Frege and the Hierarchy
Author(s): Tyler Burge
Source: *Synthese*, Vol. 40, No. 2, Logic and Linguistics, Part III (Feb., 1979), pp. 265–281
Published by: Springer
Stable URL: http://www.jstor.org/stable/20115349
Accessed: 11-04-2017 04:15 UTC

TYLER BURGE

FREGE AND THE HIERARCHY*

At the level of surface syntax in statements of propositional attitude, certain classical principles of substitutivity commonly associated with extensionality fail. The principles are

$$t = s \rightarrow (A^x/t \leftrightarrow A^x/s)$$

and

$$(A \leftrightarrow B) \rightarrow (\dots A \dots \leftrightarrow \dots B \dots)$$

where t and s are terms; A and B are formulae; 'A^x/t' signifies the result of substituting t for variable x in one or more occurrences; and '$\dots A \dots$' signifies any formula in which A occurs. These principles fail at the surface level in that, for example, from 'Bonn is the birthplace of Beethoven' and 'Anton believes that Bonn is in Germany' (on one reading), we cannot deduce 'Anton believes that the birthplace of Beethoven is in Germany'. Analogously, from 'Berg was Austrian and Brahms was German', the sentence, 'Anyone who believes that Brahms was German believes that Berg was Austrian' does not follow as a matter of logic.

The most thoroughly studied response to such failures of substitution (both in belief contexts and elsewhere) is to give up the principle of extensionality: the principle that the denotation, or extension, of an expression (including the truth-value of a sentence) is a function of the denotations, or extensions, of its semantically relevant parts as they occur in the expression. The response involves taking terms like 'believes' (or 'believes that') to be sentential operators, restricting the classical syntactical laws of substitution and quantification, and relativizing the semantical relation of denotation (or satisfaction) to a possible world. Let us call this 'the sentential operator approach'. This general approach to substitution failures in natural language has unquestionably deepened our understanding of semantical structure. But for many purposes, the key notion of possible world seems less clear than the discourse it is introduced to interpret. Moreover, it is hard to take seriously the notion at face value, as not to be explained in more primitive terms, since

its typical metalinguistic use demands that there be worlds other than the actual one and individuals other than the actual ones. The appeal to possible worlds is more naturally taken as a heuristic aid in constructing completeness and consistency proofs and in formal reasoning with modal statements — a half-way house in arriving at an intended interpretation of the discourse being studied. Moreover, as applied to statements of propositional attitude, the notion of possible world has less intuitive appeal than elsewhere.

A further problem with the sentential operator approach is that it tends to leave one without the resources to capture certain natural inferences. For example, from

(1)   Arnold believed that dominant resolves to tonic

and

(2)   The most basic point of classical harmony is that dominant resolves to tonic,

we may conclude

(3)   Arnold believed the most basic point of classical harmony.

(3) illustrates the problem most clearly. Since the sentential operator approach takes expressions like 'believes' (or 'believes that') to apply syntactically to sentences, it cannot easily account for the noun phrase that typically follows such expressions and for inferences (either substitutions or generalizations) that turn on that phrase.

The problem about generalization runs deeper. An operator approach could hold that sentences of the form 'Arnold believed something' are to be represented with a substitutional existential quantifier ['$(\exists p)$ Arnold believed that p'], which takes sentences as substituends. But such a representation assumes that all of Arnold's beliefs are expressible in English sentences. This assumption seems to me implausible for the general case, but I shall not argue the point here. Suffice it to say that unless the assumption is sound, the operator approach does not have even the beginning of an account of the quantification onto the content clause.

Frege took another approach to the substitution failures. Instead of treating them as counterexamples to extensional principles, he regarded them as evidence that the occurrences of terms and sentences within that-clauses

(and other such contexts) have different semantical functions than occurrences of those same terms and sentences outside of that-clauses. Thus 'Bonn' in the report of Anton's belief denotes not the German capital, but something else — what Frege called its oblique denotation. And the expression 'Brahms was German', as it occurs in the relevant that-clause, functions not as a sentence that is true or false, but as a complex singular term denoting (obliquely) an object of possible belief. Since 'believes' is, on the Fregean view, a predicate that applies syntactically to terms, there is no difficulty in accounting for inferences like (1) — (3). What is more, the view avoids the unintuitive elements of the possible world analysis.

My purpose in this paper is to discuss from a rather abstract point of view the effect of embedded oblique contexts on the Fregean strategy. I shall focus on the question: Do embedded oblique contexts pressure one to postulate an infinite (or indefinitely high) hierarchy of entities, with each level of the hierarchy serving to provide denotations for expressions in different degrees of embeddedness?

I

Two ways of formally accounting for the shifts of denotation by and within that-clauses are implicit in Frege's work. According to one (Method I), the formal representation allows expressions to have different denotations according to the syntactic context in which they occur. In one sense, this view allows ambiguity into the formal representation itself: a given expression has various denotations. But the view does not violate any fundamental assumption about the formality of formal systems as long as the denotation of an atomic expression can always be rigorously specified by reference to the syntactical form of complex expressions in which it occurs.

On Method I, then, the relation between an expression and its denotation is relativized to a syntactic context.[1] Thus, roughly speaking, a semantical theory for natural language would contain rules like: the denotation of 'Bonn' relative to its occurrence in any context of class T is, if anything, Bonn — where T is a rigorously defined class of syntactic contexts (what is sometimes called 'transparent contexts'). The denotation of 'Bonn' relative to ordinary oblique that-clause contexts would be something other than Bonn. Method I requires that the classical laws of substitutivity be restricted. On the

assumption that Bonn is the capital of Germany, we may interchange 'Bonn' and 'the capital of Germany' only in syntactic contexts where the denotation of the singular terms is the same as it is in identity contexts. Despite this restriction on the syntactical rule most closely associated with extensionality, the account that results from following this method is fully extensional: The truth-value of a belief sentence is determined solely by the denotations (or satisfaction values, or extensions) of its parts — as they occur in the sentence.

The syntactical notion of substitutivity and the semantical notion of extensionality are sometimes lumped together. And often the former is used as criterion for the latter. But Frege's strategy (assuming for the moment that he followed Method I) is illuminatingly described by distinguishing them. Frege's strategy was to treat apparent counterexamples to extensionality as cases of ambiguity. His primary tool for the analysis of language was to treat the truth-value of a sentence as a function of the entities denoted by the semantically relevant components of the expression. The substitution failures showed only that expressions in non-oblique contexts (say, identity contexts) denoted something different from what they denote in oblique contexts. The motivating principle of the analysis is preserved on Method I, even though the denotation of an expression can be properly specified only in the context of a containing sentence.

There is no reason to think that Frege would have held that languages involving the relevant ambiguity cannot be extensional. There is no reason for thinking that Frege would have held that extensionality fails in natural languages and can be preserved only in 'well-constructed' formalized languages where the systematic ambiguity that he attributed to natural languages is removed. Such a thought would put the matter in the wrong light: It is substitutivity that, according to Frege, fails in natural languages. The intuitive notion of extensionality, which he took to be fundamental to semantical evaluation, is unaffected. To allow for ambiguity, we must formulate the intuitive notion so that the relevant denotations or extensions of sentential components are understood to be the denotations or extensions of the components *as they occur in the relevant sentence.* But that qualification is hardly controversial. Modal logic and current belief logics, of course, flout both substitutivity and extensionality principles under their usual interpretations. So Method I contrasts with these approaches semantically, if not in its rejection of the principles of substitutivity.

The second Fregean method for explicating the shifts of denotation (Method II) is to represent expressions in natural-language oblique contexts with symbols which are different from the symbols that represent those same expressions as they occur in ordinary contexts. [2] There is no need on Method II to relativize denotation to a syntactic context and no need to restrict the classical laws of substitutivity and quantification.[3] Method II again preserves the principle of extensionality. On either method the substitution failures in natural language are no more counter-examples to extensionailty than fallacies of equivocation are counter-examples to modus ponens. We should not think of Frege as trying to translate non-extensional natural language into an extensional formal language (which is how Carnap and his students viewed the situation). Rather we should see him as arguing that, though ambiguous and syntactically misleading, natural language is covertly extensional.

In discussion of the two methods, it is often suggested that the first accounts for natural language whereas the second reforms it. This view is not clearly justified. One may regard the two methods as competing accounts of ambiguity in natural-language surface sentences. A standard device used in linguistic accounts of ambiguity is to subscript in the formal representation different representations of an ambiguous surface expression. Method II simply employs that device. What is partly at issue between the two methods is whether the ambiguity involved in the substitution failures should be resolved purely by reference to syntactic context, or whether it should be explicated, at least partly, by distinguishing the readings of particular expressions.

Interestingly enough, if Method I were applied directly to the surface sentences of natural language, it would be doomed. One cannot always determine purely on the basis of surface syntax whether or not an expression has an oblique denotation. In the sentence

(4)   Schumann believed the Polish youth to be a genius

the expression 'the Polish youth' may be taken to have either an ordinary or an oblique denotation. Thus we may understand the sentence either as admitting or as blocking substitutions of other singular descriptions of Chopin for 'the Polish youth'. Method I is therefore best regarded as a strategy for constructing and explicating *formal representations* of surface sentences. As such, it will be committed to distinguishing the different interpretations of (4) with syntactically distinct (as opposed to merely

lexically distinct) formalizaions. Now Method II is not thus committed. But I believe that the different interpretations of (4) *are* structurally distinct — and should be so construed by either Method.[4] Moreover, either Method should be able rigorously to define among the formal representations of natural-language sentences the syntactic contexts that represent oblique occurrences in surface sentences. It thus appears that Method II accounts for the shifts of denotation both by syntactical distinctions and by lexical distinctions. Is this not a sort of explanatory overkill?

The answer depends partly on whether Method II can offer compensatory advantages. It can. We have already noted that it provides a simpler seman-tical analysis for formal representations (it need not relativize the denotation or satisfaction relation to a syntactic context) and a simpler set of principles to account for deductive inferences (substitutivity and quantification laws need not be restricted).

On Method I even the grammatical categories should be relativized to a syntactic context. Thus Method II's investment in lexical complexity in accounting for ambiguity pays dividends in structural simplicity. I shall not try to judge the relative merits of this trade-off between the two methods. But I will later argue that from one point of view, Method I's lexical investment is no less substantial than that of Method II.

So far I have avoided asking after the nature of the oblique denotations of terms and sentences in natural language. This issue has been something of a sore spot for the Fregean approach. Frege himself believed that the oblique denotations are senses that terms and sentences express when they are used in ordinary, non-oblique contexts. And he regarded these senses as abstract entities, existing and applying to their associated denotations independently of any language that expresses them.

A major source of unwillingness to accept the overall Fregean approach has been refusal to think of senses as so completely independent of inten-tional agents. I approve of this refusal. But the Fregean semantical viewpoint does not of itself commit one to the Fregean ontology. One might follow Frege's general strategy for saving extensionality by appeal to shifts of denotation in natural language surface sentences (appealing to either Method I or Method II), but hold that the oblique denotation of expressions are the expressions themselves, not extra-linguistic senses.[5] And there are other alternatives. I mention the ontological question not because I want to pursue

it here, but because I want to emphasize that Frege's semantical strategy has a general interest regardless of one's ontological viewpoint.

<div align="center">II</div>

As we have described it, Method II holds that in unembedded oblique contexts like 'Bela believes Opus 132 is a masterpiece', the sentential expression 'Opus 132 is a masterpiece' is to be formally represented not by a sentence but by a term, say 'α'. For the sake of argument, let us suppose that 'α' denotes the proposition that Opus 132 is a masterpiece. According to Method II, 'α' may be exchanged for any other term (in the language of the formal representation) that denotes the same proposition. This yields reasonable results in representations of sentences containing unembedded occurrences of 'believes'. For suppose that 'π' represents 'Sergei's favorite proposition' and that Sergei is supremely enamored of the proposition that Opus 132 is a masterpiece. Then we can conclude: Bela believes α if and only if he believes π.

But embedded contexts present a new situation. Indeed, it can be argued on certain plausible assumptions that Method II must appeal to a hierarchy of some kind in order to represent embedded oblique contexts. The argument reduces to absurdity the view that 'α' may represent 'Opus 132 is a masterpiece' as it occurs in (5):

(5)   Igor believes Bela believes Opus 132 is a masterpiece.[6]

Let us assume then that 'α' (which denotes the proposition that Opus 132 is a masterpiece) represents 'Opus 132 is a masterpiece' as it occurs in (5). (5), we shall assume, asserts a relation of belief between Igor and the proposition that Bela believes that Opus 132 is a masterpiece. I shall denote this proposition by the expression '$\Gamma_1(\beta_1,\alpha)$'. Thus on our assumptions, (5) is formalized as

(6)   Believes (Igor, $\Gamma_1(\beta_1,\alpha)$)

By the principle of extensionality, the denotation of '$\Gamma_1(\beta_1,\alpha)$' is a function of the denotations or extensions of its parts. I shall assume that 'β_1' denotes the sense of 'Bela' and that 'Γ_1' denotes the sense of 'believes' — a function from β_1 and α to the relevant proposition.[7]

We assume the principle that a given sense is associated with a unique denotation or extension. Thus the proposition $\Gamma_1(\beta_1,\alpha)$ is associated with (or, in Church's terminology, is a concept of) a unique denotation or extension, its truth value.

We assume that this truth value is a function of the unique denotations or extensions associated respectively with the senses that determine the proposition. To put this another way, the truth value of the proposition is a function of the denotations or extensions of expressions that express its component senses. Let 'β' express β_1 and denote Bela; let 'Γ' express Γ_1 and denote what 'believes' denotes (or have its extension). Let 'α_0' express α and denote its truth value. (We suppose that truth value to be truth.) Then '$\Gamma(\beta,\alpha_0)$' expresses $\Gamma_1(\beta_1,\alpha)$ and denotes its truth value. 'Believes' originally applied to persons and propositions. But on our assumptions it has come also to apply to persons and truth values. This leads to absurdity in short order.

For given the classical substitution laws of Method II, we may substitute any expression that denotes truth for 'α_0' in '$\Gamma(\beta,\alpha_0)$' and preserve the truth value of '$\Gamma(\beta,\alpha_0)$'. (We speak of both sentences and propositions as having truth value.) But '$\Gamma(\beta,\alpha_0)$' supposedly expresses the proposition that Bela believes Opus 132 is a masterpiece. So it seems to follow that if Bela believes Opus 132 is a masterpiece, he believes every truth.

The argument shows that on these assumptions 'Opus 132 is a masterpiece' in (5) cannot be represented by a term 'α' denoting the proposition that Opus 132 is a masterpiece. It is prima facie plausible to assume with Frege that the expression as it occurs in (5) should be represented by a term denoting the sense of the expression that represents 'Opus 132 is a masterpiece' as it occurs in unembedded belief contexts. Given this assumption, the argument can be replicated to show that the sentential expression as it occurs in doubly embedded oblique contexts must be represented by yet another term — and so on.[8]

Let us review the assumptions of the argument. The assumptions of extensionality and classical substitutivity are basic to Method II. Distaste for the hierarchy led Carnap to relinquish both.[9] I shall not discuss this large issue here except to register the view that no non-extensional theory of belief has provided an alternative that is plausibly superior to the Fregean approach. We shall discuss giving up classical substitutivity, but not extensionality, when we come to Method I.

I assumed that β_1 is the sense of 'Bela', that Γ_1 is the sense of 'believes', and that these senses are denoted in (5). Taking Γ_1 to be a function from senses to propositions is not strictly necessary to the argument. It is possible that one would want the sense of 'believes' to be individuated more finely than identifying it with such a function would allow. Moreover it is possible that one would want to represent the sense of 'Bela' not as an individual concept, but (in some more Russellian way) as some sort of function with propositions in its range. Such variations will not affect the argument as long as the other assumptions stand.

We have been accepting for the sake of exposition the Fregean assumption that unembedded oblique sentential expressions denote a proposition, and that oblique expressions' generally denote senses. But the argument for the hierarchy does not depend on commitment to Fregean senses. For the relation between a linguistic expression and the sense it expresses, we could substitute a relation between the same expression and a standard syntactical name of it. For the relation between the sense of an expression and the denotation associated with it, we may substitute the relation between the standard name of an expression and the denotation of its denotation. The resulting argument would be equally plausible. And other ontological alternatives are open.

I assumed that the truth value associated with a proposition is a function of the denotations or extensions of the expressions that express its component senses. This assumption is almost the principle of extensionality again, but it adds that senses are associated with the denotations of expressions that express them. The addition is, I think, hardly controversial. It may be thought to be if certain distinctions are not kept in mind. It is tempting to note that (5) may be read as follows:

(7)   Igor believes of the proposition that Opus 132 is a masterpiece that Bela believes it.

(6) might be regarded as a first approximation analysis of this reading. Now one might go on to reason that the truth value of what Igor believes is a function of the ordinary denotations of 'believes' ('Γ'), 'Bela' ('β'), and 'the proposition that Opus 132 is a masterpiece' ('α') – *not 'Opus 132 is a masterpiece.* ('α_0'). In general, to find the truth value of a proposition mentioned in a belief sentence that involves transparent contexts, we do not 'drop down'

a type level for each of the component senses. Sometimes, as Russell empha-
sized, the components of such proposition will not even *be* senses: they may
be individuals. None of these considerations, however, affects the argument.
For (7) is not the relevant interpretation of (5). We may substitute coexten-
sive phrases for 'the proposition that Opus 132 is a masterpiece' in (7) and
preserve truth value. But on one reading of (5), analogous substitutions
fail.[10] It is to this reading that the assumption applies.

Finally, I assumed that for a given sense there is associated a unique
denotation. This principle has been rejected in application to certain forms of
language — for example, demonstratives. But these forms of language are not
ubiquitous. It is easy to present sentences which do not contain them. A
thorough exploration of a semantics that rejects the principle generally has
not, to my knowledge, been undertaken. In summary, the argument for a
hierarchy on Method II seems very powerful.

It would be mistaken to try to simplify the argument so as to dispense
with all assumptions except that of substitutivity, as follows. If (5) is
formalized as (6), then if α is Zoltan's favorite proposition (which latter
phrase we represent by 'π'), then (8) follows from (6):

(8)   Believes(Igor, $\Gamma_1(\beta_1, \pi)$).

But on the assumption that (6) formalizes (5), (8) would seem to formalize

(9)   Igor believes that Bela believes Zoltan's favorite proposition.

But (5) and the relevant identity do not yield (9) — Igor may never had heard
of Zoltan. So (concludes the argument) on the assumption of classical
substitutivity, 'α' cannot represent 'Opus 132 is a masterpiece' as it occurs in
(5).

The argument does not work. (8) does not formalize (9). For since
'Zoltan's favorite proposition' is in an oblique belief context (at least the
oblique context of the first 'believes'), it must denote its ordinary sense. (In
any case the argument would prove too much, for it could be adapted to
show that *no* reasonable representation could be maintained if classical
substitutivity were preserved.) So to get the argument for the hierarchy, it
seems that the longer route is necessary.

An infinite hierarchy of entities is, I suppose, theoretically unappealing
other things being equal. But the present hierarchy has considerably more

intuitive content than has sometimes been imagined. We can bring this out as follows. According to (5), Igor has a belief that makes reference to a certain belief-content (to the belief that Opus 132 is a masterpiece). As we ascribe it in (5), Igor's belief makes reference to that belief content in a certain way — it specifies it by applying a standard name. Igor might have had a belief that is best represented as making reference to that belief content in other ways, say by describing it as the first point made in a given chapter of a given book on 19th century chamber music. As we specify Igor's belief in (5), we make reference not directly (or not merely) to the belief content he ascribes to Bela, but to a certain means of ascribing it, a means that we ascribe to Igor. One means may be available to Igor — another, not. The subject matter of the hierarchy is thus certain means of ascription, means of ascribing those means, and so on.

III

Let us now view (5) through the lens of Method I. As conceived by Frege, that method too involves an infinite hierarchy of senses. There is not, however, the same pressure on Method I as on Method II to generate a hierarchy: The argument for the hierarchy that we gave in section II depended on the ability to substitute extensionally equivalent expressions in all contexts. Indeed, it has been suggested that the hierarchy is superfluous given Fregean principles and Method I.[11] Nevertheless, there is a sense in which the hierarchy is inevitable (given certain further assumptions) even on Method I.

We have described Method I as a strategy for creating a language L_I that formally represents the logical form of English. Our remarks about the semantical functions of expressions of L_I occur in a metalanguage ML_I. In order to give a systematic account of the truth conditions of sentences of L_I, we should formalize ML_I and give within it a theory of truth for L_I. If we are to give a theory that works on the structure of the sentences of L_I, we cannot construct ML_I on the plan of Method I. For to explain the denotations of complex expressions in terms of the denotations of their parts, we must be able to exchange within (translations of) the expressions of L_I different terms with the same denotations. On Method I, substitutions within expressions representing ordinary oblique contexts are allowable only if the expressions have the same sense. But the sort of expressions that need to be exchanged —

e.g. γ, $\ulcorner$the denotation of γ relative to contexts $O\urcorner$, and $\ulcorner$the sense of γ relative to contexts $T\urcorner$ — clearly do not have the same sense. So on Method I the substitutions could not be carried out in the metalanguage, nor therefore could a systematic theory of truth be given for L_I. Thus ML_I has to be constructed so as to allow substitutions of extensionally equivalent expressions which do not express the same sense.[12]

Now suppose, contrary to Frege, that L_I lacks a hierarchy. Expressions representing transparent contexts denote their ordinary denotation and express their ordinary sense; expressions representing oblique contexts denote their ordinary sense and express no further sense. (Cf. note 11.) We shall state the semantical rules of ML_I relevant to accounting for the truth conditions of (5), as represented in L_I:

(a) $(x)(x = A(\alpha, \ulcorner Igor_L \urcorner, T) \leftrightarrow x = Igor)$

(b) $Sat(\alpha, \ulcorner Believes_L(e_1, e_2) \urcorner, T) \leftrightarrow Believes(A(\alpha, e_1, T), A(\alpha, e_2, O))$

(c) $A(\alpha, \ulcorner Bela_L \urcorner, O) = Sense(\ulcorner Bela_L \urcorner)$

(d) $A(\alpha, \ulcorner Believes(e_1, e_2) \urcorner, O) = C_3(Sense(\ulcorner Believes_L \urcorner), A(\alpha, e_1, O),$
$$A(\alpha, e_2, O))$$

(e) $A(\alpha, \ulcorner Op. \ 132_L \urcorner, O) = Sense(\ulcorner Op. \ 132_L \urcorner)$

(f) $A(\alpha, \ulcorner Masterpiece_L(e_1) \urcorner, O) = C_2(Sense(\ulcorner Masterpiece_L \urcorner), A(\alpha, e_1, C$

'$A(\alpha, e_1, O)$ signifies the assignment of any sequence α to expression e_1 as it occurs in any member of the class of oblique-representing contexts O. 'T' specifies the class of transparent-representing contexts. 'Sat' is the satisfaction predicate for L_I. I have subscripted mentioned expressions of L_I with an 'L' to distinguish them from the used expressions of ML_I. Thus 'Believes$_L$' of L_I (whose second argument expression may be exchanged only for expressions expressing the same sense) is to be distinguished from 'Believes' of ML_I into which it might be translated. Substitutions within the scope of the latter are less restricted. (Cf. note 12.)

$\ulcorner C_3(Sense(\ulcorner Believes_L \urcorner, a, b)\urcorner$ signifies the composition of the sense of 'Believes$_L$' with senses, a and b, of the expressions to which 'Believes$_L$' is applied. Such a composition is the proposition expressed by the sentence produced by predicating 'Believes$_L$' of its argument expressions.[13] We

assume therefore these laws:

(A) $C_3(\text{Sense}(e_1), \text{Sense}(e_2), \text{Sense}(e_3)) = \text{Sense}(\ulcorner e_1(e_2, e_3)\urcorner)$

(B) $C_2(\text{Sense}(e_1), \text{Sense}(e_2)) = \text{Sense}(\ulcorner e_1(e_2)\urcorner)$

where ``$e_1(e_2)$'' signifies the predicative application of $e_1$ to $e_2$. ``$e_1(e_2, e_3)$'' is analogous.

By (a) – (d) it is easy to prove:

(i) $\text{Sat}(\alpha, \ulcorner \text{Believes}_L(\text{Igor}_L, \text{Believes}_L(\text{Bela}_L, \text{Masterpiece}_L(\text{Op. } 132_L)))\urcorner$
$\leftrightarrow \text{Believes}(\text{Igor}, C_3(\text{Sense}(\ulcorner \text{Believes}_L\urcorner), \text{Sense}(\ulcorner \text{Bela}_L\urcorner),$
$A(\alpha, \ulcorner \text{Masterpiece}_L(\text{Op. } 132_L)\urcorner, O)))$

By (i), (e), (f), and (B), we have

(ii) $\text{Sat}(\alpha, \ulcorner \text{Believes}_L(\text{Igor}_L, \text{Believes}_L(\text{Bela}_L, \text{Masterpiece}_L(\text{Op. } 132_L))$
$\leftrightarrow \text{Believes}(\text{Igor}, C_3(\text{Sense}(\ulcorner \text{Believes}_L\urcorner), \text{Sense}(\ulcorner \text{Bela}_L\urcorner),$
$\text{Sense}(\ulcorner \text{Masterpiece}_L(\text{Op. } 132_L)\urcorner)))$

It will be noticed that these proofs depend on allowing within the scope of 'Believes' (not 'Believes$_L$') substitutions of expressions that intuitively differ in sense.

Now from one point of view, neither (ii) nor the result of applying (A) to it can be our final explication in ML_I of the truth conditions of $\ulcorner \text{Believes}_L$ $(\text{Igor}_L, \text{Believes}_L(\text{Bela}_L, \text{Masterpiece}_L(\text{Op. } 132_L)))\urcorner$. (Call this sentence '(D)'.) For they would not explain the truth conditions of (D) to someone who understood ML_I but who did not already understand the expressions of L_I (did not understand what their senses are). Although (D) may perhaps describe the truth conditions, it does not 'give' them. To give the truth conditions, we need a sentence on the right side of the biconditional which states the relevant truth conditions purely in the terms of ML_I, without mentioning expressions of L_I. That is, we need a plausible translation of (D) into ML_I. From this viewpoint, an adequate theory of truth for L_I must satisfy Tarski's convention T, which requires that the metalanguage in which the theory is given provide a translation of the relevant sentence of L_I.[14] What might such a translation be?

It cannot be a sentence produced by translating component expressions of (D) by expressions in ML_I having the same denotation in their respective syntactic contexts. For suppose that we assume that the ordinary sense and

denotation of 'Believes$_L$' are the same as those of 'Believes'; those of 'Igor$_L$' are the same as those of 'Igor', and so on. Then since L$_I$ lacks a hierarchy, such a translation would yield a sentence amounting to (6). But an argument analogous to the one we applied to (6) earlier establishes that the purported translation is not equivalent to the original English sentence (5). So it would not be equivalent to any adequate representation of (5) in L$_I$.

Of course, we can reasonably question our assumption that the ordinary sense and denotation of 'Believes$_L$' are the same as those of 'Believes' (and so on). But the grounds for questioning make reference to the same mismatches in structure between L$_I$ and ML$_I$ that we used to show that a term by term translation will not work. Giving up the assumption would make such a translation even less likely. It thus appears that there may be no way to 'give' the truth conditions of (D) in ML$_I$.

Let us summarize the difficulty. A language of the form of L$_I$ cannot give a systematic truth theory for L$_I$ — one needs laxer substitution principles than L$_I$ countenances. Explications of truth conditions like (ii) do not satisfy the translation requirements of Tarski's Convention T. But term by term translations of embedded belief sentences of L$_I$ into the 'laxer' meta-language are prevented by the argument of section II. On the other hand, to attribute a hierarchy to L$_I$ would be to controvert the hypothetical *informal* explication of embeddings in L$_I$, according to which no hierarchy is generated. Thus one might well feel that no such attribution would provide an intuitively acceptable *translation* of the relevant sentences of L$_I$. So Tarski's requirement is once again not met.

This is, I think, an interesting situation. It suggests a pair of options. The first is to emphasize the importance of Tarski's Convention T and hold that since (apparently) no informative and systematic truth theory for L$_I$ meets it, L$_I$ is an unlikely model of natural language since natural language should be capable (modulo the paradoxes) of interpreting itself. On this option, one would hold that an interpretation of a Method I language (as modeling natural language) would inevitably follow Frege's own lines and involve a hierarchy. For only thus will Tarski's requirement be met. A second option is to relax the requirement that an adequate theory of truth always provide translations (in some intuitive sense of good translation) of all the sentences of the object language. One might settle for semantical explications like (ii). These explain semantical structure without giving content. Assuming that one

could master the content of the sentences of the object language by learning them directly, a theory of truth need, on this option, do no more.

A choice between the two methods must surely depend on global considerations governing the ultimate aims of one's semantical theory and one's representation of logical form. Insofar as lexical representation is our main concern, Method I has a strong claim on our attention. Insofar as we aim at a semantical theory which provides both a simple account of the formal structure of our language and a plausible representation of its content, Method II will remain dominant.

University of California at Los Angeles

NOTES

* I am indebted to Nathan Salmon for comments on an earlier version, and to the National Endowment for the Humanities for its support.

[1] Method I appears to be suggested in Frege, 'On Sense and Reference' in P. Geach and M. Black (eds.), *Philosophical Writings of Gottlob Frege* (Blackwell's, Oxford, 1966). It is perhaps the most straightforward implementation of Frege's slogan: "Never ask for the meaning (Bedeutung) of a word in isolation, but only in the context of a sentence", *Foundations of Arithmetic*, trans. J. L. Austin (Northwestern University Press, Evanston, Illinois, 1968) p. (x).

[2] Michael Dummett, *Frege – Philosophy of Language* (Duckworth, London, 1973), chapter IX, suggests that Method I is clearly the method favored by Frege. But this is misleading. Frege's general insistence on avoiding ambiguity in a well-constructed language (perhaps the representing language) and a passage in a letter to Russell December 28, 1902 militate against Dummett's suggestion: "Eigentlich musste man ja, um Zweideutigkeit zu vermeiden, in ungerader Rede besondere Zeichen haben, deren Zusammenhang aber mit den entsprechenden in gerader Rede leicht erkennbar wäre." ("Actually, in order to avoid ambiguity, one must in indirect speech have special signs whose relation to the corresponding signs in direct speech would be easily recognizable.") quoted in James M. Bartlett, *Funktion and Gegenstand* (Munich M. Weiss, 1961, p. 19); also in John Wallace, *Philosophical Grammar* (Dissertation, Stanford, 1970), pp.105–6. The chief proponent of Method II is Alonzo Church, 'A Formulation of the Logic of Sense and Denotation' in Henle, Kallen and Langer (eds.), *Structure, Method, and Meaning* (Liberal Arts Press, New York, 1951); 'A Revised Formulation of the Logic of Sense and Denotation', *Nous* 7 (1973), pp. 24–33; (1974), pp. 135–156.

[3] If we wished to apply semantical analysis to the natural-language surface sentences, then on Method II, we would have to relativize the chief semantical relation to a formal reading (e.g.: 'Bank' relative to the reading '(Money)bank' denotes (Money)banks). The laws of substitutivity may be expected to remain unrestricted, since they work essentially on formal representations in any case. In general, it is easier to apply the truth predicate to canonical or formal readings of the surface sentences.

⁴ For examples of representations that give syntactically distinct treatments to the different readings of the sentence, see John Wallace, 'Belief and Satisfaction', *Nous* 6 (1972), pp. 85–95; David Kaplan 'Quantifying In', in D. Davidson and J. Hintikka (eds.), *Words and Objections* (Dordrecht, D. Reidel, 1969); and my 'Belief *De Re*', *The Journal of Philosophy* 74 (1977), pp. 338–62.

⁵ I discuss one objection (the Church-Langford translation test) to taking expressions as oblique referents in 'Self-Reference and Translation', M. Güenthner-Reutter and F. Güenthner (eds.), *Meaning and Translation* (Duckworth, Cambridge, 1978). For a systematic exploration of several aspects of the formal relations between the syntactical and Fregean ontologies, see David Kaplan, *Foundations of Intensional Logic* (Dissertation, UCLA, 1964).

⁶ In what follows I shall confine myself to embedded contexts that do not contain semantical terms like 'true', or terms whose use carries semantical implications, like 'necessary' or 'knows'. These expressions complicate representations of embedded contexts in ways that I prefer to consider separately.

⁷ Already we have an inconvenience. In representing the second occurrence of 'believes' in 'Igor believes Bela believes Zoltan's favorite proposition', 'Believes' would most naturally be used to denote a two-place function from the sense of a term and the sense of a term to a proposition. But in this context, it has senses of sentences in its domain. But worse is to come.

⁸ The preceding argument is nowhere, to my knowledge, explicitly stated. Its conclusion and premises (near enough) are assumed in Alonzo Church, 'A Formulation of the Logic of Sense and Denotation', *op. cit.*, and David Kaplan, *Foundations for Intensional Logic, op. cit.* I should note that the appeal to a hierarchy of senses in the Fregean system is motivated not only by embedded oblique contexts, but also by higher-level extensions of the 'paradox' of identity.

⁹ Rudolf Carnap, *Meaning and Necessity* (University of Chicago Press, Chicago, 1956), pp. 129–144, 232.

¹⁰ Wallace, who presupposes Method II, attempts in 'Belief and Satisfaction', *op. cit.* to avoid a hierarchy. The idea briefly is to treat (5) as one would treat (7). For Wallace a singular term denoting a first-level proposition or propositional function always represents a sentential expression in oblique contexts regardless of whether the context is embedded. The attempt to avoid the hierarchy fails for the reason stated above: (7) is not the relevant reading of (5). Another approach originally developed by Wallace in *Philosophical Grammar, op. cit.* is held by Wallace to avoid the hierarchy and is so advertised by Davidson in 'On Saying That', *Synthese* 19 (1969), pp. 136–137. This approach is not tested in these passages on embedded contexts. If it is, it will be seen to be subject to our argument.

¹¹ This language is proposed by Dummett, *Frege, op. cit.* pp. 266–269, as a means of avoiding a hierarchy. Dummett does not, however, investigate formalization of the metalanguage, which I shall argue urges a hierarchy at least from a certain point of view. Dummett's proposed language is not as compatible with fundamental Fregean principles as Dummett implies. In embedded oblique contexts, it controverts the principle that the denotation associated with a proposition is a function of the denotations associated with the component senses of the proposition.

¹² Actually, the argument I am giving does not require that ML_I have the *classical* substitution laws, though I shall assume here that it does. It might – as far as my

argument goes – allow in certain contexts substitution only of 'necessarily' or 'logically' equivalent expressions *which differ in sense*, as long as the language is extensional in its semantics in the sense that the argument involving sentence (3) required.

[13] For heuristic purposes, I have taken ML_I to be first order. As a consequence 'Sense ('$Believes_L$')' does not occur in function sign position. This is contrary to Frege's would-be intentions. But the argument I am giving does not depend on this point. I shall assume that C_3 and C_2 are defined on the senses of syntactically appropriate expressions.

[14] Alfred Tarski, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics* (Cambridge Press, Cambridge, 1956), pp. 187–188.