are denoted by the same word, and so the name of a name “tells” us the name. (Tajtelbaum 1957, 53)

Despite the lack of unequivocal evidence, the paper has all the hallmarks of being authored by Tarski (Sundholm 1993). This fact is not historically unmotivated. It is well known that Tarski himself was rather reluctant to express his philosophical views in articles (Sundholm 1993; Murawski 2011), sometimes deliberately (Tarski 1930/1983). Viewed from that angle the fact that the paper was published under the name Tajtelbaum is not that

\(^2\) The paper is not mentioned in (Burdman Feferman, Feferman 2004) who explicitly state that ‘after that [1924] all his papers were published under the name Alfred Tarski because, shortly before receiving his Ph.D. degree in 1924, he officially changed his surname and concomitantly converted to Catholicism’ (Burdman Feferman, Feferman 2004, 37–38).
surprising. Notwithstanding these historical questions, a more pressing task is to determine how Tajtelbaum’s account differs from Tarski’s functional approach. Note that Tarski took quotational names to be the values of the function defined on names (expressions of the object language). By contrast, Tajtelbaum’s idea is that a name and its quotational name are expressed by one and the same expression, which means that there is no mapping as in (1).

As it stands, Tajtelbaum’s idea invites further questions. First of all, it cannot be the case that a name and its name are one and the same. And this, of course, is not Tajtelbaum’s claim; he clearly states that the two are denoted by the same word, not that they are the same object. Still, the author does not make explicit how the proper meaning—the name or its quotational name—denoted by a given word can be discriminated within a formal semantic system.

The idea sketched by Tajtelbaum has been given a new life in the form of so-called identity theory of quotation (aka use theory of quotation) developed by Geach (1980), Washington (1992), Saka (1998), a.o. (see also Saka 2013; Maier 2014 for some critical overviews). The idea is that every expression is primarily ambiguous. Accordingly, the meaning of the expression $E$ is the set $\{e_1, ..., e_n\}$, where $e_i: 1 \leq i \leq n$ stands for the extension, intension, lexical entry or the phonic/written form. It is the role of quotes to (partially) disambiguate the expression. When flanked by quotes, $E$ is interpreted as denoting something else than its extension, most standardly (though not exclusively) a lexical entry or a phonic/written form. Put more formally, every use of an expression $E$ in the context $C$ maps this pair onto an element of the power set of $E$. Thus $use: (E, C) \rightarrow P(E)$. If the particular utterance is unambiguous, the obtained value is a singleton.

If, then, it is the role of quotes (the quotation functor $Q$) to disambiguate the use of expressions, the situation roughly looks as follows:

\[
\text{use}(Q(E), C) = \{e_i, ..., e_k\}, \text{ where: }
\{e_i, ..., e_k\} \text{ is an element of } P(E) \& i \leq j \leq k \& j \neq \text{the extension of } E
\]

Within this approach Alfred in both (6) and (7), repeated below, is a primarily ambiguous single expression:
(6) His name is *Alfred*.

(7) Alfred is reading a book.

It is the context of its use that is responsible for disambiguation. In (7) the *use* function maps it onto its extension, perhaps as a default value chosen due to the lack of other markers (e.g. modal operators or a marker of quotation). By contrast, in (6) the operation of enquoteation (here marked by the italic) results in the *use* function yielding the expression of the form *Alfred*.

The picture sketched in (8) is different from the one in (1). Here quotation-function is not defined on expressions of an object language, as was crucial for Tarski (1933/1983). In this regard it is safe from the threat exposed in (5). Instead, it takes as its argument an underspecified linguistic object $E$, a sort of root in the sense of Marantz (1997; 2000), as investigated within the generative inquiry. When additionally mapped by quotes, $E$ denotes the word at hand, so that *the name of a name “tells” us the name*, as put by Tajtelbaum in the passage quoted above. Accordingly, an expression of the metalanguage is an output of operation defined on an underspecified item, rather than on an expression of object language.

So, Tajtelbaum’s approach to quotation is orthogonal to the account of quotation provided within Tarski’s truth theory. This fact, when viewed in the context of Reach’s puzzle, deserves a more careful treatment. As concluded by Tajtelbaum (1957, 53), the problem *draws attention to a limitation of formal analysis*. What the discussion taken up by Tajtelbaum suggests is that there are serious doubts whether defining the metalanguage does not extend those limitations. Put differently, there are doubts whether the relation between an object language and the metalanguage can be expressed in an informative and formal way. In this sense they are more severe than those cited from Tarski (1933/1983) in Section 3. In the next section I will discuss some issues following from these doubts.

5. The emergence of the metalanguage: reactivation

The above concern raised by Tajtelbaum with respect to the limitation of formal analysis unearthed a deeper problem than just a formal aspect of
a theory of quotation. It shows that the present understanding of how the metalanguage is derived from an object language might be much weaker than it has standardly been assumed.

It might be apt at this point to recall a different yet closely connected problem raised by Tarski and tackled by Soames (1999). Assume the following definition of truth (cf. Tarski 1933/1983, 159–60):

\[(9) \text{ for every } t (\text{‘}t\text{’ is true iff } t)\]

The puzzle is that the substitution of \(t\) in (9) by any sentence cannot be blind. In particular, it cannot target the letter \(t\) occurring in the word \textit{true}. The approach proposed by Soames (1999) is based on the fact that (9) works perfectly fine if we are able to distinguish a set of substitutional variables that we wish to use to quantify into such constructions (Soames 1999, 88). Thanks to this distinction, \textit{true} is somehow frozen and thus protected from the unwanted substitution.

Note, however, that that kind of solution works insofar as the relevant distinction is possible. In the case presented above, the distinction separates variables from constants, being thus hardly questionable. But the problem given in (5) looks different, mainly because it does not boil down to the logical distinction as the one above. The problem in (5) arises if what is flanked by quotes undergoes substitution \textit{salva veritate}. Thus for (5) not to arise, the material within quotes must have been somehow frozen, on a par with \(t\) in \textit{true} being frozen for the unwanted substitution. Still, regardless of the reasoning behind this effect, it would mean that freezing is absolute, targeting equally each expression flanked by quotes. In this sense whatever is flanked by quotes (i.e. taken by the quotation-function as an argument) would be treated as a purely material string. But if such a blindly overall freezing is assumed, a reasonable question arises whether within this approach the domain of quotation-function is not shifted from expressions of object language into material strings they are represented by.

If this reasoning is on the right track, then Tajtelbaum’s doubts unearth an important problem, pointed out by Dummett:

The presence in our language of various meaning-theoretic terms forces us, as we saw, to \textit{impose} on it a distinction between object-language and metalanguage which is not there in reality. And we
shall want to draw the line so as to put into the metalanguage only those terms, and those uses of such terms, which really do serve the purpose of expressing some imperfectly formed ideas we have about how our language functions—or, to put it differently, which could be understood only as having a place in a meaning-theory for the rest of the language. Now, if one of these terms, considered as subject to a certain type of characterisation, would not play any useful role in such a meaning-theory, it is either useless or belongs (in so far as it is so characterised) on the other side of the line, to what we ought to take as constituting the object-language. And that is how we ought to view the term ‘true,’ considered as characterised either directly by the requirement that each instance of the (T) schema holds, or by a Tarskian truth-definition to which the fact that the metalanguage is an expansion of the object-language is taken as essential. (Dummett 1991, 71–72)

That is, leaving aside the exact way one formalises the metalanguage, it is crucial for Tarski’s semantics that the metalanguage is defined as being derived from an object language. In a similar vein Simchen (2003) points out that the a priori character of T-sentences, crucial for Tarski’s theory, is secured insofar as the metalanguage extends an object language. Of course, a functional account as the one in (1) is not the only way of defining this relation, and neither Dummett nor Simchen claims so. As pointed out by Shapiro (1997, 49), the minimum condition is that of metalanguage being a faithful representation of the object language. Still, the way one is related to the other must be established. If, on the other hand, this extends the limitation of formal analysis, then the problem is quite severe. This limitation was suggested in Tajtelbaum’s note and can be inferred from (8). According to this approach, an object language and the metalanguage are just

---

3 The quote from Dummett is important for yet another reason. In his seminal paper Tarski (1933/1983) proposes a number of other approaches to quotation, many of which have been developed in the literature (see Cappelen & Lepore 2007; Saka 2013; Maier 2014 for some relevant surveys). Nevertheless, as pointed out by Dummett, it is the functional approach, according to which the metalanguage is derived from the object language, which is crucial for Tarskian semantics.

*Organon* F 27 (4) 2020: 480–493
two different values of the *use* function defined on underspecified linguistic objects. Still, the function does not relate expressions of an object language with their metalinguistic names. Put differently, we know how to obtain expressions of an object language from some underspecified objects. We also know how to obtain expressions of the metalanguage from such underspecified objects. Perhaps we can even stipulate criterions of faithfulness mentioned by Shapiro.\(^4\) However, what remains beyond the present understanding is whether it is possible to set up (and, if yes, how) a relation securing a systematic, extensional connection between the two. In this sense while the truth theory itself is secured by the fact that we can derive an object language and the metalanguage, the understanding of truth theory is weaker to the extent the relation between the two extends the limitation of formal analysis, and thus a precise understanding.

6. Conclusion and future prospects

Tajtelbaum’s idea, developed as the identity theory of quotation, is not the only way for circumventing\(^5\) the unwanted effects discussed in Section 2. Perhaps Davidson’s (1979) demonstrative theory is the most established one in the current literature (see Reimer 1996; Predelli 2008; Maier 2014 for some relevant comments). Still, the theory has hardly been discussed from the point of view of Tarski’s semantics in the context of the abovementioned comments provided by Dummett (1991) and Simchen (2003). And the problem is relevant, bearing in mind that various authors (Goldstein 1985; Kamp 1995) pointed out that under this theory the operation of enquotation is defined not on words of object language, but on their inscriptions. This,

\(^4\) It is also not at all unproblematic whether *faithfulness* is the right notion. Various linguistic tests show that quotation is a vague, scalar (*Jones said something very similar to the Japanese ‘Kanpai’*) and highly context-dependent object (De Brabant 2017). In this regard linguistic data suggest that it is conditions of a successful demonstration, rather than faithfulness, that should be investigated in this context. Accordingly, the problem of how successful demonstration relates to the problem of the a priori character of T-sentences is pressing and challenging (Simchen 2003).

\(^5\) Some of them mentioned in Tarski (1933/1983, 162).
of course, provides a solution to the puzzle laid out in (5). To see this, let capital letters stand for inscriptions of expressions and consider the following line of reasoning:

(10) i. The inscription of the expression *the first word of this paper* is THE FIRST WORD OF THIS PAPER.

ii. The meaning of ‘THE FIRST WORD OF THIS PAPER’ is an object of the form of THE FIRST WORD OF THIS PAPER.

iii. Quotation-function takes inscriptions as arguments. Thus, the following is true:

The meaning of ‘Q’ maps the meaning of ‘THE FIRST WORD OF THIS PAPER’ onto the meaning of ‘‘THE FIRST WORD OF THIS PAPER’.’

iv. Given ii., the meaning of THE FIRST WORD OF THIS PAPER cannot by substituted *salva veritate* by the word ‘could.’

Given that the quotation-function maps inscriptions of expressions, not expressions as such, onto quotational names, it is not vulnerable to the puzzle laid out in (5). However, since inscriptions and expressions are two different classes of objects, quotation-function as described in (10) no longer relates an object language with the metalanguage. This, in turn, suggests that problems arising for Tarski’s semantics as those discussed above in the context of comments given by Dummett and Simchen are not just a matter of the particular approach to quotation, but rather of the nature of quotation as such.

The lesson from these observations is that building semantics upon the metalanguage as functionally derived from an object language is misguided. It is not a matter of the way linguistic expressions are mapped onto expressions of the metalanguage. Rather, it is a matter of how expressions are evaluated, so that they can be interpreted as extensional, metalinguistic or other expressions. In this sense the quotes seem to be part of modality, i.e. part of grammar securing various evaluations of particular expressions, rather than of functions mapping one object onto another. The present squib was aimed at showing that observations underlying this approach have strictly logical motivations, whose importance did not remain unnoticed by Tarski.
Acknowledgments

The present paper benefited much from discussions with Andrzej Bogusławski and Richard Gaskin, as well as two anonymous reviewers of Organon F. I am solely responsible for all errors and shortcomings.

References


*Organon F* 27 (4) 2020: 480–493