1. Tarski

“The present article is almost wholly devoted to a single problem – the definition of truth. Its task is to construct – with reference to a given language – a materially adequate and formally correct definition of the term ‘true sentence’.” (“The Concept of Truth in Formalized Languages” (1933), 152.)

A. Tarski on truth, the liar, metalanguage and object language

“The attempt to set up a structural definition of the term ‘true sentence’ – applicable to colloquial language is confronted with insuperable difficulties. The breakdown of all previous attempts leads us to suppose that there is not satisfactory way of solving our problem. Important arguments of a general nature can in fact be invoked in support of this supposition...” (“The Concept of Truth in Formalized Languages,” 164.)

“A characteristic feature of colloquial language ... is its universality. It would not be in harmony with the spirit of this language if in some other word a language occurred which could not be translated into it... But it is presumably just this universality of everyday language which is the primary source of all semantical antinomies ... These antinomies seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent.” (“The Concept of Truth in Formalized Languages,” 164-5.)

“If these observations are correct, then the very possibility of a consistent use of the expression ‘true sentence’ which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression.” (“The Concept of Truth in Formalized Languages,” 165.)

“Accordingly, we decide not to use any language which is semantically closed in the sense given.” (“The Semantic Conception of Truth” (1944), 349.)

“... we have to use two different languages in discussing the problem of the definition of truth... The first of these is the language which is ‘talked about’ and which is the subject-matter of the whole discussion; the definition of truth we are seeking applies to the sentences of this language. The second is the language in which we ‘talk about’ the first language, and in terms of which we wish, in particular, to construct the definition of truth for the first language. We shall refer to the first language as ‘the object-language,’ and to the second as ‘the meta-language.’” (“The Semantic Conception of Truth,” 349-350.)
B. The thesis of the essential richness of the metalanguage

“It should be noticed that these terms ‘object-language’ and ‘meta-language’ have only a relative sense. ... In this way we arrive at a whole hierarchy of languages.” ("The Semantic Conception of Truth,” 350.)

“The solution turns out to be sometimes positive, sometimes negative. This depends upon some formal relations between the object-language and its meta-language; or, more specifically, upon the fact whether the meta-language in its logical part is ‘essentially richer’ than the object-language. It is not easy to give a general and precise definition of this notion of ‘essential richness.’ If we restrict ourselves to languages based on the logical theory of types, the conditions for the meta-language to be ‘essentially richer’ than the object-language is that it contain variables of a higher logical type than those of the object-language.” “It should be noticed that these terms ‘object-language’ and ‘meta-language’ have only a relative sense. ... In this way we arrive at a whole hierarchy of languages.

If the condition of ‘essential richness’ is not satisfied, it can usually be shown that an interpretation of the meta-language in the object-language is possible; that is to say, with any given term of the meta-language a well-determined term of the object-language can be correlated in such a way that the assertible sentences of the one language turn out to be correlated with the assertible sentences of the other. As a result of this interpretation, the hypothesis that a satisfactory definition of truth has been formulated in the meta-language turns out to imply the possibility of reconstructing in that language the antinomy of the liar; and this in turn forces us to reject the hypothesis in question.

Thus we see that the condition of ‘essential richness’ is necessary ... If we want to develop the theory of truth in a meta-language which does not satisfy this condition, ... We have then to ... express the fundamental properties of the notion of truth in a series of axioms. ... This procedure can be avoided. For the condition of the ‘essential richness’ of the meta-language proves to be, not only necessary, but also sufficient...” ("The Semantic Conception of Truth,” 351-2.)

“To define recursively the notion of satisfaction, we have to apply a certain form of recursive definition which is not admitted in the object-language. Hence the ‘essential richness’ of the meta-language may simply consist in admitting this type of definition. On the other hand, a general method is known which makes it possible to eliminate all recursive definitions and to replace them with normal, non-recursive ones. If we try to apply this method to the definition of satisfaction, we see that we have either to introduce into the meta-language variables of a higher logical type than those which occur in the object-language; or else to assume axiomatically in the meta-language the existence of classes that are more comprehensive that all those whose existence can be established in the object-language.” (“The Semantic Conception of Truth,” 372, fn. 16.)

“If the system \(L\) is based upon the theory of types, this enlargement will consist, in the first instance, in an enrichment of its linguistic forms by the introduction of variables of a higher level... The postulates and the rules of inference must be adapted to the enlarged resources of the
linguistic forms, but their content, roughly speaking, does not change. If the system \( L \) is constructed after the manner of the system of Zermelo, neither the linguistic forms not the rules of inference are changed, but the change consists entirely in strengthening the set of postulates. The essential point here is the introduction of a new postulate, the content of which may be described as follows: There is a set \( X \) the elements of which provide a model for the system of postulates of the original system \( L \).” (“On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth” (1939), 110.)

C. Convention T

**CONVENTION T.** A formally correct definition of the symbol ‘\( Tr \)', formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:

(a) all sentences which are obtained from the expression ‘\( x \in Tr \) if and only if \( p \)' by substituting for the symbol ‘\( x \)' a structural-descriptive name of any sentence of the language in question and for the symbol ‘\( p \)' the expression which forms the translation of this sentence into the metalanguage;

(b) the sentence ‘for any \( x \), if \( x \in Tr \) then \( x \in S \)' (in other words, ‘\( Tr \subseteq S \)'.” (“The Concept of Truth in Formalized Languages,” 187-88.)

2. Devedi and Solomon’s critique

“Our contention, put bluntly, is that this claim is false, unless ‘essentially richer’ means nothing else than ‘sufficient to contain a truth-definition for the object-language’; that is, the claim is false unless it is merely empty. Somewhat less bluntly, we know by Tarski’s theorem that there must be something differentiating the metalanguage from the object language if the former contains a truth definition for the latter. Our contention is that the only property which could be a necessary and sufficient such something in general would be wildly disjunctive, hence not a natural property that might make Tarski’s claim informative, and that the content of Tarski’s claim therefore reduces at best to a mere reformulation of Tarski’s theorem.” (“Tarski on ‘Essentially Richer Metalanguages,” 1-2.)

“Tarski was mistaken about, or at least insufficiently cautious in claiming, the need for a hierarchy.” (“Tarski on ‘Essentially Richer Metalanguages,” 2.)

A. Devedi and Solomon’s attempt at providing an argument for Tarski

“A translation of \( L \) into \( L' \) must be a map of the well-formed expressions of \( L \) into the well-formed expressions of \( L' \) such that the translation of the conjunction of two formulas must be the conjunction of their translations; the translation of the negation of a formula must be the negation of its translation; the translation of a quantified formula must be a quantified formula; and the translation of a theorem of \( L \) must be a theorem of \( L' \).” (“Tarski on ‘Essentially Richer Metalanguages,” 9.)
“L₁ is stronger than L₂ if there is a translation of L₂ into L₁ but not conversely, while they are of the same strength if each is translatable into the other.” (“Tarski on ‘Essentially Richer Metalanguages,”’ 9.)

“... ‘usual’ cases will need to be ones in which ... the translation of a translation of a sentence of ML must be provably (in ML) equivalent to the original sentence (and of course, similarly, mutatis mutandis, for a sentence of OL). As a convenient name tag for this sort of case, let’s say that in such a situation t is a strong translation of ML into OL.” (“Tarski on ‘Essentially Richer Metalanguages,”’ 10.)

Their idea, though not stated explicitly, then seems to be to try the suggestion:

L₁ is essentially richer than L₂ if there is a strong translation of L₂ into L₁ but not conversely, while they are equally rich if each is strongly translatable into the other.

They then consider an argument which attempts to show that if ML and OL are equally rich, and ML contains a truth-definition for OL, then ML will contain a truth-definition for itself. They begin (I use ‘[’ and ‘]’ for their corner-quotes):

“Suppose we have strong translations t of ML into OL and s of OL into ML, in particular that if σ is a sentence of ML,

(*) \[ \vdash_{\text{ML}} \sigma \leftrightarrow \vdash_{\text{OL}} t(\sigma) \]

holds. Suppose also that we have a truth-definition in ML for OL under the translation s, i.e., for some one-place predicate Tr of ML.

\[ \vdash_{\text{ML}} \text{Tr}([s(\psi)]) \rightarrow s(\psi) \]

for every sentence ψ of OL.” (“Tarski on ‘Essentially Richer Metalanguages,”’ 11.)

This is confused; what they want is rather:

\[ \vdash_{\text{ML}} \text{Tr}([\psi]) \rightarrow s(\psi) \]

Adjusting for this, and correcting another slip, their argument proceeds:

“...for every sentence φ of ML ... we have

\[ \vdash_{\text{ML}} \text{Tr}([t(\phi)]) \rightarrow s(t(\phi)). \]

... In general, though, we only have that \[ \vdash_{\text{ML}} \text{Tr}([t(\phi)]) \rightarrow \phi \] In order to be able to conclude on that basis that ML will have a truth-definition for itself, we would need to appeal to the existence of a
predicate \( \text{Tr}' \) in ML such that

\[
\text{Tr}'([\varphi]) \equiv \text{Tr}([t(\varphi)]).
\]

We have no guarantee that such a predicate is available, and its existence would depend on something like the formalizability of the function \( t \) in ML.” (“Tarski on ‘Essentially Richer Metalanguages,” 10-11, altered so as to make sense.)

**B. Examples – extended vocabulary**

Devidi and Solomon go on to consider other attempts to explain “essential richness,” along the way considering some examples of cases in which ML contains a truth-definition for OL. The first set of examples are meant to show that “essentially richer” cannot mean “can prove the consistency of OL.” These include examples in ML can define truth for OL but cannot prove the consistency of OL, in various senses of “prove the consistency of OL.” One example here is that of the set theories ZF (as OL) and GB (as ML). GB can define truth for ZF, but cannot prove consistency of ZF. The key point for us here is that in this case, ML extends OL by adding vocabulary:

“...we get GB from ZF by moving from a one-sorted to a two-sorted language by adding large variables \( X, Y, Z, \ldots \). These are intended to range over the classes defined by the predicates of the original theory...” (“Tarski on ‘Essentially Richer Metalanguages,” 12.)

In this case ML is richer in modes of expression, but not in theorems.

**C. Examples – new theorems**

“The predicative conservative extensions we have described make an interesting contrast with Tarski’s other sort of example in which we extend OL to ML, not by adding new well-formed expressions, but by adding new theorems.” (“Tarski on ‘Essentially Richer Metalanguages,” 20.)

The example here is that of the set theories Z (as OL) and ZF (as ML). The definition of truth here goes through a translation in which the quantifiers of Z are mapped into bounded quantifiers of ZF.

In this case ML is richer in theorems, but not in modes of expression.

**D. Their interim conclusion**

“Summarizing and simplifying a bit, then, we might say that while the fact that a conservative extension can contain a correct truth definition shows that we cannot simply mean stronger by ‘essentially richer’, the present case shows that we cannot mean richer in modes of expression, either. ... the vagueness of the term ‘mode of expression’ leaves open the possibility of a rescue
attempt on which enumerating these modes involves not just the stock of well-formed expressions, but consideration of what can be proved in the language as well. ... it’s not easy to see how this revised version is going to avoid collapsing into merely being the disjunctive property of either containing more well-formed expressions or being stronger. But if this is what ‘essential richness’ boils down to the necessity half of Tarski’s claim is merely a restatement of a trivial consequence of Tarski’s theorem.” (“Tarski on ‘Essentially Richer Metalanguages,” 21.)

E. Examples (3): languages which contain definitions of truth for each other (Wang)

A disjunctive account would leave it open that two languages might contain truth definitions for one another, if the first had more modes of expression than the second, while the second was stronger than the first. Such is the case with their last example. In this case we have the set theories $Z$ and $P$, the impredicative extension of the theory of finite sets. Wang’s result is that each contains a truth-definition for the other. Here $Z$ has more theorems than $P$, but $P$ has more modes of expression than $Z$.

We’ll return to this example below and question whether we really have a definition of truth for $Z$ in $P$.

“In conclusion, there are two points that we think deserve special emphasis. First, all the loose talk about a ‘hierarchy of languages’ being the natural upshot of Tarski’s work on truth can be seriously misleading. ... Moreover, we certainly do not get, from the requirement of correct truth definitions alone, a hierarchy of ever stronger metalanguages, as is sometimes suggested.” (“Tarski on ‘Essentially Richer Metalanguages,” 23.)

3. An account of essential richness

Here we’ll return to the rejected suggestion of basing our account of essential richness on translation, but with a better account of translation.

A. Convention T and “language”

“... 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.” (“The Concept of Truth in Formalized Languages,” 166-7.)

Convention T requires:
(1) a notion of translation from OL into ML
(2) a notion of naming applicable to both OL and ML
(3) a notion of provability and inference applicable to both OL and ML.
Suggestion: think about this in terms of Frege’s distinction between sense and meaning (reference). Think about (2) under the heading of reference; (3) under the heading of sense; and think of translation as having to preserve both meaning and sense.

“... as has frequently been remarked, in his writing on truth Tarski does not distinguish, as logicians are typically careful to do today, between a language and a theory.” (“Tarski on ‘Essentially Richer Metalanguages,” 5.)

DeVidi and Solomon often seem to think of a language as merely a set of well-formed expressions, that is an uninterpreted language (e.g. in their discussions of “normal” and “non-normal” truth-definitions).

B. Translation and essential richness again

Suppose that translation has to preserve sense and meaning of the expressions of a language. Translation has then to be conceived of as term to term, rather than sentence to sentence, as DeVidi and Solomon’s account has it.

Suppose the following constraints on translation. A translation from $L_1$ to $L_2$ will be a map $t$ from the well-formed expressions of $L_1$ to the well-formed expressions of $L_2$ such that

$(\alpha)$ $t$ respects syntactic type: the translation of a singular term is a singular term, the translation of a predicate is a predicate (which may be complex, i.e. we can translate a simple predicate with an open formula), the translation of a connective is a connective, the translation of a sentence is a sentence, the translation of a quantifier is a quantifier (allowing unbounded quantifiers to be translated into bounded quantifiers) etc.

$(\beta)$ $t$ respects the compositionality of meaning; in particular, for a simple predication $P\tau$, $t(P\tau) = t(P)t(\tau)$, and for a biconditional $\phi \leftrightarrow \psi$, $t(\phi \leftrightarrow \psi) = t(\phi) t(\leftrightarrow) t(\psi)$.

$(\gamma)$ the translation of a logical connective is the corresponding logical connective (negation into negation, etc), and the translation of a quantifier is the corresponding quantifier (except that it may be bounded).

$(\delta)$ the translation of a theorem is a theorem. (This is because translation has to preserve sense and we are thinking of provability as reflecting sense, thought of as inferential role.)

$(\epsilon)$ where we have back-and-forth translation, the translation of the translation of any sentence is provably equivalent to the original sentence. (This is because translation is supposed to preserve sense, and sameness in sense should be captured here by provable equivalence.)

$(\zeta)$ translation preserves reference.

This notion of translation is stronger than DeVidi and Solomon’s “strong translation.” But some of its clauses derive from their account, clearly.

Now I propose to revive the explanation of “essential richness” in terms of translation:

$L_1$ is essentially richer than $L_2$ if there is a strong translation of $L_2$ into $L_1$ but not conversely, while they are equally rich if each is stongly translatable into the
other.

C. An argument for essential richness

We will now argue that if ML and OL are equally rich, and ML contains a truth-definition for OL, then OL contains a truth-definition for itself. For simplicity assume that ML and OL share the same logical vocabulary, in particular ‘↔’ is the biconditional in each.

Suppose ML contains a truth-definition of OL.

Then OL is translatable into ML, since ML together with the definition proves all instances of schema T, and these instances require translations in ML of all sentences of OL. So suppose t translates OL into ML.

Further ML contains names for every OL-sentence. Let [φ] be the ML name of the OL sentence φ. We can assume that [φ] is a structural-descriptive name of φ (since Convention T is stated in terms of structural-descriptive names).

Further there is some open formula $Tr(x)$ of ML such that for every sentence φ of OL

$\vdash_{ML} Tr([φ]) \rightarrow t(φ)$.

Now suppose that s translates ML back into OL. Note that by reference-preservation (ζ), s([φ]) is a name of φ, and by compositionality of translation, (β), s([φ]) is in fact a structural-descriptive name of φ.

Now we get, by (δ) (the translation of a theorem is a theorem)

$\vdash_{OL} s(Tr([φ])) \rightarrow s(t(φ))$

Now from compositionality of translation, (β), we get

$\vdash_{OL} s(Tr([φ])) \rightarrow s(t(φ))$

Since the translation of a translation is equivalent to the original sentence, (ε), it follows that

$\vdash_{OL} s(Tr([φ])) \rightarrow φ$

From compositionality of translation again, (β), we get

$\vdash_{OL} s(Tr([φ])) \rightarrow φ$

But now, as noted above, s([φ]) is a structural-descriptive name of φ. Thus we have found a truth-definition for OL in OL itself. And so, assuming enough syntactic resources in OL, we will
get the Liar paradox.

D. Analysis of examples (1)-(3)

First, if we consider the examples under (1) and (2) we easily see that in these cases ML is essentially richer than OL by our account.

Second, what about the case in which Wang is said to have proved that P and Z contain truth-definitions for one another? In this case, it seems we should deny that P contains a truth-definition for Z. For after all, the so-called “truth-definition” is not going to make various theorems of Z, e.g. the axiom of infinity, come out as provably true. The appearance that there is a truth-definition here depends on the assumption that, as the language of P is an extension of the language of Z, we can replace Convention T with the requirement that every instance of

\[ \text{Tr}([\varphi]) \rightarrow \varphi \]

be provable in ML. But this tacitly assumes that \( \varphi \) (in ML) is a correct translation of \( \varphi \) (in OL), and this assumption is unwarranted. In effect, DeVidi and Solomon are here relying on the notion of a language as an uninterpreted formal system. They think that it is to such languages that Convention T “applies in the first instance” (“Tarski on ‘Essentially Richer Metalanguages,” 26, fn 18), but this is the sort of case for which Tarski has told us the problem of truth “has no relevance.” (“The Concept of Truth in Formalized Languages,” 166-7.)

This involves DeVidi and Solomon’s distinction between “normal” and “non-normal” truth definitions; in effect we are denying that the latter are truth definitions at all. We can discuss this further after the presentation.

4. Conclusion

Tarski was right!