THE TARSKIAN HIERARCHY OF LANGUAGES

A STUDY OF THE TARSKIAN HIERARCHY OF LANGUAGES

By

JOHN THOMAS MOORE, B.A., B.ED.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

November, 1993

MASTER OF ARTS (1993)
(Philosophy)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: A Study of The Tarskian Hierarchy of Languages

AUTHOR: John Thomas Moore, B.A. (Trent University)

B.Ed. (University of Toronto)

SUPERVISOR: Professor Nicholas Griffin

NUMBER OF PAGES: v, 107

ABSTRACT

Alfred Tarski's formulation of the hierarchy of languages was intended to prevent semantic paradoxes from occurring in a formal language, and thus to preserve consistency. This thesis attempts to explain how Tarski accomplishes this, and to identify logical problems which occur in the hierarchy. Chapter one outlines the parameters of study. Chapter two outlines a simple version of the hierarchy of languages. Chapter three provides an exegesis of 'The Concept of Truth in Formalized Languages', explaining the notion of "essential richness", which is central to understanding how paradoxes are avoided by Tarski. Chapter four introduces problems relating to infinite regress, and a direct, general definition of truth is presented using Tarski's own notation. Questions relating to expressibility and proving consistency are addressed in chapter five. In the appendix, a weakness in Tarski's attempt to reduce semantic concepts to non-semantic concepts is shown using his own definitions and notation.

ACKNOWLEDGEMENTS

This thesis has taken an incredibly long time to complete. That I have been able to complete it at all is due to the patience of many people. Nick Griffin has been helpful in every stage of the development of this document. My wife Megan, and my children have all been quite forbearing over the years this has taken. Finally, my sister Audrey went above and beyond the call of family support to retype the first three chapters - no small task.

TABLE OF CONTENTS

Chapter One: Introduction

The notion of a hierarchy of languages has been used by logicians and philosophers for some time now as a solution to the semantic paradoxes. For all its importance, however, it has not received much direct attention, although there are a number of subtleties within it which are worth careful consideration. In what follows I intend to look at some of the features of this framework, with the goal of explaining its logical structure, and analyzing it for flaws of a logical nature.

First, however, it is necessary to indicate exactly what notion of a hierarchy is to be studied, and how it is to be studied. Hierarchies of languages have been employed for a number of purposes, not all of which involve the same sort of a hierarchy; for example:

A – Hilbert's program of proving mathematics consistent;

B – Russell's means of avoiding Wittgenstein's notion of showing in the _Tractatus_;

C – as a solution to the semantic paradoxes;

D – as a means of dealing with Lewis Carroll's infinite regress in "What the Tortoise said to Achilles", in Mind 1895, relating to the justification of deduction;

E – as a means of providing proofs for formulas which are undecidable in a given system;

F – as a tool for clearly distinguishing a language

> being discussed from the language used to perform the discussion.

The basic common element in all of these hierarchies is roughly that they all involve some sort of distinction between an object language (OL for short) and a metalanguage (ML), whereby the OL is "talked about" in the ML. There are, however, a number of differences. For example, Hilbert suggested only an OL (mathematics as an uninterpreted system) and an ML, while Russell's proposal involved an infinite series of languages (OL, ML, meta-ML, etc.) in his version. The focus of this thesis will be on C above, and generally, I will not be concerned with its compatibility with the other possibilities. (Hereafter, the term 'HL' will be used specifically to refer to the notion of the hierarchy of languages designed to deal with C above.) The standard source regarding the HL is Tarski's 'The Concept of Truth in Formalized Languages' (CTFL)[1] and consequently I will take Tarski's work as the point from which to begin,

---

[1]. See for example, the Encyclopedia of Philosophy, vol. 5, p. 49 on the Logical Paradoxes. See also, the Kneales' The Development of Logic; there they say that, on the basis of Ramsey's distinction between syntactic and semantic paradoxes, "it seems more natural to deal with the semantic paradoxes by special semantic measures, and the suggestion which has won most favour is the distinction of language and metalanguages by Tarski in his work 'The Concept of Truth in Formalized Languages'" (p. 665). Another example of Tarski's importance is the fact that, in his 'Outline of a Theory of Truth' Kripke takes Tarski's position as his main adversary (pp. 57-63 in Martin, ed.).

supplementing it with the work of other logicians/philosophers where necessary.

In order to formulate the goals of the thesis it will be useful to differentiate carefully purposes C and F above. The standard distinction between OL and ML is often used in a very loose manner which, though satisfactory for F, is insufficient for C. It is often said that when a language is used to talk about itself (as in a grammar text for example) it is both OL and ML at the same time.[2] Thus, presumably a sentence such as

1- This sentence has five words.

could be treated as being both an OL and an ML sentence at the same time. However, on this construal of the notions of OL and ML, the sentence

2- Phrase numero deux est fausse.

would also be accepted as both an OL and an ML sentence, and the paradox of the Liar would not be avoided. Furthermore, it would not be enough simply to look at sentence (2) in another language, say English, making French our OL and English our ML. In this situation we would still be forced into saying (in our English ML) that if sentence (2) is true, it must be false, and if it is false, then it must be true. If a hierarchy of languages is to be designed to

---

[2]. See, for example, Carnap, Introduction to Semantics, p. 4.

preclude the semantic paradoxes, then a simple distinction of use and mention is not enough.

Clearly then, the notions of OL and ML must be somewhat stricter for C than for F. As mentioned above, the notion of a hierarchy of languages as used to avoid the semantic paradoxes involves restricting the expressive capacity of a language; something not involved in cases where, as with F, we simply wish to observe the difference between use and mention.

In spite of this distinction between C and F, C is often, if not always, presented in a manner quite similar to F. In the first section of his CTFL, for example, Tarski goes about rejecting the use of natural languages for precisely explicating various semantic concepts, but then in the remainder of the article, he uses a natural language supplemented with special notation as his ML. The problem here isn't the use of natural language _per se_, but rather that natural languages (superficially at least) violate the restrictions placed upon expressive capacity present in a hierarchy designed to deal with C (but note that for F this state of affairs is perfectly acceptable). The nature of Tarski's use of natural language needs to be clarified, in order to understand the use of the HL to be studied here. There are two issues intertwined here; first, this use of natural language may give the impression that there are only two levels to the hierarchy (an OL and a natural language ML). Secondly, it may give the impression that only the OL needs to be formalized, and that the ML can remain unformalizable.

With regard to the first issue, there appear to be two ways of explaining the situation.[3] On the one hand, it can be argued that it is simply the semantics (and perhaps the syntax also) of the OL which are of interest to Tarski, and so long as the semantic concepts for this language are not logically flawed (i.e., they don't lead to the paradoxes),then it doesn't matter whether or not the semantic concepts applicable to the ML are logically flawed. On the other hand, it can be argued that the ML is capable of being formalized just as the OL is, in such a way that if it were expressed formally the paradoxes would clearly be avoided. The ML is not expressed formally because it is the OL which is (almost always) of direct interest, and thus there is no need to formalize the ML. The first alternative would in effect be an "abridged" version of the HL, consisting of only two levels, one formalized, and the other not,[4] while the second alternative tacitly leads to an

---

[3] Both of the alternatives outlined here seem legitimate; the question at hand is what approach is taken by the "standard" approach, as typified by Tarski.

[4] This appears to be the form of hierarchy used in A above.

infinite hierarchy.

An infinite hierarchy is needed for C above, because otherwise the ML would contain the semantic paradoxes - and thus no complete solution would be provided - or the ML would use some other means of precluding the paradoxes - in which case the HL would be only a partial solution. Tarski's system is in fact an infinite hierarchy (see for example 'The Semantic Conception of Truth' p.60.). Consequently, "abridged" hierarchies are not of interest here.

Another issue to clarify with respect to Tarski's use of natural language is that of whether or not the ML and higher levels can be formalized, and why this should be possible. By using natural language, Tarski is able to outline his system quite easily. At several points in CTFL however, he mentions that his ML <u>could</u> be formalized (e.g., CTFL, p. 170, p. 188, fn. 1, p. 195). Tarski gives no clear explanation of why it is important that the ML be formalizable. However, it is important to try and piece together what reasons he may have had for taking this alternative.

One reason for it seems to be that in order to prove certain metatheoretic claims (about the OL) we must use a meta-ML, and formalize our ML. Tarski suggests this regarding the proof of the assertion that his definition of truth is materially adequate (i.e., that it meets the conditions of his "convention T"; see p. 195, CTFL). A more fundamental and philosophical reason may be suggested by Tarski in 'The Semantic Conception of Truth' (SCT), where he says that

> "[i]f we take our work [i.e., semantics] seriously, we cannot be reconciled with [the presence of the antinomy of the Liar]. We must discover its cause, that is to say, we must analyze the premises upon which the antinomy is based; we must then reject at least one of these premises, and we must investigate the consequences for <u>the whole domain of our research</u>" (pp. 58-59, my emphasis).

Perhaps with regard to this last clause, he goes on, on the next page, to say that

> "if ... we become interested in the notion of truth applying to sentences, not of our original object language, but of its metalanguage, the latter becomes automatically the object language of our discussion; and in order to define truth for this language, we have to go to a new metalanguage - so to speak, a metalanguage of higher level. In this way we arrive at a whole hierarchy of languages." (p. 60)

If by "the whole domain of our research" (in the first quotation above) Tarski meant to include languages capable of referring to linguistic expressions, then the result is an infinite hierarchy in which the semantic concepts for each level can be analyzed in higher levels.

The question then is why such languages should be within the domain of the logician's research. Two possible reasons are: (1) if one were to attempt to interpret natural language as tacitly operating on the basis of a hierarchy of languages,[5] one would have to deal with languages capable of referring to linguistic expressions, such as natural languages; (2) (a reason probably more to Tarski's liking, given his indifference to natural language) it is desirable from a philosophical (or scientific, for that matter) perspective to have as complete an understanding as possible of semantic concepts such as truth, and consequently it is necessary to see how they apply to different kinds of languages, including those capable of referring to linguistic expressions.

Based on either or both of these purposes, we can conclude that the HL is a hierarchy which consists of an infinite number of languages (as opposed to an "abridged" hierarchy, which need contain only two levels), with each language above the OL having semantic concepts which are not logically flawed by the semantic paradoxes.[6] The ML, and

---

[5]. See for example, Charles Parsons `The Liar Paradox', or the Kneales' The Development of Logic, p. 665.

[6]. As will be explained at the end of chapter two, the simple theory of types is standardly introduced to keep the syntax free from the logical flaws manifest in the syntactic antinomies. For example: "such paradoxes are now generally dealt with by assuming not only a hierarchy of `parts of speech' (this is what the simple theory of types amounts to) but also a hierarchy of languages - a basic language, a `metalanguage' in which we discuss the meaning and truth of expressions in the basic language, a "metametalanguage" in which we deal similarly with the metalanguage, and so on." - A.N. Prior's article on Russell in The Encyclopedia of Philosophy, vol. 7, p. 251.

all higher levels, must be formalizable because only then can they be shown to be consistent[7] and the paradoxes clearly avoided. There is still an open question, however, regarding whether or not the hierarchy as a whole manages to avoid serious logical flaws.[8] For the two purposes mentioned above, this is just as important as the formal adequacy of any specific language within the hierarchy. It is also a point which, as I mentioned in the opening paragraph, seems to have received very little attention in spite of its importance. What I propose to do in this thesis is make a very modest attempt at examining the HL as a whole, to see if any serious logical flaws may be found in it.

This study will be related to the problem of applying the HL to natural language, but basically concerns a separate issue. I will not be concerned with whether the HL can adequately deal with the variety of natural language sentences involving semantic concepts (for example, Kripke's Nixon/Jones sentences in his 'Outline of a Theory of

---

7. In SCT, Tarski says that "the problem of consistency has no exact meaning with respect to this [i.e., natural] language" (p. 60).

8. Such flaws could be, for example, an inconsistency in the rules governing the construction of such hierarchies, an inconsistency between these rules and other logical principles, harmful infinite regress, and problems with our ability to express such a semantic theory at all, and thus an inability to demonstrate the consistency of the HL.

Truth'); the HL's (apparent) inability to deal with such sentences says nothing of its formal adequacy. On the other hand, if some logical flaw is present in the HL (or some version of it), this would obviously have implications for the attempts to apply the HL to natural language, given that such attempts are intended to overcome the apparent flaws in our intuitive understanding of semantic concepts such as truth. Similarly, this study will not address the question of whether or not Tarski's interpretation of the concept of truth (correspondence theory) is correct. The internal logical consistency of the hierarchy of languages is a separate issue from whether or not it accurately explains the concept of truth.

It is worth noting that in order to examine the HL for any logical flaws it will be necessary to treat it as, in a sense, a purely <u>formal</u> hierarchy. As mentioned above, for certain purposes the ML (and all higher levels) is assumed to be formalizable, meaning in effect that its syntactic and its semantic rules <u>can</u> be formally expressed. This, I believe, is Tarski's approach and consequently it is what I also will follow. For those interested in mapping the HL onto natural language, this is probably an overly strong demand. However, even they will agree that every language in the hierarchy must be treated as strictly following the restrictions on expressive capacity required to preclude the paradoxes, and for present purposes, this amounts to the same thing.

There is a very large variety of formal languages, and it would be quite impossible to perform an analysis comprehensive enough to encompass all of them. Consequently, in keeping with the focus on Tarski, this study will limit itself to the range of languages Tarski limits himself to.[9]

In order to see whether or not the HL as a whole contains any logical flaws, it will be necessary first to go over the restrictions it places on the syntactic and semantic structures of the languages in the hierarchy, and secondly, the consequences of strictly following these restrictions must be considered. The essential thing to be avoided in this enterprise is a ML (or any higher level language) which is treated informally in such a way that any possible logical flaws of the HL could remain hidden in this "informality". The main task of the next two chapters will be to provide a technical explanation of the HL, and then

---

[9] This range is roughly outlined in section two of CTFL. The only substantial additions in later sections are predicates and variables of increasingly higher orders. In his Postscript, Tarski states "I shall consider only those languages in which occur, in addition to the universal and existential quantifiers and the constants of the sentential calculus, only individual names and the variables representing them, as well as constant and variable sentence-forming functors with arbitrary numbers of arguments." (CTFL, p.268)

later on some consequences of strictly following this structure will be examined.

Chapter Two:
The Simple Hierarchy of Languages

Paradoxes are standardly construed as resulting from a set of assumptions, at least one of which must be altered or removed to avoid contradiction. For example, Tarski offers two "essential" assumptions for the construction of the Liar. The first of these is a conjunction of three sub-assumptions:

I (a) the language in which the antinomy is constructed contains names for its own expressions;

(b) it also contains semantic terms such as "true", referring to sentences of this language;

(c) all sentences which determine the adequate usage of the semantic predicates can be asserted in this language.

"A language", says Tarski, "with these properties will be called semantically closed"[1] The second assumption is

II for the language in question, the ordinary laws of logic hold.

Tarski doesn't explain what he means by II, but he dismisses out of hand the possibility of rejecting it, suggesting only that there is no possibility of "changing our logic". Consequently, he decides "not to use any language which is semantically closed in the sense given" (SCT, p.59).

One thing that is immediately worth noting is that,

---

[1]. Tarski, SCT, p. 59.

even though rejecting any of I(a), (b), or (c) could prevent the semantic paradoxes, this solution in and of itself does not have to lead to a hierarchy of languages. All that need follow from restricting the expressive capacity of a language in these ways is that a "silence" would result in which, for example, the semantic attributes of a language simply can't be talked about.[2] A hierarchy of languages is introduced as a means of overcoming the restrictions on the expressive capacity of a language imposed by this solution. The ML, in its ability to talk about the OL, accomplishes what the OL cannot (without leading to contradiction). The ML, in its turn, is limited in its capacity to talk about itself, and thus we have a meta-ML, which can safely talk about the ML, and etc. Thus, properly speaking, it is the restrictions built into the languages in the HL which prevent the paradoxes, and not the hierarchy itself.

Tarski offers three routes (rejecting either 1(a), (b) or (c)) and in SCT at least, he does not specifically choose

---

[2]. The relation of this "silence" to Wittgenstein's notion of showing versus saying in the Tractatus Logico Philosophicus is far from clear; what can be said is that in the Tractatus, Wittgenstein's reasons for the limits on the expressive capacity are related to a theory of representation, and not to the antinomies. This does not, however, preclude a significant relation between the two.

one of the three.[3] Given the presence of different options, it is appropriate to consider some of the more basic ones.

A basic distinction may be drawn between Tarski's 1(a) above and all others. In choosing to reject 1(a), we find that a language cannot refer to its own linguistic expressions at all; "names" in 1(a) above must be construed as including definite descriptions and other forms of reference (as they might be applied to linguistic expressions), for otherwise the paradoxes are not prevented (e.g., "This sentence is false"). This is the simplest form the HL solution to the paradoxes can take, and accordingly, it will be referred to as the SHL - simple hierarchy of languages.[4]

In contrast to the SHL, there are a number of ways of allowing some degree of self-reference into a language without the return of the semantic paradoxes. For the sake of convenience, these will be called "liberal" hierarchies, or LHL's. The solution Tarski uses in CTFL is to reject

---

[3]. Tyler Burge seems to collapse 1(b) and 1(c) into one option, and he says Tarski chose this one. See Burge's "Semantic Paradox" in Martin, ed. pp. 84-85. However, 1(b) and 1(c) need not be so joined.

[4] This solution, as presented by Tarski, is actually a bit too simple; as will be shown below, unless further restrictions are placed on the semantic predicates, contradictions will return at the meta-ML.

assumption 1(b) above; precluding the semantic predicates for a language from being expressions in that same language.[5] Before examining Tarski's own solution, the structure of the SHL will be discussed.

The SHL has, on occasion, been taken as the standard solution to the semantic paradoxes.[6] However, it is not generally used by logicians, except to offer a quick, simplified explanation of how the paradoxes may be avoided. Its main use here is as an introduction to the general structure of HL's.

Since the HL avoids the paradoxes by restricting what a language "talks about" or refers to, the rules governing these restrictions may best be drawn out of an analysis of the name relation. In what follows, 'w' and 'x' will be used as variables ranging over names of expressions, 'Φ' and 'Ψ' will be used as variables ranging over predicates. The sentence form

---

[5] Another form of LHL, for example, proposed by K. Reach, involves precluding the names (and all synonyms of the names) of these same predicates from occurring in the language to which they apply (in effect, rejecting assumption 1 (c) above). See 'The Name Relation and the Logical Antinomies', in The Journal of Symbolic Logic, vol.3, 1938, pp.97-111.

[6] For example: "It follows from the familiar levels of language approach to the paradox [of the liar] that all self referential sentences are illegitimate (ill-formed). For to avoid being ad hoc, such an approach is based generally on the distinction between use and mention of expressions." Martin, 'Towards a solution to the Liar Paradox' in Philosophical Review, LXXVI, 1967, p. 280.

$$x \mathrm{Des} \Phi$$

can be taken as meaning "The name 'x' designates the property Φ. For example, "The name 'human' designates the property of being human."[7]

This name relation, in combination with the Grelling paradox, will be used to examine the SHL.[8] The Grelling, or Heterological paradox, is the easiest to formalize; an expression is said to be _autological_ if it has the property it refers to - thus 'short' is an autological word because it is short, and 'English' is autological because it is an English word. On the other hand, an expression is said to be _heterological_ if it does _not_ have the property it refers to; thus 'long' is heterological because it is not long, etc. The antinomy follows from asking whether 'heterological' is heterological; if it is, then it isn't and if it isn't then it is. The antinomy may be formalized in this manner (using 'Het' for heterological);

---

[7]. My own use of quotation mark names should be explained: single quote marks around an expression indicate the name of that expression. Thus _two single_ quote marks on either side of an expression indicate the name of the name of an expression. For example, ' 'petit' Des small ' means the word 'petit' designates the predicate small, while ' ' 'petit' ' Des 'petit' ' means ' 'petit' ' is the name of the expression 'petit'. _Double_ quotation marks are used when quoting someone.

[8] The examination of the Grelling paradox to follow is based on that of Copi, found in _Symbolic Logic_, fifth edition, 1979, appendix C.

Het(x) = df(∃Φ)[xDesΦ . (ψ) (xDesψ iff ψ=Φ) . ~Φ(x)]

Thus if we ask whether 'Het' is Het we find:

Het ('Het') > 'Het'Des Het . ('Het'DesHet iff Het=Het) .
~Het('Het') EI, UI.

Now, given the truth of both ' 'Het'Des Het ' and ' ('Het'DesHet iff Het=Het ', we can assert:

Het('Het') >~Het('Het')

(by the rule: A > (B . C) ⊢ A > B)

Conversely, starting with

~Het('Het') > ~{(∃Φ)['Het'DesΦ . (ψ) ('Het'Desψ iff ψ=Φ) .
~Φ('Het')]}

>(Φ)~['Het'DesΦ . (ψ)('Het'Desψ iff ψ=Φ) . ~Φ('Het')]
Quantifier Negation.

>~['Het'DesHet . ('Het'DesHet iff Het=Het) .
~Het('Het')] UI.

>~('Het'DesHet) or ~[('Het'DesHet iff Het=Het) .
~Het('Het')]
DeMorgan's

>('Het'DesHet) > ~[('Het'DesHet iff Het=Het) .
~Het('Het')]
Material Impl.

Since ' 'Het'Des Het ' is true we can get

>~[('Het'DesHet iff Het=Het . ~Het('Het')]

>('Het'DesHet iff Het=Het) > Het('Het') DeMorgan's,
Mat'l Impl.

Again, given the truth of '(ψ)('Het'Desψ iff ψ=Het)', we can reduce this to

>Het('Het')

Thus, we have

Het('Het') iff ~Het('Het')

In showing how the SHL avoids this contradiction, it is helpful to make use of what is called an "absolute" OL[9]. An absolute OL, in contrast to a "relative" OL, is one in which there are no names of expressions at all (a relative OL is one which refers to itself or some other language(s), but is the object of discussion in a particular context).[10] If we assume that the first level in the SHL is an absolute OL, then clearly the Het paradox cannot occur in this OL. In such a language, the name relation would not occur because there are no terms in the language which fall within the range of the variable x in 'xDesΦ', i.e., there are no names of expressions in the language. This holds for the Het predicate as well, and for all such semantic predicates. Consequently, none of the semantic paradoxes are derivable in an absolute OL.

In the ML, it is possible to name all expressions of the OL, but none of the expressions of the ML itself. The ML also contains translations, or synonyms, of all OL expressions.[11] In this case, the ML contains the

---

[9] If we reject Tarski's assumption I (a) - as the SHL does - then either the SHL begins with an absolute OL, or it recedes to infinity downwards, as well as upwards.

[10]. See The Dictionary of Philosophy, A. Flew, p. 253.

[11]. In SCT Tarski suggests the ML may contain the OL (pp.60-61).

predicate Het, but since the only names it contains are names of OL expressions (and names of things), and it contains no name for the Het predicate itself, no contradiction can be derived in the ML.

However, the situation gets rather more complicated at the next level, the meta-ML, and it will be necessary to introduce a small amount of specialized notation to explain the matter. First, it will be convenient to have a means of referring to a given level of language without having to place any number of "meta's" in front of it. The following abbreviations will be used to accomplish this: OL and ML will be used as before, and the meta-metalanguage will be represented by ML-2, the meta-meta-metalanguage by ML-3, etc. The expression 'ML-k' (where k is greater than or equal to 0) will be used to refer to any level of language in a hierarchy (the OL and ML may also be treated as ML-0 and ML-1 respectively).

The second piece of notation is essential for avoiding the paradoxes at the ML-2 and higher levels. The name relation introduced earlier as

$x\mathrm{Des}\Phi$

will have a subscript attached to it such that for

$x\mathrm{Des}_k\Phi$,

'k' indicates that the range of x is <u>names</u> of expressions of the <u>ML-k</u>. This subscript will apply to all semantic predicates.[12] The need for such a subscript can be seen by looking at the consequences of not having it in the ML-2. First of all, the ML-2 contains names for all OL and all ML expressions, and translations for them as well. As with all levels in the SHL it contains no names for its own expressions. However, because it does contain names for all ML expressions, it also contains a name for the Het predicate which occurs in the ML. With this, it would be possible to derive the contradiction, by substituting this name into the definition of the Het predicate as it occurs in the ML-2.

The contradiction is prevented here by means of the restriction underlying the subscript introduced above. The effect of this subscript is to create two Des relations in the ML-2, Des_0 and Des_1, such that in

$x Des_0 \Phi$

the range of the variable x is names of expressions of the ML-0 (i.e., the OL), and in

$x Des_1 \Phi$

the range of x is names of expressions of the ML-1. Thus the Des relation used in the ML-1 is actually a translation of Des_0, and Des_1 does not occur in the ML-1 because the ML-1 contains no names for its own expressions (similarly Des_2

---

[12]. A specific definition of `semantic predicate' can be found on p. 252 of CTFL.

does not occur in the ML-2, because the latter does not contain names for its own expressions). Corresponding to these two name relations, there are also two Het predicates, Het_0 and Het_1. $Het_0(x)$ means roughly that the <u>OL</u> expression named by a given value of x does not possess the property which it names. Similarly, $Het_1(w)$ means that the ML expression named by a given value of w does not possess the property which it names. The definitions of these two predicates display this difference in meaning clearly:

$$Het_0(x) =_{df} (\exists\Phi)\ x\ Des_0\Phi\ .\ (\psi)(x Des_0 \psi \ \underline{\text{iff}}\ \psi=\Phi)\ .\ {\sim}\Phi(x)$$

$$Het_1(w) =_{df} (\exists\Phi)\ w\ Des_1\Phi\ .\ (\psi)(w Des_1 \psi \ \underline{\text{iff}}\ \psi=\Phi)\ .\ {\sim}\Phi(x)$$

Het_0 cannot be defined in terms of Des_1 (nor Het_1 in terms of Des_0) because the ranges of the arguments in the functions are names of expressions of different languages.[13]

Now, no contradiction can be derived from Het_0 because the range of its argument is names of expressions of the OL, and there is no Het function in the OL to be named. However, the ML does contain a version of the Het function, and since the ML-2 can name all expressions of the ML, it is

---

[13]. Perhaps a clearer way of expressing this is to say that extensionally, they are by definition completely different sets, given the restrictions imposed by the subscripts.

possible to substitute ' 'Het' '[14] for w in $Het_1(w)$.

Thus we have:

Het_1('Het') >'Het'$Des_1\Phi$.(Ψ)('Het'$Des_1\Psi$ <u>iff</u> $\Psi=\Phi$) .

~Φ('Het')

There are two predicates, Het_0 and Het_1, which may be substituted for Ψ (and Φ). Using Het_0 we get

Het_1('Het') > 'Het' Des_1 Het_0 . ~Het_0('Het')

The expression ' 'Het'Des_1 Het_0 ' is true, since Het_0 is a translation of the Het function found at the ML-1, and consequently we get

Het_1('Het') > ~Het_0('Het')

This, however, leads to no contradiction. Given the nature of the restrictions emplaced by the subscript, the conditional may be construed as either true or meaningless. The consequent of the conditional, ' ~Het_0('Het') ' is either true or ill-formed; Het_0 is by definition satisfied only by names of OL expressions, and here we have substituted the name of an ML-1 expression. Whether this is said to be a true expression or an ill-formed expression is a side issue; the fact is that no contradiction results.

If Het_1 is substituted for Φ, we get

Het_1('Het') > 'Het'Des_1Het_1 . ~Het_1('Het')

Now, unlike the previous case, nothing is ill-formed here;

---

[14]. Where this is the name of the Het predicate operating in the ML.

' 'Het' ' is one of the values which can satisfy Het_1 (while it cannot satisfy Het_0 above). From this statement, it is possible to assert

Het_1('Het') > ~Het_1('Het')

(recall the rule: A > (B . C) ⊢ A > B). This, however, is only half of the biconditional needed to form the contradiction; the converse of this would require the ability to assert that 'Het'Des_1Het_1. But given the rules governing the subscripts, ' 'Het' ' refers to a function which first occurs in the ML-1, while Het_1 has no synonym in the ML-1. Consequently, it is simply <u>false</u> that 'Het'Des_1Het_1. If ' 'Het' Des_1 Het_1 ' were true, the contradiction could be derived, however, this expression could only be true if the subscripts were removed. We are still left with the conditional

Het_1('Het') > ~Het_1('Het').

This is not a contradiction; given the meaning of this conditional in formal logic, it simply indicates that the antecedent of the conditional must be <u>false</u> i.e., that ~Het_1('Het') is true (a false statement can imply both true and false statements).

This method of preventing the Grelling paradox can be generalized to any levels in the SHL: taking j and k as representing any given ML-j and ML-k, we can employ the following general formula, based on the definition of Het:

Het_j ('Het_k') <u>iff</u> 'Het_k' Des_j Het_k . ~Het_k('Het_k')

The first thing to be noticed is that j must be greater than k, for otherwise the expressions 'Het_j ('Het_k')' and ' 'Het_k' Des_j Het_k ' are not well-formed. The expression ' 'Het_k' Des_j Het_k ' is obviously true (if j > k), and so we derive the biconditional

Het_j ('Het_k') <u>iff</u> ~Het_k('Het_k')

This is not a contradiction; it merely states that '$Heterological_k$' is $Heterological_j$ if, and only if, '$Heterological_k$' is <u>not</u> $Heterological_k$. In point of fact, 'Het_k' <u>cannot</u> be Het_k because the predicate Het_k takes as the range of its argument names of expressions of the ML-k, and the predicate Het_k does not occur in the ML-k.

The restrictions introduced above appear to be sufficient to avoid the semantic paradoxes as a whole. The rules governing these restrictions may be formulated as follows:

(1) Names for ML-k expressions can be contained <u>only</u> in languages ML-j where j>k (k is greater than or equal to zero)

(2) For any semantic predicate (Heterological, Designates, True ...), the range of values for the named expressions is expressions of <u>only one</u> language, and that language must be of <u>lower</u> level than the level of the language in which that semantic predicate first

occurs. (Thus there can be no semantic predicates which can take names of expressions of the language in which these semantic predicates occur).

Rule (1) prevents the contradiction from arising within a given level of language, and (2) prevents them from arising between levels. This second restriction only becomes apparent in the ML-2 and higher levels, and it embodies the restriction placed by the subscripts attached to the semantic predicates. Provided the "spirit" of these rules is not violated, the vocabulary and syntax of the languages within the hierarchy may be of any number of forms.

Rule (2) operates by restricting the range of the variable(s) of semantic predicates to one level of language (thus $Het_3(x)$ is attributable only to expressions of the ML-3, $xDes_2\Phi$ correlates only expressions of the ML-2 to their designata). A significant consequence of this is that extensionally, the semantic predicates for the different levels of language are completely distinct, by definition. For example, at the ML-3 there will be three distinct truth predicates, T_0, T_1, T_2; each attributable to different sets of sentences. There appear to be no problems with partially "collapsing" the semantic predicates in the following manner. As rule (2) indicates, the predicate Het_3 takes as the range of its argument only names of expressions of the ML-3, but the ML-3 itself either contains translations of all expressions of ML-n, n<3, or it actually contains these languages. Thus, Het_3 is effectively capable of being satisfied by names of expressions of all ML-n where $n \leq 3$. Thus the semantic predicates can be collapsed to the extent that the extension of a given predicate with subscript n may be a <u>subset</u> of the extension of a "parallel"[15] predicate with subscript x where n<x. Allowing this collapse of the hierarchy does not reintroduce the paradoxes because the predicates remain distinct due to the increasing range of their application. For example, Het_n('Het_n') is still prohibited by rules (1) and (2), while any other combination likely to produce a contradiction will prove <u>false</u> with respect to ' 'Het_x'$Des_n Het_y$ ' where $x \neq y$.

This method of avoiding the semantic paradoxes is based on rejecting Tarski's assumption 1 (a) stated at the start of this chapter. It is not based on the theory of types; the restrictions enforced by rules (1) and (2) do not prevent the syntactic paradoxes. For example, given only rules (1) and (2) it is still possible to ask whether the property of being impredicable is impredicable (Russell's paradox), because there is no use of the <u>name</u> of the predicate. Rules (1) and (2) only restrict the use of names of expressions. Consequently, in addition to rules (1) and (2), it is also

---

[15]. For example, Het_2 and Het_3 are parallel predicates, and Het_2 can be interpreted as a subset of Het_3.

necessary to incorporate other means (such as the theory of types) in order to preclude the syntactic paradoxes.[16]

This is substantially different from the standard HL to be explained in chapter three. Contrary to Martin's claim (fn. 6, this chapter) that this SHL avoids being ad hoc, this version actually appears more ad hoc than the standard view. While the SHL requires rules 1 and 2 plus the simple theory of types (or some other method of preventing the syntactic paradoxes), the standard solution is a direct consequence of the theory of types and the specific nature of certain semantic concepts such as satisfaction, as will be seen in the next chapter.

---

[16] See footnote 6 in chapter 1.

Chapter Three: The Tarskian Hierarchy of Languages

As was explained in chapter two, the SHL is not a system which has been commonly advocated by logicians. The presentation of the SHL was intended as a base from which to develop a more complex hierarchy which allows a greater expressive capacity (the "standard" version of the HL). This is the goal of the present chapter.

The first and foremost source of such a system is to be found in Tarski's 'The Concept of Truth in Formalized Languages' (CTFL)[1]. Tarski's CTFL is "almost wholly devoted to a single problem - <u>the definition of truth</u>" (p. 152). He introduces and elucidates the concept of an HL only to the extent necessary to allow a "formally correct" definition of truth. In the CTFL there is no single prescription regarding the "best" form of HL for avoiding the semantic paradoxes, nor is there a succinct, detailed explanation of the structure of Tarski's HL, or how it avoids the paradoxes. Rather, Tarski provides a brief analysis of the causes of the paradoxes in natural language, and some general comments as to the nature of an HL. Since his goal is the construction of a consistent semantics for a

---

[1] All page references in this chapter are to CTFL, unless otherwise indicated.

language, it is only natural that he presents no single hierarchy. However, a fairly generic, "Tarskian" hierarchy can be derived from his work, as will be seen by the end of this chapter.

In order to arrive at this "Tarskian" hierarchy, it will be necessary to take a fairly close look at exactly what Tarski does in CTFL.

Having initially stated his goal as providing a materially adequate and formally correct definition of truth, Tarski looks into the possibility of finding such a definition in natural language. After considering a number of other problems, he then introduces "important arguments of a general nature" [i.e., the paradoxes] which support his contention that no such definition of truth is possible in natural language. In the light of the paradox of the Liar he makes the following claim:

> ... no language can exist for which the usual laws of logic hold and which at the same time satisfies the following conditions: (I) for any sentence which occurs in the language a definite name of this sentence also belongs to the language; (II) every expression formed from [x is a true sentence if and only if p] by replacing the symbol 'p' by any sentence of the language and the symbol 'x' by a name of this sentence is to be regarded as a true sentence of the language ...(p.165)[2]

---

[2]. A third "empirical" condition is also included, which need not be considered here. It amounts to stating that a form of the Liar paradox is a sentence of the language. These conditions are effectively a special instance of the conditions quoted at the beginning of chapter two.

In a footnote here Tarski makes an addition to these conditions, considering the "heterological" paradox, such that the conditions differ from those above "in that they treat not of sentences but of names, and not of the truth of sentences but of the relation of denoting." (p. 165) In and of themselves, these conditions are rather weak (in the present context), since from them we may conclude only that _if_ conditions I and II are satisfied, _then_ that language is inconsistent. This is simply a material implication; where the antecedent is _false_, and conditions I and II are _not_ met, we still cannot conclude that a language will avoid the semantic paradoxes.

Tarski's point in formulating these conditions is simply to preclude natural languages from further consideration in his pursuit of a definition of truth. A natural language is able to name the expressions found in it, and consequently, is always able to violate the conditions quoted above. Tarski proceeds to focus his attention on formalized languages only. "Formalized languages", he says, "can be roughly characterized as artificially constructed languages in which the sense of every expression is uniquely determined by its form" (pp. 165-66). Furthermore, "the formalized languages do not have the universality [of natural languages]" (p. 167). In saying they are not universal, Tarski is pointing out that restrictions are placed on their expressive capacity. The question of exactly what these restrictions are, however, is not clearly explained. The final "general" comments Tarski makes in this direction are that, first, "most [formal] languages possess no terms belonging to the theory of language, i.e., no expressions which denote signs and expressions of the same or another language or which describe the connections between them." (p. 167) Thus, he says, we must distinguish "the language _about_ which we speak and the language _in_ which we speak ...." Finally, he explains that

> the names of the expressions of the first language, and of the relations between them, belong to the second language, called the _meta-language_ (which may contain the first as a part). The description of these expressions, the definition of the complicated concepts, especially of those connected with the construction of a deductive theory (like the concept of consequence, of provable sentence, possibly of true sentence), the determination of the properties of these concepts, is the task of the second theory which we call the _metatheory_. (p. 167)

I have quoted at length partially because of the importance of these comments, and partially because they are just about all Tarski _explicitly_ has to say about the HL and how it avoids the paradoxes.

Beyond this it will be necessary to take a careful look at the specific language systems Tarski uses in CTFL, and glean what detail we can about his HL from them. One of the most significant ambiguities to come out of his general comments is the issue of interpreting an ML in an OL. Prima facie, such pronouncements as on p. 167 appear to <u>preclude</u> the possibility of interpreting the ML in the OL, as would be the case in an SHL. However, this clearly is not Tarski's intent.[3] The issue that is not immediately clarified in CTFL is the extent to which such OL interpretation of the ML is possible without contradiction, and special care must be taken in this regard.

In section 2 of CTFL, Tarski's goal is to outline the basic structure for a specific formal language he wishes to study, and its metatheory (which is never completely formalized). While outlining symbols for his OL and ML, he makes use of natural language. There appear to be two possible interpretations of this use. First, Tarski may be assuming that the portion of natural language which he uses itself need not involve any contradictions and so is formalizable. Second he may intend this use of natural language to be <u>outside</u> or beyond his ML, and not part of it at all. Considering the care with which he goes about

---

[3] See, for example, p. 262: "The languages of infinite order, thanks to the variety of meaningful expressions contained in them, provide sufficient means for the formulation of every sentence belonging to the arithmetic of natural numbers and consequently enable the metalanguage to be interpreted in the [object] language itself." The method of interpretation Tarski refers to here is arithmetization: see fn. 1, p.247. This issue will be dealt with later this chapter.

setting up notation for the ML, this second interpretation seems more likely. Tarski's only comment on the issue is to say that aside from some terms relating the ML to the OL, "the process of formalizing the metatheory shows no specific peculiarity" (p.170). In either case, he does make extensive use of natural language in explaining his system.

Tarski's OL is a simple variation of the calculus of classes. The language includes the following four constants: 'N' for negation; 'A' for disjunction; 'Π' for the universal quantifier and 'I' for set inclusion. Finally, for variables, he says he will "use for variables only such symbols as 'x_1', 'x_{11}', 'x_{111}', and analogous signs which consist of the symbol 'x' and a number of small strokes added below." (p. 169) These variables are taken to represent "names of classes of individuals".

The ML contains <u>names</u> for OL expressions, as well as translations of them, as indicated below:

OL	ML name	ML transl
'N'	'ng'	'not'
'A'	'sm'	'or'
'I'	'in'	'is included in' or '$\subseteq$'
'Π'	'un'	'for all'
'x_1'	'v_1	'a'

Other ML symbols include 'a' and 'b', which range over classes of individuals, 'f', 'g', 'h', which range over sequences of such classes, 't', 'u', ' w', ' x', 'y', and 'z', which range over names of OL expressions and sequences of expressions. To indicate concatenation of names in the ML, '^' is used. Thus the structural descriptive name of 'NIx_1x_{11}' is '$((ng\text{^}in)\text{^}v_1)\text{^}v_2$',[4] and its translation is 'b is not included in a'. The ML also contains expressions for identity, set membership, and much more (see pp. 170-173 for a more complete list).

Tarski's presentation of his notation is somewhat less clear than this; he appears to have set up certain symbols at one point, and then started up a more systematic approach, as outlined above. On p. 169 he says that he will "use as variables only such symbols as 'x_1', 'x_{11}', 'x_{111}'..." in his OL. However, he then introduces expressions such as '$\prod xp$' which reads 'for all classes x we have p' - which contain variables which range over sentences or sentential functions ('p' and 'q') or over variables ('x' and 'y'). He goes on to mention that "the proper domain of the following considerations is not the language of the calculus of classes itself but the corresponding metalanguage" (p.169). It is unclear whether Tarski intended 'p', 'q', 'x', and 'y', as introduced here to be part of his

---

[4]. Using Definitions 1 and 2 (p. 175) a simpler name for this expression is '$\overline{\iota_{1,2}}$'. '$\bigcap_2$' is defined (definition 6) as a simplified version of '$un\text{^}v_2$', and '+' is used to indicate disjunction (definition 3). The relation between "x's" and "v's" is maintained throughout CTFL - i.e., the former being OL variables, the latter being their corresponding structural-descriptive ML names.

object language or his metalanguage.[5] In any case, these symbols do not play a major role in his formulas, and so this should not pose a real problem.

One point which becomes evident when going through Tarski's presentation is that he makes use of two different ways of naming OL expressions; he uses his own "structural-descriptive" names (which form part of his ML), as well as conventional quotation mark names. This lends support to the opinion that his use of natural language is not intended to be part of his ML, especially when one considers that he consistently uses <u>only</u> his structural-descriptive names in formulating his axioms and definitions in the metatheory. His use of quotation mark names appears to be intended to help us understand the complex issues being introduced and discussed, since quotation mark names are intuitively obvious, and sole use of his structural-descriptive names would be rather confusing at times.

There are one or two instances, however, where he seems to inappropriately mix the two different ways of naming OL expressions. For example, on pp. 178-179, he makes reference to certain OL axioms such as '<u>ANAppp</u>', using the symbol 'p' as a sentential variable ranging over OL

---

[5] It would be interesting, if it were possible, to examine Tarski's early drafts or working papers for CTFL to see if these different notations indicate different approaches to his topic.

expressions. Now, the expression ' '<u>ANAppp</u>' ' is a quotation mark name, and according to his own line of argument (p. 159) such names <u>cannot</u> contain variables. Unfortunately, 'p' as used above, is intended to be exactly that; a variable ranging over sentences of the object language.

The problem is solved simply enough though, by making use of Tarski's ML names. Thus the OL axiom in question should be named (in the ML) as '((((sm^ng)^sm)^y)^y)^y'.[6]

The OL used in sections 2 and 3 does not contain any terms which refer to linguistic expressions. This OL contains only logical operators, and variables whose extension is classes of individuals. Thus between this OL and its ML there exists a situation similar to that found in the SHL discussed in chapter 2. The issue of allowing partial self reference, then, only occurs at higher levels, or when OL's of greater expressive power are considered. Tarski makes few references to either strictly formalizing this ML, or to its relation to an ML-2 (see for example pp. 170, 188 (fn), 195, 246, 249 (fn. 1)), and so the most

---

[6]. Or more simply, using definitions 2 and 3 (p. 175) '$\overline{y+y}$ +y '. 'y' is used as a variable ranging over names of OL expressions (see CTFL, p.173, or fn. 4, this chapter) and consequently it can be used as a name of the expression which would be instantiated for p.

There is one minor problem with this solution, however: 'y' is to range over names of any OL expression, while 'p' ranges over only sentences or sentential functions. This problem could be corrected with minor modifications.

fruitful source of information is to be found in sections 4, 5 and 6, where he considers far more complex OL's.

It is all but impossible, however, to analyze the structure of Tarski's hierarchy without also examining the routes he must follow in constructing his definition of truth, and maintaining this definition in languages of increasing complexity. Most of his references to the hierarchy are in relation to the development of the definitions of truth and satisfaction.

In section 4 Tarski differentiates four types of languages. To understand these distinctions, it is necessary clearly to understand the concepts of <u>semantical category</u> and <u>order</u>. Semantical category is explained as being an "extension" of the notion of type in the simple theory of types (p. 215, fn.). All expressions of an interpreted language which are parts of sentential functions can be divided into mutually exclusive semantical categories. For example: names of individuals, names of classes of individuals, names of two-termed relations between individuals, and sentential functions themselves (p. 216). Semantical categories are classified on the basis of their <u>order</u>.[7] He offers this convention regarding 'order'

---

[7]. Tarski stresses that his definitions of semantic category and order are not complete; quantifiers and logical constants, for example, are not adequately explained here. Cf. p. 218, and 218, fn.

(as it applies to variables):

> (1) the first order is assigned only to the names of individuals and to the variables representing them; (2) among expressions of the $n+1^{th}$ order, where n is any natural number, we include the functors of all those primitive functions all of whose arguments are of at most the n^{th} order, where at least one of them must be of the n^{th} order. (p. 218)

Thus every expression of a given semantical category has one and the same order, but one order may contain many semantical categories (for example, names of classes of individuals, and names of two-termed relations between individuals are both expressions of order 2, but are different semantical categories (p. 219)).

The four categories of languages Tarski identifies are as follows. There are languages in which variables:

(1) all belong to one and the same semantical category;

(2) are of more than one but still a finite number of categories, and a finite order;

(3) are of infinitely many semantic categories, but are of finite order;

(4) are of arbitrarily high (infinite) order (p. 220).

Tarski distinguishes languages on this basis because each different level of complexity requires more and more complex apparatus in the metatheory for clearly fixing the meaning of semantic concepts such as satisfaction and truth. By first addressing the simpler cases, he is able to build new apparatus into his metatheory to deal with complications as they arise in more complex OL's.

Though there is no need to go through this work in detail, a basic understanding of the concept of "essential richness", as well as the significance Tarski attaches to the concept of satisfaction, is appropriate. Through an examination of these issues as they arise in more complex OL's, it will be possible to address the question of how much self-reference is possible in Tarski's hierarchy. The OL Tarski uses in sections 1 and 2 fits into group (1) above. With this OL, he is unable to construct a general definition of truth firstly due to the mechanical restrictions of quotation mark names. For example,

'(x)('x' $\in$ Tr if and only if x)'

fails because, according to Tarski, a name such as ' 'x' ' names the 24th letter of the alphabet, and cannot remain a variable within quote marks (p. 159 ff). Secondly, a general definition of truth might be possible if all sentences of the form 'x is true if and only if p' (where 'x' is the structural descriptive name of sentence p) could be listed. However, the OL in use contains infinitely many variables and thus potentially infinitely many sentences, and so such a list is not possible.[8] Finally, Tarski also finds it impossible to define truth directly by recursive means. A

---

[8] See chapter four for more detail on this definition.

recursive definition would first define truth conditions for simple sentences of the OL, and then outline how the truth of composite sentences is a function of simple sentences. However, this won't work because "in general composite sentences are in no way composed of simple <u>sentences</u>" (p. 189). The elementary formulas in all of Tarski's OL's are sentential functions, containing free variables - and thus are neither true nor false. Sentences are constructed by binding all variables in a sentential function by quantification (Cf. definitions 10, 11, 12, pp. 177-178). A recursive definition cannot be based on the truth conditions of the elementary sentential functions, because they are neither true nor false.

It is possible, however, for a sentential function to be <u>satisfied</u>, or not satisfied, by objects. Tarski thus proceeds to define truth <u>indirectly</u>, using the notion of satisfaction of a sentential function by given objects (in this case the "objects" are classes of individuals). As an intuitive example of the satisfaction relation, Tarski offers the following: "for all a, a satisfies the sentential function 'x is white' if and only if a is white" - thus snow satisfies the sentential function 'x is white'. An example from within his ML is "for all a, a satisfies the sentential function $\cap_2 \iota_{1,2}$ if and only if for all classes b we have a $\subseteq$ b"[9] (p. 190). In both of these examples, there is only one free variable in the sentential function. Where there are *two* free variables, we have a relation between an ordered pair (of "objects") and a sentential function. For a sentential function with an arbitrary number (say n) of free variables, it is necessary to have an ordered n-tuple, or a *sequence* of objects to satisfy it. With this in mind Tarski says "For the sake of a uniform mode of expression[10] we shall from now on not say that given objects but that *a given infinite sequence of objects satisfies a given sentential function* " (p. 191). Since we are dealing with *ordered* sequences of objects, it is possible to correlate each free variable in a sentential function with a distinct object of every sequence.[11] The

---

[9]. Recall '$\bigcap_2 \iota_{1,2}$' is the ML name for the OL expression '$\prod x_{11}$ Ix_1 x_{11}', and the translation of this OL expression is 'for all classes b we have a $\subseteq$ b'.

[10]. By "a uniform mode of expression" Tarski means that, by using infinite sequences of objects, he will be able to use only one satisfaction relation. If the length of the sequence were to correspond to the number of free variables in a given sentential function, he would effectively have an infinite number of satisfaction relations, each of different semantic category, though all of the same order. Sentential functions with one free variable would have a satisfaction relation of different semantical category from the satisfaction relation for sentential functions with two free variables, etc. (Cf. p. 224, and pp. 225-26). Further discussion of this follows below.

[11] At this point Tarski introduces a modification to his notation to permit this correlation. As introduced on p.173,

sequences are infinite so that there will always be enough objects to correlate with the variables in a given sentential function. The excess objects are simply to be disregarded. As an example Tarski again offers the sentential function '$\bigcap_2 \iota_{1,2}$' (i.e., for all classes b, a $\subseteq$ b). This function contains only one free variable, so only the first objects (which are, in this case, classes) of sequences will be used. In this example, "we say that the infinite sequence f of classes satisfies the sentential function

$\bigcap_2 \iota_{1,2}$ if and only if the class f_1 satisfies this function ... [where] for all classes b, we have $f_1 \subseteq$ b" (pp. 191–92). This formula is generalized in the following scheme:[12]

> f satisfies the sentential function x if and only if f is an infinite sequence of classes, and p. (p. 192)

In this formula 'x' is replaced by an ML structural

'f' is a ML variable ranging over sequences of classes, but now he adds 'f_k', where k is any natural number, such that 'f_k' ranges over classes of individuals (just as 'a' and 'b'), and "the term f_k will be correlated with the variable v_k" – p. 191. 'v_k' is the ML name of an OL variable with k strokes after it.

[12]. This scheme will have a significant impact on the concept of essential richness, to be discussed later, specifically with regard to the notion of "semantical correlation" between free variables in a sentential function and the names of objects which satisfy that sentential function (see p. 48 ff below).

descriptive name of an OL sentential function. All free variables occurring in this sentential function (e.g., v_k, v_l, where k and l are natural numbers) are replaced by the corresponding classes in the sequence f (e.g., f_k, f_l) and 'p' is replaced by the resulting instantiation of the sentential function (or, more exactly, its translation in the ML).

Using this scheme, Tarski then states the conditions for the satisfaction of any elementary sentential function (e.g., $\iota_{k,l}$; an inclusion), and inductively from this, the satisfaction conditions for negation, disjunction, and quantification (Cf. Definition 22, p. 193). Thus he completes the recursive definition of satisfaction of sentential functions. With this semantic concept defined, Tarski claims it is possible to define "a whole series of concepts in this field [i.e., semantics]" (pp. 193–94). Included among these are the concepts of denotation, definability and truth.

To define 'true sentence' Tarski need only point out that a <u>sentence</u> can be interpreted as a sentential function with <u>no free variables</u>; all variables have been quantified, and thus either <u>all</u> sequences will satisfy it, or none will.[13] The indirect definition of truth then, is simply "x is a true sentence - in symbols $x \in Tr$ - if and only if $x \in S$ and every infinite sequence of classes satisfies x" (Cf. Definition 23, p. 195). '$x \in S$' means x is a sentential function with no free variables, i.e., a sentence (Definition 12, p. 178).

When we come to more complex OL's, those with variables of more than one semantical category (but still finite in number), and those with an infinite number of semantical categories but still of finite order, Tarski is required to introduce much more complex notions of satisfaction upon which to base a definition of truth. Two problems present themselves. The first has already been addressed; namely the use of infinite sequences to avoid having satisfaction relations of different semantical categories for sentential functions with different numbers of free variables (Cf. footnote 10 of this chapter). The second problem is unique to the more complex OL's, and results from the presence of free variables of differing semantical categories.

To understand this problem, and its solution, it is useful to start by considering the extension of the satisfaction relation used in sections 2 and 3 of CTFL

---

[13]. For example "for all classes a..." can only be satisfied where all classes fit the conditions involved - and thus all sequences must fit them also.

(i.e., where all variables are of the same semantical category). There, the satisfaction relation is a set of ordered pairs, where the domain is "objects" (in that case sets of individuals) and the range, or counter domain is sentential functions which are satisfied by the objects listed in the domain. For example,

$$\langle a, \iota_{1,1} \rangle \in Sat.$$

i.e., the class a satisfies the sentential function $(I^\wedge v_1)^\wedge v_1$ – (a is included in itself). A problem was encountered there in that sentential functions with more than one free variable required ordered sequences of objects to satisfy them. For example,

$$\langle \langle a,b \rangle, \iota_{1,2} + \overline{\iota_{1,2}} \rangle \in Sat$$

(i.e., either set v_1 is included in set v_2 or it is not – my example). Where an ordered pair of objects is required we end up with a satisfaction relation which is of a different semantical category from the satisfaction relation required when only one free variable is present in the sentential function. In sections 2 and 3, for the calculus of classes OL, this problem is avoided by the use of infinite sequences of objects regardless of the number of free variables in a sentential function.[14]

---

[14]. Tarski admits that if the ML were to be strictly formalized there would be difficulties maintaining this "intuitive" approach, and that "if we had formalized the metalanguage it would have been necessary to use infinitely

Now, once variables of more than one semantical category occur in sentential functions of the OL, not only is the number of free variables in a sentential function a problem, but the differing semantical categories of the free variables becomes an even larger problem.[15] Tarski considers it in the context of two increasingly complex OL's, the first belonging to the second type of language, the other belonging to the third type. The first is the logic of two-termed relations, consisting of variables representing names of individuals (first order) such as 'x_1' 'x_{11}', etc., and variables representing names of two-termed relations between individuals (second order) such as 'X_1', 'X_{11}', etc. The second OL he uses is the logic of many termed relations, which is similar to the logic of two termed relations, except that instead of a finite number of second order semantical categories (in the case of the logic of two termed relations there is only one second order semantical category), there are an infinite number of second order semantical categories (e.g., one for one-termed relations, another for two-termed relations, and so on).

---

many distinct terms instead of one term `satisfies'." (p. 224)

[15]. The entire problem disappears for Tarski once he allows formal languages not based on the theory of semantical categories. (Cf. p. 272) I will provide an explanation of the problem and its solution simply for the purpose of exegesis of CTFL.

For both of these OL's, there are variables of different semantical categories. Thus, in both cases there are sentential functions with variables of different semantical categories. For example, 'Xyz', where 'X' may be replaced by any variable of the second order, and 'y' and 'z' can be replaced by variables of the first order (the expression reads: "individual y stands in relation X to individual z" - cf. p. 223).

The significant difficulty here results from what Tarski calls the "semantic correlation" between the free variables of the sentential function and the objects which satisfy the sentential function. This semantic correlation has a significance beyond this specific problem because, coupled with the theory of orders, it forms the basis of the concept of "essential richness". Specifically, the free variables and the <u>names of the objects</u> which satisfy the sentential function in which the variables occur must be of the <u>same semantical category</u> (p. 224). This is so <u>because the variables ranging over classes of objects must fill the place of the free variables when substituted for them in the sentential function</u>.

Now, when we look again at the extension of the satisfaction relation, we find a relation between sequences of objects of more than one order, and a sentential function. However, "the theory of semantical categories does not permit the existence of such heterogeneous sequences." "Consequently", Tarski says, "the whole conception collapses" (p. 226).

Some examples may prove useful here. To begin with, recall the examples mentioned previously, for an OL with variables of only one semantical category:

$$< \langle a,b \rangle, \ \iota_{1,2} + \iota_{1,2} > \in \text{Sat}.$$

This ordered pair consists of 1- an ordered pair of objects (properly speaking it should be an infinite ordered sequence), and 2- the name of a sentential function. The variables ranging over the classes of objects, i.e., 'a' and 'b' must be of the same semantical category as the free variables in the sentential function named in the ordered pair, because their substitution into the sentential function results in a well-formed expression. In this example there is no problem because 'a' and 'b' are of the same semantical category. However, take the expression

$$p_{1,2,3} + \overline{p_{1,2,3}}$$ [16]

To say that certain individuals G and H and relation R satisfy this sentential function would amount to saying that

$$< \langle R,a,b \rangle \ , \ p_{1,2,3} + \overline{p_{1,2,3}} \ > \in \text{Sat}.$$

However, the ordered sequence '<R,a,b>' is not permitted by

---

[16]. '$p_{1,2,3}$' names the OL expression '$X_1 x_{11} x_{111}$'. The entire expression reads "either individual x_{11} stands in relation X_1 to individual x_{111} or it does not." See p. 225.

the theory of semantical categories because it is a heterogeneous sequence of first and second order variables, and thus Tarski must make use of more complicated logical procedures.

The only solution Tarski finds for this problem which works for languages of both types 2 and 3 is the method of "semantical unification of the variables" (p. 228). Basically, this unification involves correlating in 1-1 fashion each individual to a unique two-termed relation. For example individual a is correlated with a*, where a* is an ordered pair "whose terms are identical with a, i.e., the relation R which holds between any two individuals b and c if and only if b=a and c=a." (p. 228) Thus lower order expressions can be "raised" to the higher order(s) found in a given language. The unifying category must be of the highest order found in a given OL, since all lower order variables are raised to the highest order (cf, p. 230, p. 235).

It is worth noting, also, that just as using infinite sequences of objects was an informal, intuitive solution, so too we now find that "all the variables of the language belong to one and the same semantical category, not indeed from the formal but from the intuitive point of view ..." (p. 229). To allow such correlations formally could presumably allow Russell's "impredicable" paradox to reoccur. There is a difference, however, between this informality and the informality of using infinite sequences of objects. With the problem of infinite sequences, a completely formalized ML could still contain the semantic concept of truth, but _many_ truth predicates would be required, instead of one. In this case however, without a way of avoiding heterogeneous sequences, no definition of satisfaction is possible for Tarski without abandoning the theory of semantic categories.

In any case, with a basic understanding of the concept of a unifying category, we are one step away from completing the rationale behind the notion of essential richness. Thankfully, too, it is a relatively simple step.[17] Satisfaction is a two termed relation with a domain of sequences and a range of sentential functions. Given the "semantic correlation" between free variables in sentential functions and names of objects satisfying these sentential functions (outlined above), and the fact that the unifying semantical category is of the highest order found in a given OL, we can conclude that the domain of the satisfaction relation for that OL will be of the highest order found in that OL. Consequently, given the definition of 'order' (p.

---

[17]. For all that this final step appears necessary in explaining the need for essential richness, Tarski does not explicitly state it.

218) the satisfaction relation for any specific OL must be one order higher than any found in the OL. For example, if an OL's variables are of order n or lower, then the domain of the satisfaction relation for that OL will consist of names of objects of order n, and therefore the satisfaction relation will be of order n+1. Since the truth predicate (or any other semantic predicate) is defined using the satisfaction predicate, this condition of "essential richness" applies to it also. When Tarski says that the ML must be essentially richer than the OL, this is what the claim amounts to; the ML contains predicates of higher order than any found in the OL.

The reasons why essential richness is necessary alter in the Postscript, where Tarski abandons the need for the theory of semantic categories. However, the significance of essential richness does not change. In the Postscript, written five years later, Tarski admits the validity of formal languages which are not based on the theory of semantic categories. Consequently, "we can freely operate with sequences whose terms are of different orders" (p. 272), and there is no need for the complicated devices outlined in section 4 (many-rowed sequences, and the unifying semantic category). However, for all that the theory of semantic categories need not be used, "the concept ... of the order of an expression plays a part which is no less essential than before ..." (p. 269). Since the theory of semantic categories is not used, "it may happen that one and the same sign plays the part of a functor in two or more sentential functions in which arguments occupying the same places nevertheless belong to different orders." (p. 269) 'Order' is therefore redefined by considering <u>all</u> arguments in <u>all</u> sentential functions for a given sign:

> If the order of all these arguments is smaller than a particular natural number n, and if there occurs in at least <u>one</u> sentential function an argument which is exactly of order n-1, then we assign to the symbol in question the order n. (p. 269)

Note that the "semantic correlation" in the satisfaction relation remains given this definition of order; the order of variables in a given sentential function must also be the order of the names of the objects which satisfy the sentential function. Now, given this definition of 'order', and the semantic correlation, it is still the case that the satisfaction relation applicable to a given OL must be one order higher than the highest order variable found in the OL. Consequently, even for those formal languages Tarski uses which are not based on the theory of semantic categories, the condition of essential richness still applies.

In section five, where the subject is OL's of infinite order, Tarski argues that it is no accident of his method that the ML must be essentially richer than the OL in order to construct a definition of truth (p. 246, and pp. 253-54). He offers a reductio ad absurdum argument using an OL of infinite order.[18] Since the OL is of infinite order it appears[19] impossible to construct a ML which is essentially richer. Basically, the reductio argument runs as follows: assume a truth predicate can be found for the OL. Since this OL is of infinite order, and the ML is not essentially richer, the ML can be interpreted in the OL in such a way that for each sentence of the ML an equivalent one is constructed[20] in the OL. Since the ML has the tools needed for constructing structural-descriptive names for every expression of the OL, the ML now becomes able to construct names for its own expressions (or those equivalent to them) - including those using the truth predicate. The ML thus becomes a "universal" language in which the paradoxes can be derived (Theorem I, pp. 247-251). Given the resulting contradiction, Tarski concludes that the ML <u>must</u> be essentially richer for it to be able to define truth

---

[18]. It is worth noting that Tarski admits this reductio argument properly belongs in the ML-2.

[19]. In the Postscript Tarski suggests using transfinite orders to get around this difficulty.

[20]. This process involves using Godel's method of arithmetization, and therefore only applies where the OL is complex enough to contain arithmetic. However, Tarski does suggest that such "technical devices" can be avoided - p. 248, fn. 2.

for its OL.

In sections 5 and 6 Tarski concludes that no comprehensive truth predicate is possible for languages of infinite order. Instead he proposes, for such languages, limited definitions of truth such that for a given truth predicate Tr_n, of order n, its extension consists of expressions with variables of orders less than n (Theorem II, p. 255). At this point it is possible to see what has been called a "Tarskian hierarchy of languages"[21] where, for a given language of infinite (or transfinite - see below) order, each fragment of that language containing all expressions of a specific order n or lower, there is a distinct truth predicate of order n+1. The entire infinite order language can be interpreted as a hierarchy of languages in which each level of language has a specific highest order, and each level contains expressions of lower levels, as opposed to translations of them. Tarski in fact hints at such an interpretation when he construes a fragment of his infinite-order OL as being a language of finite order

---

[21]. For example, in 'Predicative Logics', Allen Hazen, in Handbook of Philosophical Logic. vol.I, D. Gabbay and F. Guenthner, eds. p. 384. Also Dictionary of Logic. W. Marciszewski, ed. p. 388. "We obtain a Tarskian hierarchy of languages by considering those fragments of the language of (Russell's) ramified system containing, for some n, only variables of level n or lower and predicates of level n+1 or lower."

(pp. 255-56).[22]

As an example of such a "Tarskian" hierarchy, it should be sufficient simply to use the OL of infinite order that Tarski uses in section five, the general theory of classes (pp.241-243)[23]. The constants used are negation, disjunction, and the universal quantifier, as before. For variables he uses such signs as 'X^{1}_{1}', 'X^{1}_{11}', 'X^{11}_{1}', etc., where n strokes above and k strokes below indicate the k^{th} variable of the n^{th} order, named by 'V^{n}_{k}'. Thus where n=0 the variable represents names of individuals ("objects of the first order"); where n=1 the variable represents classes of individuals ("objects of the second order"), and so on. Now, the Tarskian hierarchy involves taking that fragment of the general theory of classes which contains variables of order 1, and none higher, as constituting the OL (effectively, this is a first order predicate calculus). The truth predicate and all other semantic predicates for this language are found in that fragment of the general theory of classes which includes all expressions of order 2 or less, and this constitutes the ML. Since the general theory of

---

[22] See also comments on p. 270.

[23] One modification to his explanation of the general theory of classes on pp.241-243 will be made: following his recommendation in the Postscript (p.269), names of individuals and the variables representing them will be given the order 0, instead of 1.

classes is of infinite order, the HL created from it is also infinite in extent. In summary:

OL:
- zero order variables such as 'X^{0}_{1}', 'X^{0}_{11}','X^{0}_{111}', etc., first order variables such as 'X^{1}_{1}', 'X^{1}_{11}','X^{1}_{111}', etc., plus logical constants and quantifiers.

ML:
- all expressions of the OL;
- second order variables 'X^{11}_{1}', 'X^{11}_{11}', and appropriate constants and quantifiers;
- structural descriptive names of all OL expressions ("the morphology of language") e.g., 'V^{0}_{1}', 'V^{0}_{2}', etc., and 'V^{1}_{1}', 'V^{1}_{2}', etc.;[24]
- derived semantic predicates as found in sections two and three of CTFL.

ML-2:
- all expressions of the ML;
- third order variables 'X^{111}_{1}', 'X^{111}_{11}', and appropriate constants and quantifiers;
- structural descriptive names of all ML expressions, e.g., 'V^{2}_{1}', 'V^{2}_{2}', etc.;
- derived semantic predicates as found in section three of CTFL.

Metalanguages of higher order repeat this pattern indefinitely.

With this system in place, the "semantic correlation" present in the satisfaction relation is evident in the notation itself. Take for example the expression used on p. 46 above:

---

[24] Other ML variables as introduced on pp. 172-173 could also be included here, with appropriate superscripts attached.

$$\langle \langle a,b\rangle, \iota_{1,2} + \overline{\iota_{1,2}} \rangle \in \text{Sat}.$$

Using the new notation, superscripts must be used which clearly indicate the order of the variable being used:

$$\langle \langle X^n{}_1, X^n{}_{11}\rangle, \iota^n{}_{1,2} + \overline{\iota^n{}_{1,2}} \rangle \in \text{Sat}$$[25]

With the superscripts in place, it is clear that the satisfaction relation must be at least of order n+1, given that the order of one of its arguments - specifically '$\langle X^n{}_1, X^n{}_{11}\rangle$' - is n.[26] Note however that the order of expressions such as '$V^2{}_1$' is not stipulated by '2' as present in the expression; this '2' simply indicates exactly what variable is being named (i.e., '$X^2{}_1$' as opposed to, say '$X^1{}_1$'). Similarly, the 'n' in '$\iota^n{}_{1,2} + \overline{\iota^n{}_{1,2}}$' does not indicate the order of this expression. This is the essence of the semantic correlation found in the satisfaction relation; the variables '$X^n{}_1$' and '$X^n{}_{11}$' are clearly of the same order as the expression named by '$\iota^n{}_{1,2} + \overline{\iota^n{}_{1,2}}$'. The

---

[25] I have taken the liberty to adapt Tarski's definition of 'ι' (Def. 1, p.175), so that it may be used with variables of any order; the 'n' indicates the order of the variables involved. The complete structural descriptive name equivalent to '$\iota^n{}_{1,2}$' is:
'$(in^\frown V^n{}_1)^\frown V^n{}_2$'

[26] The exact order of this semantic predicate depends on how ordered pairs and ordered n-tuples are defined. Where '$\langle X^n{}_1, X^n{}_{11}\rangle$' is defined as $\{ \{X^n{}_1\}, \{X^n{}_1, X^n{}_{11}\} \}$ the order of the satisfaction relation is greater than n+1. Tarski offers no explanation of how he interprets ordered pairs/n-tuples, and so it isn't clear how he would address this problem.

order of the name of the expression is not involved.[27]

In the Postscript to CTFL, Tarski considers the use of languages not based on the theory of semantic categories (as mentioned earlier). He also considers the use of transfinite orders, to construct a comprehensive definition of truth for a language of infinite order. He concludes that there is no insurmountable difficulty in doing so, again provided the ML can remain essentially richer by using transfinite orders (pp.271-272). The reductio argument referred to earlier is here modified to imply that, where the metatheory is not of higher order than the OL, no definition of truth is possible with consistency. He concludes the issue with the following general theses:

A'. The semantics of any formalized language can be established as a part of the morphology of language based on suitably constructed definitions, provided, however, that the language in which the morphology is carried out has a higher order than the language whose morphology it is.

B'. It is impossible to establish the semantics of a

---

[27] There is, however, a de facto restriction involved; given the fact that the satisfaction relation of order n+1 can only take arguments of order n and lower - e.g., X^{n}_{1}, X^{n}_{11}, obviously the only names it can take as arguments will be ones which name sentential functions whose order is n or lower - e.g., $\iota^{n}_{1,2} + \overline{\iota^{n}_{1,2}}$ - because the variables will effectively determine the order of the sentential function being named.

This issue will have implications in chapter four, when a direct definition of truth, without use of satisfaction, is considered.

> language in this way if the order of the language of its morphology is at most equal to that of the language itself. (pp.273-274)[28]

Presumably, the "Tarskian hierarchy" can be extended to the levels of transfinite order as well. It should also be noted that although such a use of transfinite orders may allow a single, comprehensive definition of truth for an OL of infinite order, A' and B' make it clear that such a definition cannot be comprehensive enough to also range over expressions of transfinite order as well.

A summary of this journey through Tarski's CTFL is appropriate, as a prelude to a critique of this solution to the semantic paradoxes. No direct definition of truth was possible given 1- the limitations of quotation mark names and 2- the fundamental role of sentential functions in Tarski's OL; sentences are derived from sentential functions and not primitive sentence forms and so a direct recursive definition was not possible. Thus Tarski makes fundamental use of the concept of satisfaction of sentential functions. The satisfaction relation must be

---

[28] John R. Myhill has demonstrated that, provided a language lacks "classical negation", it is possible for it to contain its own satisfaction and truth predicates. Myhill's system also lacks a universal quantifier. Cf. 'A Complete Theory of Natural, rational, and Real Numbers', Journal of Symbolic Logic, vol 15, (1950), pp.185-196, especially p. 194.

of higher order than any expression of the language to which it applies, and since other semantic concepts are defined using the satisfaction relation, this condition of "essential richness" applies to all semantic predicates.[29] Essential richness itself is ultimately derived from what might be called the "theory of orders" (which Tarski maintains even after abandoning the need for a theory of semantic categories), combined with the "semantic correlation" which is a feature of the satisfaction relation. By using transfinite orders, Tarski is able to define semantic concepts which can range over languages of infinite order. However, he uses a reductio ad absurdum argument to suggest that consistent semantic predicates for a language must be of higher order than any found in the language to which the predicates apply.

It is important to stress the distinction between the "Tarskian" hierarchy of languages and any specific hierarchy which Tarski actually uses. Tarski's interest is to construct a model by which the semantics for any given formal language can be constructed without contradiction, whether that language be a first order

---

[29] This claim is not made explicitly by Tarski, however it does seem to be a necessary step. If one were to substitute its definition in the place of 'true', the connection becomes clear.

predicate calculus or a language of infinite order. The Tarskian hierarchy, on the other hand, emerges only when focus shifts from constructing a consistent definition of truth to the issue of preventing the semantic paradoxes. From <u>this</u> perspective, a systematic, comprehensive and simple solution to the paradoxes can be constructed on the basis of Tarski's work - and this is the Tarskian hierarchy of languages. This hierarchy is systematic in the sense that the rules which prevent contradictions are roughly the same for all levels of language; it is comprehensive in the sense that it is applicable to many formalized languages, based on the theory of orders, and it is simple because its form is easily seen; by dividing languages on the basis of order, the function of essential richness is clear.

The final step in completing this exegesis of the standard HL is to examine the degree of self-reference possible in a Tarskian hierarchy.[30] To do this, it will be useful to examine the methods used in the reductio argument referred to above (Theorem 1 in

---

[30] In his article 'On Decidable Statements in Enlarged Systems of Logic and the Concept of Truth', Tarski claims to have shown, in CTFL, that "metalogical statements about the system L can, at least in part, be formalized, or rather interpreted in the system L itself" p.106, <u>Journal of Symbolic Logic</u>, vol. 4, number 3, 1939.

section five). This reductio argument involves considering a ML which is of at most the same order as the OL to which it applies. In such a case, Tarski suggests, the OL may interpret (i.e., can name and translate) the ML[31] - and thus since the ML is capable of naming all OL expressions, both OL and ML become capable of naming *all* their own expressions. Given the presence of the semantic predicates, the result is a reintroduction of the semantic paradoxes.

What Tarski accepts, without stating or explaining, is that an OL of given order n (where n could equal the smallest transfinite ordinal ω, or for that matter ω+1 etc.) *logically cannot* interpret expressions of order n+1 or higher. Otherwise, a statement such as the following makes no sense:

> It is impossible to give an adequate definition of truth for a language in which the arithmetic of natural numbers can be constructed, if the order of the metalanguage in which the investigations are carried out does not exceed the order of the language investigated (cf. the relevant remarks on p.253). - p.272, CTFL.

If the order of the ML *does* exceed the order of the OL, then an adequate definition *is* possible (he goes on to state as much in Thesis A, p.273). This is only possible *because the truth predicate is essentially*

---

[31] Generally, the OL must be capable of arithmetization before this is possible. However, see p.248, ft. 2, and p.272.

richer, and _cannot_ be interpreted in the OL, even though the OL is capable of arithmetization. The obvious question is why, using arithmetization, it is impossible to interpret the semantic predicates in the OL. The "relevant remarks" Tarski leads us to are somewhat unclear, though they offer some assistance.[32] On p. 253 he refers back to the languages of finite order he used in sections two, three and four:

> ...the methods there applied required the use in the metalanguage of categories of higher order than all categories of the language studied and are for that reason fundamentally different from all grammatical forms of this language.

The question at hand is what implications there are to this "fundamental difference". The most reasonable interpretation of his statement is that, using the method of arithmetization it _is_ possible to interpret predicates of order n+1 in an OL of order n, but in so doing the OL is "fundamentally changed" into an ML of order n+1. No contradictions will result from such interpretations, because the interpreted predicates of order n+1 will take arguments of highest order n, and

---

[32] Further assistance may come from a reference to Godel's work: p. 274, fn. 1. Perhaps the inability to prove consistency for a language is analogous to the inability to define truth in such circumstances, and the same reasons apply. This issue, however, takes us too far afield.

so no-self reference is possible.

There is the fact that in section five, Tarski does manage to construct a contradiction in his reductio argument. But there, the ML was _not_ of higher order and so when the truth predicate was interpreted in the OL, it could indirectly take itself (or more exactly its name) as its own argument - and Tarski thus concluded no such predicate could exist to be so interpreted. In a Tarskian hierarchy which regresses to infinity, the semantic predicates for a given level can always be of higher order, and interpreting the ML-n+1 in the ML-n is not a danger. Either only predicates of the highest order found in ML-n will be interpreted, or if those of the highest order of the ML-n+1 are also interpreted, then what was the ML-n is now the ML-n+1, by definition. The latter case is uninteresting. The former provides us with a good indication as to what degree of self-reference is possible in a given ML-n in the Tarskian hierarchy.

It would seem that, along this line of conjecture, a given ML-n is capable of containing names for _all_ of its own expressions. The Tarskian hierarchy places no direct restriction on the use of names of expressions, as was explained on pp.58-59 above. No contradiction results because the semantic predicates for the ML-n cannot, by definition, be contained in the ML-n. It is important to remember that the <u>name</u> of an expression does not have to be of the same order as the expression it names. If it did, then for a metalanguage of order n, even though it could contain the name of an expression of order n, it could assert virtually nothing of that expression, because any attribute of such an expression would have to be of order n+1 – i.e., one higher than the name of the expression. Perhaps the <u>name</u> of an expression could even be treated as order zero (if signs could be treated as "objects"), or order one. In any case, the <u>syntactic</u> attributes of a language need not be excluded from the language in this manner. The <u>semantic</u> attributes are excluded through the fact that, for Tarski, all semantic concepts are defined using the satisfaction relation, and given the "semantic correlation" explained earlier, a satisfaction relation of order n will only take names for expressions of orders lower than n as arguments. Since syntactic attributes are not defined using the satisfaction relation, this restriction does not apply to them. Only those attributes of an ML of order n which require the use of satisfaction in their definition are precluded from being expressed in the ML.

The Tarskian hierarchy is a far cry from a "simple hierarchy of languages", where all self-referential sentences are ill-formed. The increased expressibility of the Tarskian hierarchy is not, however, ad hoc[33]; it is based on the use of the theory of orders, combined with the properties of the satisfaction relation.[34] The only restrictions placed on expressive capacity here are placed on the semantic predicates, otherwise, any self-reference permitted by the theory of orders is possible.

---

[33] Recall the statement quoted in chapter two: "For to avoid being ad hoc, such an approach is based generally on the distinction between use and mention of expressions." Martin, 'Towards a solution to the Liar Paradox' in Philosophical Review, LXXVI, 1967, p. 280.

[34] The properties found in the satisfaction relation are also present in the designation relation, as will be shown in the next chapter.

Chapter Four:
Problems with the Tarskian Hierarchy of Languages

Having completed an outline of the standard, "Tarskian" hierarchy, it is now appropriate to take a closer look at some of the attributes of this system. In particular, the semantic relations between different levels of the hierarchy, and the logical significance of these relations, may offer some fruitful insights on both the hierarchy of languages as well as on Tarski's definition of truth.

Tarski first discusses the relations between his OL and ML in section 2 of CTFL, while briefly outlining the structure of the ML. He focuses on two particular issues, and says that aside from these, "the process of formalizing the metatheory shows no specific peculiarity. In particular, the rules of inference and of definition do not differ at all from the rules used in constructing other formalized deductive sciences" (CTFL, p.170). The two points which require special care are first, the enumeration of all signs and expressions used in the ML, and second, the setting up of a system of axioms which can at least form a foundation for Tarski's results. The first of these features is of particular interest here, since in enumerating these signs, their referents in the OL are stipulated, and thus the semantic relation between the two levels of language is established. Tarski certainly considered the topic important: "were we to neglect...[this enumeration], we should not be able to assert either that we had succeeded in correctly defining any concept on the basis of the metalanguage, or that the definition constructed possesses any particular consequences." (CTFL, p.170) Exactly <u>why</u> it would be impossible to make the assertions Tarski refers to without this "enumeration" is an important question to be addressed later. For the time being, it is enough to note that for Tarski this enumeration is of fundamental importance.

This enumeration of expressions of the ML is divided into A- "expressions of a general logical character" (CTFL, p.170), and B- "specific terms of the metalanguage of a structural-descriptive character" (CTFL p.172). Within A, there is a series of expressions which have the same meaning as the constants of the OL. The structural-descriptive expressions (B above) are described as "names of concrete signs or expressions of [the object language]"

(CTFL, p.172)[1] It is these expressions, and the establishment of their relation to OL expressions which is of special interest here. Effectively, his intention is to provide explanations of what (if anything) the signs of the ML designate, or name, in the OL. Thus, "the symbolic expression `$((ng^\frown in)^\frown v_1)^\frown v_2$' can serve as a name of the expression `NIx_1x_{11}'." (CTFL p.172) Following the conventions outlined on p. 34 of chapter 3 above, we may also <u>translate</u> this OL expression into the ML as

a is not included in b

All things considered, it may be more appropriate to describe this not as an "enumeration of signs and expressions" , but rather as stipulating the semantic connection between ML and OL expressions; Tarski is not simply enumerating a set of signs, he is also establishing the meanings of these signs.

One curious feature of these connections is that it is necessary to use <u>names of ML expressions</u>. Since Tarski refuses to allow quotation mark names in his formalized systems, they cannot be used. However, whatever considerations necessitate stipulating the connection between ML names of OL expressions and OL

---

[1] The system Tarski sets up was outlined at the beginning of chapter three, above.

expressions would presumably also apply to correlating ML expressions and _their_ names.

It is worth asking what, from _within_ a completely formalized hierarchy, such statements would look like. They would _not_ look like:

((ng^in)^v1)^v2 _designates_ a is not included in b

because this statement _uses_ the ML name and its translation; this statement actually correlates an OL expression and a "state of affairs" which it designates. Recall the semantic connection of Tarski's quoted earlier: "the symbolic expression \`((ng^in)^$v_1$)^$v_2$' can serve as a name of the expression \`NIx_1x_{11}'." The ML name is _mentioned_ here; that is to say it is _named_, using a quotation mark name. To express the semantic connections between OL and ML _without_ using quotation mark names would require developing a system of structural-descriptive names which would designate _ML_ expressions[2]. Presumably, a system similar to that which Tarski uses for naming OL expressions could be developed, however, the same problem would then reassert itself; in establishing the

[2] Furthermore, this system would have to be constructed in the ML-2; these statements talk _about_ ML expressions, "and about their relation to what they mean" consequently, such a system must be part of the meta-metalanguage. (Cf. _The Encyclopedia of Philosophy_, 'Correspondence Theory of Truth' p.230.)

semantic connection between these structural descriptive names and the ML expressions they designate, we must <u>mention</u> the structural descriptive names - and to do this we must use their names thus requiring another set of semantic connections to establish the names for the name of the ML name[3] of an OL expression!

Taken in these general terms, this problem is (obviously) rather difficult to follow. It may be helpful to review this point with an example. Take the OL expression 'Ix_1x_{11}' : its ML name is '$((in^\wedge v_1)^\wedge v_2)$', and its ML translation is: a is included in b. By analogy, if we were to talk about a person's name, we would not say

Biff has four letters.

because we do not attribute letters to a person. Rather, we use the name of the person's name:

'Biff' has four letters.

So, to correlate Biff's name with Biff, we would use the name of his name:

'Biff' is Biff's name.

---

[3] Technically, all these names could be located in the ML (as opposed to the ML-2, etc.). This is so because there is no need for any of them to be of higher order than the highest order found in the OL. The regress, then, is located <u>within</u> the ML. Parallel regresses could arise when the semantic connections between the ML-2 and the ML are established.

Similarly, to correlate the ML name '$((in^\wedge v_1)^\wedge v_2)$' with the OL expression it names ('Ix_1x_{11}'), we must use the name of the ML name. However, since Tarski has rejected quotation mark names in his formalized systems, he has a tougher time than Biff. A new system of structural descriptive names must be constructed to make it possible to name ML expressions. Let '<u>in-v1-v2</u>' function as the name of '$((in^\wedge v_1)^\wedge v_2)$'. Thus, to correlate the ML name ['$((in^\wedge v_1)^\wedge v_2)$'] with the OL expression ['Ix_1x_{11}'], we would say

<u>in-v1-v2</u> designates $((in^\wedge v_1)^\wedge v_2)$

But now the same problem surfaces: if it is necessary to state the connection between an expression and its referent - as Tarski says (p. 69 above, "were we to neglect... p.170 CTFL) - then it should be equally necessary to state the connection between '<u>in-v1-v2</u>' and '$((in^\wedge v_1)^\wedge v_2)$'. To do this we must first have a name for '<u>in-v1-v2</u>', and so the process of stating semantic connections continues forever.

Thus we are faced with a messy infinite regress, and before considering any solutions to it, we must first see whether it is a harmful or harmless regress. Tarski suggests that without such a semantic connection of ML and OL expressions, it would be impossible to make any assertions regarding a correct definition of truth. Specifically, the question is what _logical_ importance is to be attributed to such designation relations, and exactly _why_ it would be impossible to make the assertions Tarski refers to without such designations.

It may prove easiest to answer these questions in a round about way, by looking at how these semantic connections come to bear on Tarski's definition of truth. In constructing his definition, Tarski uses what he calls "convention T" (CTFL, pp.187-188) as a measure of the material adequacy for a definition of truth. This convention is in turn based on what may be called "T sentences"; sentences of the form:

> 1- 'Snow is white' is true _iff_ Snow is white.

Sentences of this form are taken to exemplify the correct use of "true". Tarski considers the possibility of constructing a general definition of truth directly on the basis of these T sentences. For example:

> 2- for all x, x is a true sentence _iff_ for a certain p, x='p' and p.[4]

However, following normal conventions for using such quotation mark names, the consequence of 2 is that the letter 'p' is the only true sentence. Tarski considers using "quotation variables", but for various reasons

---

[4] This is Tarski's number 6, see pp.159-160, CTFL.

this is also rejected.[5] He then proceeds to construct a recursive definition without any further consideration of a general definition of truth, as was explained in chapter three.

In statement 1 above, there appears to be no need to explain the fact that the expression on the left is a name of the sentence on the right. However, in 2, it is necessary to explain this, by using a quotation mark name on the right hand side. With this in mind, consider the following example of a T sentence, provided by Tarski, which occurs within his formalized system:

3- $\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1})$ is a true sentence <u>iff</u> for any classes a and b we have $a \subseteq b$ or $b \subseteq a$.[6]

In this case, there is no need to include an explanation of the naming relation between the left and right hand sides of the biconditional, in spite of the fact that quotation mark names are not used. The difference between 2 and 3 is that in 3 Tarski has already stipulated the structural-descriptive system of naming which is being used.

---

[5] See pp. 161-162, CTFL. I will not consider this option here, but it is worth noting that it seems more promising than Tarski thinks. See 'Use, Mention, and Quotation', L. Goddard and R. Routley, in <u>The Australian Journal of Philosophy</u> vol. 44, 1966, pp.1-49.

[6] CTFL, p.187.

At this point it is somewhat easier to explore the question of why Tarski says it would not be possible to claim to have defined any concept on the basis of the ML without these semantic connections. Without an understanding of the semantic connection between

' $\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1})$'

and

'for any classes a and b we have $a \subseteq b$ or $b \subseteq a$'

(i.e., that the former is the name of the latter), statement 3 would make no sense – just as statement 2 would make no sense if not for the semantic connection achieved by ' x= 'p' '. The difference between 2 and 3 is that the place of this semantic connection has moved from <u>inside</u> the statement (as with 2) to what Tarski calls "the enumeration of...the signs and expressions which will be used in the metalanguage" (CTFL, p.170).[7] Thus ML statements such as 3 can be understood because of the explanations of their semantic connection. If we don't know what the symbols mean, then of course we are hardly in a position to assert that a certain sequence of signs is a correct

[7] As will be demonstrated later in this chapter, it is possible, using only the tools Tarski already employs, to place the semantic correlation desired in 2 in such an enumeration of ML signs, and thus create a direct, general definition of truth.

definition of truth.

However, as the infinite regress has shown, Tarski has not avoided the problem; by rejecting the use of quotation mark names, a completely formalized hierarchy of languages is unable to "tie down" the meanings of its terms.

Now, for all that there is a certain level of "harm" to this regress, it does minor damage only. The problem can be solved by using a modified version of quotation mark names such as that provided by Goddard and Routley (see footnote 5, this chapter), or some similarly generalized method of naming. Furthermore, it is important to differentiate a <u>cognitive</u> problem from a <u>logical</u> one. Notice that the problem as outlined above is stated as a conditional: "<u>If</u> we don't know what the symbols mean...". If we <u>do</u> know what they mean, then there is no problem with the logical integrity of the hierarchy of languages. The issue then is the cognitive state of the reader, and not the formal structure of the system. Similarly we could say that if we didn't know what 'then' meant, the sequence

if a then b
a
therefore, b.

would make no sense, and it would be impossible to assert the validity of the argument. There is no *logical* problem here, only a *cognitive* one. Consequently, the infinite regress present in Tarski's enumeration of ML expressions is not a serious one; no *logical* flaw is present, and only minor changes are needed to avoid the problem altogether.

An interesting corollary of this analysis is the availability of a direct, general definition of truth using only tools Tarski already employs. Consider that in 1 above, there appears to be no need to explain the fact that the expression on the left is a name of the sentence on the right. In 2, it seems to be necessary to explain this connection, by using a quotation mark name on the right hand side. In the case of number 3, there is no need to include an explanation of the naming relation between the left and right hand sides of the biconditional, in spite of the fact that quotation mark names are not used. The difference between 2 and 3 is that Tarski has already stipulated the structural descriptive system of naming which is being used in 3. What is curious about this situation is that this appears to be the only significant difference between 2 and 3. Consequently, by building the appropriate designation relations into the enumeration of ML signs (along with those already stipulated by Tarski) which are capable of replacing the " x= 'p' " in 2, a general, direct definition of truth does seem to be possible.

Tarski uses the letters 't', 'u', 'w', 'x', 'y', and 'z' as metalanguage variables representing the names of object language expressions and sequences of expressions (CTFL, p.173) - thus 'x' as it is found in 2 above need not be changed. Two minor modifications to Tarski's notation will be required, however: first, a variable performing the function of 'p' in 2, and secondly, a systematic means of correlating a sentence and its name. On p. 169 Tarski introduces variables ranging over sentences or sentential functions: 'p' and 'q' are described as "any sentences or sentential functions". Given the fact that Tarski uses 'p' for different purposes elsewhere,[8] 'r' will be used here to represent any sentences or sentential functions (i.e., it will replace 'p' and 'q' as they are used by Tarski on p.169, CTFL).

Finally, we need to have some means of correlating a variable representing an expression with a variable representing the name of that same expression. Once again, we need only use tools already used by Tarski, applied in a new situation. Recall that in the OL he

---

[8] Tarski gives 'p' a different function as stipulated on p.173, CTFL.

uses in section two of CTFL, 'x_1', 'x_{11}', etc are variables ranging over classes of individuals, and that corresponding to these are ML structural descriptive names 'v_1', 'v_2', etc. By employing the same system of subscripts to ML variables which represent expressions ('r' as introduced above), as well as the ML variables representing names of expressions (for example 'y' above), we can correlate the variable representing an expression with the variable representing the name of that expression. Thus an expression 'r_1' would be named by 'y_1', and 'r_2' would be named by 'y_2', etc. The only difference in this use of the connection is that, instead of correlating a variable ('x_{11}') with a constant ('v_2') we are correlating a variable ('r_2') with a variable ('y_2') - not a logically significant difference.[9] Finally then, the following statement constitutes a direct, general definition of truth within Tarski's system (where S is the set of all meaningful sentences[10]):

---

[9] Tarski actually does correlate variables in exactly this fashion, on p. 191 of CTFL. There the variable v_k (ranging over names of OL variables) is correlated with f_k (a translation of the variable named by v_k). The only difference between this situation and the one introduced above is that the variable f_k ranges over classes of individuals, while r_k ranges over OL expressions.

[10] See definition 12, p.178, CTFL.

4- for any y_k, y_k is a true sentence <u>iff</u> y_k is a member of S and r_k.

Any "T-sentence" can be instantiated into this definition. For example, Tarski's expression quoted earlier (statement 3 above). 'y_k' is instantiated by '$\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1})$', and '$r_k$' is instantiated by the expression named, i.e., 'for any classes a and b we have $a \subseteq b$ or $b \subseteq a$'. Once instantiated into definition 4, the expression reads exactly like statement 3, with the addition that 'for any classes a and b we have $a \subseteq b$ or $b \subseteq a$' is a sentence - which it is because it contains no free variables.

This definition is derived entirely within Tarski's system, making only minor modifications to his notation. It completely avoids the need to use a satisfaction predicate, and avoids many of the complications Tarski wrestles with in section 2, and it virtually eliminates the need for sections 3, and 4, since these two sections are concerned with the substantial modifications required to adapt his definition to languages of higher order. Definition 4, however, does not encounter these difficulties.[11] Furthermore, without having to use satisfaction, it

---

[11] To be fair to Tarski, he had already found ways of avoiding most of those complications by the time he wrote his Postscript.

would be much easier to derive the various theorems he needs to provide for his system of semantics.

However, by eliminating the need for the satisfaction relation, the relevance of "essential richness" comes into question. Essential richness (that the ML-n be of higher order than the ML-(n-1)) was based on two elements: first, the theory of orders, and second, the fact that all semantic predicates are defined in terms of satisfaction and the satisfaction relation must be of higher order than the _expression named_ in a given relation. Tarski is able to apply essential richness to the truth predicate (and all other semantic predicates) because he defined them using the satisfaction relation. As a result the substitution instances for the truth predicate are determined by the satisfaction relation, and so the only sentences its argument can name are those of lower order than the truth predicate itself.

The problem is that this new definition of truth does not make use of the satisfaction relation. As a result, it would appear that the truth predicate need only be of higher order than the _name_ of a sentence, rather than the sentence itself.[12]

This is a very significant change, for if the name of a sentence can be of a lower order than the sentence it names, then the truth predicate Tr^n (of order n) could be named by an expression of order n-1, and thus the predicate could take its own name as argument, reintroducing the paradoxes. The name of a sentence containing a truth predicate of order n could be of order n-1, and thus it could have the truth predicate of order n attributed to it. For example:

5- let $l^n = {\sim}Tr^n(m^{n-1})$

6- let m^{n-1} Des^{n+1} l^n

Here 'l^n' is a constant which could be instantiated for r_k (in formula 4 above), and 'm^{n-1}' is a constant which would be substituted for the name of r_k, i.e., y_k. Using 4 above:

7- $Tr^n(m^{n-1})$ <u>iff</u> l^n

or, given 5,

---

[12] Recall the example given in chapter three:

$< \langle X^n_1, X^n_{11} \rangle, \iota^n_{1,2} + \overline{\iota^n_{1,2}} > \in Sat$

In this expression, it is the order of the first argument ('$\langle X^n_1, X^n_{11} \rangle$') which necessitates that the satisfaction relation be of order n+1 ("essentially richer"). The order of the second argument '$\iota^n_{1,2} + \overline{\iota^n_{1,2}}$', is not an issue (see p. 58, above). Once the truth predicate is not defined using satisfaction, the only restriction on its expressive capacity is its own argument -which would be of the form '$\iota^n_{1,2} + \overline{\iota^n_{1,2}}$' - and this term itself does not necessitate essential richness.

8- $Tr^n(m^{n-1})$ iff $\sim Tr^n(m^{n-1})$

Superscripts have been used to indicate the order of the symbols being used. The designation relation used must be of order n+1, since '1^n' contains a predicate of order n (i.e., 'Tr^n'). The superscripts show that in all of this, there is no violation of Tarski's "theory of orders".

One possible objection is that although truth may be defined directly, designation must still be defined via satisfaction. However, the notion of designation presupposed in definition 4 was also presupposed by Tarski when he defined satisfaction.[13] Tarski saw no reason to include this form of designation in his formulas (as with 3 above).

It is still possible, however, to impose restrictions on the accepted range for the variable 'y_k' for definition 4. Tarski's rejected definition (2 above) was altered by replacing ' x='p' ' with two minor modifications to Tarski's notation: a variable ranging over OL expressions ('r_k'), and a subscript correlating the variable ranging over names of OL expressions ('y_k') with the expression named. In

---

[13] See H. Field, 'Tarski's Theory of Truth', in The Journal of Philosophy, vol. 69, pp.347-375. 1972. Field's article is based on an analysis of Tarski's enumeration of ML symbols.

effect, ' x='p' ' was replaced by ' y_k Des r_k '. Notice that in statement 6 above it was necessary to use a designation relation of order n+1. Now, the definition of truth in 4 tacitly uses a designation relation correlating 'y_k' and 'r_k', and this designation relation is of order n.[14] This is all that is needed, given the desired range of r_k. If we replace the designation relation used in 6 above with one of order n, then '1^n', which is of order n, is an invalid substitution for 'r_k', and the contradiction cannot be derived.

In his formulation of 2 Tarski was very close to a much simpler definition of truth. The question which comes to mind is: how did he miss it? If one were to include the designation relation introduced above directly into the definition of truth, even the need for subscripts vanishes, and the definition can be written as follows, where S is the set of meaningful sentences and using 'x' and 'q' as variables ranging

---

[14] If we abandon Tarski's attempt at reducing semantic concepts to non-semantic concepts, this relation can be explicitly included in the definition - as in 9 below. Once the role of the designation relation is shown explicitly, it is clear that, by following the derivation in lines 5 to 8, the closest we get to a contradiction is:

$Tr^{n+1}(m^{n-1})$ <u>iff</u> $\sim Tr^{n}(m^{n-1})$

This situation is similar to that demonstrated for "heterological", in chapter two, and does not involve a contradiction.

over expressions and names of expressions respectively – as Tarski uses them:

> 9- for any x, x is a true$^{\text{n}}$ sentence <u>iff</u> ($\exists$q) x is a member of S, x designates$^{\text{n}}$ q, and q.

The problem Tarski would have with this definition is it violates one of his prime goals: "...[in defining truth] I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts" (CTFL, p.153). In 9 truth has been reduced to designation, but designation hasn't been reduced to any other non-semantic concepts. It would be interesting to look in Tarski's early notes for CTFL, to see whether he attempted to find a way to reduce designation as he did satisfaction. He may have considered ' x='p' ' as a failed attempt to do exactly this.[15]

Using the modifications to Tarski's notation introduced above does make it possible to "reduce" designation, as follows ('y_k' and 'r_k' will be used as introduced above):

> 10- x Des q =df. $(\exists r_k)(\exists y_k)$ $q=r_k$ and $x=y_k$

Consequently, definition 9 is also reduced to non-

---

[15] The expression ' x='p' ' fails for Tarski in two ways: first due to the limitations of quotation mark names, and second because it is an identity relation between two names, similar to "Eric Blair = George Orwell". Consequently the designation relation is neither reduced or avoided, but tacitly present in the form of a quotation mark name.

semantic concepts. Any objections Tarski might raise against 4, 9 or 10 (vis-à-vis reducing semantic concepts to non-semantic concepts) would also affect his definition of satisfaction; all three are based on the enumeration of signs and expressions used in the ML. If the existence of this enumeration implies that truth, as defined in 4 or 9 has not been reduced to non-semantic concepts, then it has the same implication for Tarski's definition of satisfaction, and also for his definition of truth. What the "definition" of designation in 10 above shows is that Tarski's enumeration of ML expressions may not allow him to reduce semantic concepts as easily as he would like. The definition seems to beg the question; in the process of explaining the meaning of the variables 'y_k' and 'r_k' we effectively presuppose exactly the form of designation we wish to define using these variables. Of course, if we abandon the need to reduce semantic concepts to non-semantic concepts, then Tarski's theorems regarding completeness, consistency, provability, etc. are still intact - but now with a much simplified definition of truth.

In any case, the issue of reducing semantic concepts to other non-semantic concepts does not have an impact on the question of the consistency of the hierarchy of languages. Definitions 4 and 9 are derivable within Tarski's notation, and both appear to avoid contradictions. The regress discovered at the beginning of this chapter also has little impact on the HL; it appears harmless, and could probably be avoided altogether by using some form of quotation mark variables.

Aside from establishing the semantic connections between levels of the HL, there is little evidence of internal weakness in the Tarskian hierarchy. There are, however, serious problems for the field of semantics in general, given the nature of the restrictions imposed by the Tarskian hierarchy. These problems relate to the inability of the system to be expressed in a language capable of meeting the restrictions the system would impose, and secondly, the inability to prove its consistency, or freedom from contradictions.

Chapter Five: Expressibility

The original purpose of this thesis was to examine the standard version of the hierarchy of languages, to see whether there were any logical difficulties present in the hierarchy as a whole. Although some minor complications have been observed, none are insurmountable. Probably the greatest problem relates to the expressibility of the principles of the hierarchy as a whole:

> ...semanticists are in the anomalous position of being unwilling or unable to apply to their own field and to their own assumptions the same standards of scientific exactness and linguistic analysis which they rightly insist on applying to the fundamental assumptions of other branches of knowledge.[1]

The entire hierarchy is structured in such a way that any semantic predicate of order n cannot take the name of an expression of order n as its argument. For any ML-n (where the highest order of any predicate or variable is n), the truth predicate applicable to its

[1] Frederic Fitch, review of Tarski's 'The Semantic Conception of Truth and the Foundations of Semantics', in Journal of Symbolic Logic, volume 9, 1944. p.68.

See Kleene's review of Carnap's semantics for a similar point: "Semantic rules for the language of the whole science have to be formulated in ordinary language..." p.117, in Journal of Symbolic Logic, volume 4.

expressions - and the entire semantics of that language - can be constructed only in an ML greater than n. Consequently, there is no level inside the hierarchy capable of expressing the semantics for the entire hierarchy. Whatever level of language we attempt to express the structure of the entire hierarchy in, we must concede one of three points: either 1- the semantics of this level (it may be appropriate to call it "omega") can be "shown" only; no descriptions or definitions can be stated, or 2- this level is not immune to the semantic contradictions found in the universality of natural language or finally, 3- some other method of avoiding the contradictions is used. The first of these options would result in an incomplete science of semantics[2]. It would be an admission that there were some aspects of language (more specifically, certain semantic concepts) incapable of analysis. The second option results in the failure to produce a completely consistent science of semantics. This would be an admission that such aspects of language are incapable of avoiding contradiction.

---

[2] In 'The Establishment of Scientific Semantics' (pp.401-408 of Logic, Semantics, and Metamathematics), Tarski defines the task of 'scientific semantics' as "characterizing precisely the semantical concepts and of setting up a logically unobjectionable and materially adequate way of using these concepts" (p.402).

The third option would effectively eliminate the need for the hierarchy in the first place.

None of these options would be acceptable. A completely different solution would be to deny the existence of the problem; reject the need for any single ML-"omega" language. There are two possible reasons for this. First, it could be argued that the entire system, when properly constructed, would obviously be consistent, and simply by constructing such a hierarchy this could easily be seen. This option, however, offers no <u>proof</u> that the hierarchy avoids contradictions. It would also mean that much of what Tarski has written in CTFL (including his enumeration of ML expressions and their meanings) would have to be eliminated, leaving behind only the actual statements (definitions, theorems, etc.) of the ML and OL.

The second reason is it is clear that any given ML-n <u>can</u> have its semantic concepts appropriately formalized if necessary. Regardless of the value of n, it is always possible to construct its semantics in the ML-n+1. Consequently it can safely be said that the semantic predicates are free of contradiction for every level in the hierarchy. Once again, however, this argument offers no proof of consistency which is expressible <u>within</u> any level of the hierarchy. The expression 'ML-n' has a universality about it; it ranges over <u>all</u> levels of the hierarchy. Whatever language this argument is given in, it cannot be within the hierarchy.[3] We can see the reasonableness of this argument, but only within our universal natural language, which contains the contradictions.

The hierarchy of languages cannot itself be constructed within a hierarchy of languages framework. It cannot prove that it avoids contradictions resulting from semantic concepts, without making use of exactly those semantic concepts which are capable of universality and creating contradictions. Strictly speaking this is not an "internal" inconsistency, but it does indicate a failing of the theory, given the ultimate goals Tarski has for the science of semantics.

In spite of all this, Tarski does have a few suggestions which in some measure assist in overcoming the problem of expressibility. In his Postscript, Tarski suggests using semantic predicates of <u>transfinite</u> <u>order</u>. Now, what the use of semantic

---

[3] In effect this argument would involve attributing something to the semantic predicates of every level -i.e., the property of being "formalizable". To attribute this, we would need a predicate of higher order than any found in the hierarchy.

predicates of transfinite order allows is predicates which can range over a language of infinite order, or, following the "Tarskian" hierarchy of languages, it allows semantic predicates which can range over an infinite number of levels of language. At first glance, this may seem to offer exactly the kind of comprehensive language needed to provide a "formally correct" semantics for the Tarskian hierarchy as a whole. In some ways, in fact it does; it provides a transfinite metalanguage and an infinite order object language. The system is immune from contradictions, because it still follows the theory of orders, and the ML is essentially richer than all the levels to which it applies. It would be conceivable, then, to prove the consistency of this OL[4] in its transfinite ML. Furthermore, an infinite order OL can be interpreted as a Tarskian hierarchy; each new order of predicates and variables representing a new level of language. In this light, a transfinite "level" could range over an entire infinite hierarchy of languages, and provide proofs of its consistency.

Of course, this transfinite language could not apply to itself (at least it doesn't contain its own

---

[4] Subject to the incompleteness theorems and other consequences shown by Godel.

semantics). This is where the expressibility problem reintroduces itself. We now have a language of transfinite order, and the question is: where is its semantics stated? Tarski suggests that there is also a hierarchy of transfinite languages:

> To those signs of infinite order which are functors of sentential functions containing exclusively arguments of finite order we assign the number ω as their order. A sign which is a functor in only those sentential functions in which the arguments are either of finite order or of order ω (and in which at least one argument of a function is actually of order ω), is of the order $\omega+1$. (CTFL, p.270)

Such a continuation of the hierarchy has no impact on the arguments raised above; the problem of expressibility as outlined simply reasserts itself, applied to this new transfinite hierarchy.

Conclusion

No decisive logical flaws have been identified in the standard version of the hierarchy of language solution to the semantic paradoxes. The fact that the theory has survived this long suggests there probably are no serious problems of this sort. Ironically, the greatest logical problem this solution faces may well be that it is unable to prove that it faces no such logical problems.

It has been shown that the Tarskian hierarchy is able to allow a very substantial range of self reference within a language, with regards to syntax. The restrictions it places on the expression of semantics appear inevitably to lead to the inability to include the science of semantics in a language which follows the structure of the hierarchy.

The greatest internal problems discovered in Tarski's CTFL relate to his enumeration of symbols used in his ML. A weak infinite regress was identified in this enumeration of symbols, but this regress is peculiar to his method of constructing names, and not an inevitable feature of the hierarchy itself. In studying this enumeration, a general definition of truth was discovered which both simplified his work there, but also pointed out a weak link in his ability to reduce semantic concepts to concepts known to be free of contradictions.

All in all, the standard version of the hierarchy of languages is a rigorous solution to the semantic paradoxes, but its limits seem to be predetermined, and a complete science of semantics must go beyond these limits.

Appendix I

On Tarski's Reduction of Semantic Concepts

In constructing his semantic theory of truth, Tarski had two important goals: providing a system which was free from contradictions, and reducing semantic concepts to logical and physical concepts only.[1] Analyzing the first of these goals was the main focus of this thesis. However, in discovering simpler definitions for truth and designation in Tarski's notation, the opportunity presented itself to show clearly how Tarski in fact did not reduce semantic concepts to non-semantic concepts. This will be shown first by pointing out the vicious circle involved in the definition of designation, and secondly by demonstrating how Tarski's definition of satisfaction relies on this undefined use of designation. Finally, this argument will be compared to that of Hartry Field.

Using only slight modifications to his notation, the following definitions were presented in Chapter four:

4- for any y_k, y_k is a true sentence <u>iff</u> y_k is a member of S and r_k.

---

[1] See 'The Establishment of Scientific Semantics', p.406.

9- for any x, x is a true^{n} sentence iff x is a member of S, x designates^{n} q, and q.

10- x Des q =df. $(\exists r_k)(\exists y_k)$ $q=r_k$ and $x=y_k$

When these definitions were presented, it was pointed out that any objections which might be raised against 4, 9 or 10 (vis-à-vis reducing semantic concepts to non-semantic concepts) would also effect Tarski's definition of satisfaction. Definitions 4, 9, 10, and Tarski's definition of satisfaction are all based on the enumeration of signs and expressions used in the metalanguage.

When definition 10 is considered in the light of the enumeration of metalanguage expressions, it is clear that the definition is circular. Tarski explains that a vicious circle in definition "arises only when the definiens contains either the term to be defined itself, or other terms defined with its help."[2] In definition 10 or, for that matter, Tarski's own definition of denotation (CTFL, p. 194, fn.), the definiens contains expressions which were introduced using an unreduced denotation relation when the semantic theory was set up. This unreduced denotation relation is the same one as that defined inside the semantic theory, so consequently there is a vicious

[2] 'The Semantic Conception of Truth', p. 67.

circle in the definition of designation in Tarski's theory of semantics.

The significance of this unreduced denotation relation on other semantic concepts can be shown explicitly in the notion of satisfaction used by Tarski[3]. Tarski offers a convention for satisfaction similar to the one he offers for truth:

11- for all a, a satisfies the sentential function x if and only if p (CTFL, p. 190)[4]

A definition of satisfaction will be materially adequate if all equivalences of this form follow from it. Tarski offers the following application of this schema in natural language:

12- for every a, we have a satisfies the sentential function 'x is white' if and only if a is white (CTFL, p. 190)

The role of Tarski's unreduced denotation relation in his definition of satisfaction can be made perspicuous if we formulate 11 using the notation Tarski uses for the general theory of classes. He explains his notation as follows:

> As variables we use... signs composed of the symbol ' X 'and a number of small strokes above

---

[3] Since Tarski defines all his semantic concepts using satisfaction, if it relies on an unreduced notion of denotation, then all the others will as well.

[4] Where 'x' represents the name of a sentential function and 'y' represents a sentence formed from it.

> and below. The sign having n strokes above and k below is called the k-th variable of the n-th order and is denoted by the symbol ' V^n_k '.[the symbol ' ^ ' is also used to indicate concatenation in structural-descriptive names - CTFL, p. 242]

Using this notation, we can reformulate 11 as follows:

13- for every X^o_1, we have X^o_1 satisfies the sentential function V^1_1^V^o_k if and only if X^1_1 X^o_1

The denotation relation between ' V^1_1 ' and ' X^1_1 ' is essential to the proper interpretation of the satisfaction predicate. Without it, we could offer the following as a legitimate application of 13:

14- for every X^o_1, we have X^o_1 satisfies the sentential function 'V^o_k is white' if and only if X^o_1 is purple

In this example, the semantic correlation between ' V^1_1 ' and ' X^1_1 ' is absent. Obviously, its presence is necessary, for to allow sentences such as 14 precludes the material adequacy of Tarski's definition of satisfaction.[5] The only other way to prevent such inappropriate substitutions would be to somehow include the necessary correlation between name and expression named in the formation rules, transformation rules, or

---

[5] Without any rules within the system governing the relation between expression and expression named, it would be equally possible to reintroduce the semantic paradoxes.

axioms[6] for the metalanguage. In point of fact, Tarski does not do this - indeed, it is difficult to see how it could be done without relying on the notion of denotation he tacitly presupposes. Consequently, the unreduced notion of denotation introduced by Tarski (in the quote from p. 242 above) must be treated as an internal part of his semantic theory, and not simply a part of the process of constructing this theory.

This result is similar to that of Hartry Field; Field claims that "Tarski succeeded in reducing the notion of truth to certain other semantic notions" [7]. Field's criticism applies to the internal operations specific to Tarski's truth theory. He criticizes the definition of denotation inside Tarski's semantic theory, pointing out that the definition simply consists of an enumeration of denotation relations and is not what he calls a "real explication of denotation in non-semantic terms"[8]. Tarski's definition does not

---

[6] Tarski describes expressions such as ' V^1_1 ' as "primitive expressions of the metalanguage" (pp. 210-211, CTFL). He also constructs "axioms which determine the fundamental properties ...[of these primitive expressions]" (CTFL, p. 211). Such axioms are presented on pp. 173-174. However, there is nothing present in these axioms which would meet the present requirements.

[7] H. Field, 'Tarski's Theory of Truth', in The Journal of Philosophy, vol. 69, pp.347-375. 1972, p. 347.

[8] Field, p. 365.

explain denotation, it merely stipulates it, and thus Field argues that Tarski's definition is "trivial", and does not reduce the concept to non-semantic terms.

There is a significant difference between Field's conclusion and mine however. His criticism of Tarski is really a disagreement over what it means to "reduce" semantic concepts. Field expects some form of a theory of representation to be an essential part of a theory of semantics, while Tarski considers this an issue beyond the purview of semantics.[9] On the other hand, he accepts the position (held by Tarski) that "the notion of an adequate translation [of object language expressions into the metalanguage] is employed in the methodology of giving truth theories, but is not employed in the truth theories themselves".[10] The

---

[9] See Tarski's "Polemical Remarks" in 'The Semantic Conception of Truth':

> "...the semantic conception of truth implies nothing regarding the conditions under which a sentence like
>
> (1) snow is white
>
> can be asserted. It implies only that whenever we assert or reject this sentence, we must be ready to assert or reject the correlated sentence (2):
>
> (2) the sentence 'snow is white' is true."
>
> (p. 71.)

Surely Tarski would make a similar claim for other semantic concepts, including satisfaction and denotation.

[10] Field, p. 355. Tarski considered his "primitive expressions" - such as ' V^1_1 ' - to be "undefined" (See SCT, p. 73). Considering the situation presented above for his notion of satisfaction, such expressions must be defined to preserve the material adequacy of the system.

problem outlined above is a more general one, applying to what Field calls "the methodology of giving truth theories" (what Tarski calls the enumeration of metalanguage expressions). It shows that Tarski's reduction isn't simply trivial, it is circular and thus not a reduction at all.

In the definition of designation (10), and in the analysis of Tarski's satisfaction relation, it is quite clear that the distinction between "the methodology of giving truth theories" and the truth theories themselves is illegitimate in this case.[11] Field's claim that "the notion of an adequate translation is never built into the truth characterization and is not, properly speaking, part of a theory of truth"[12] isn't always correct. The direct definitions shown to be easily constructed in Tarski's semantic theory indicate that when one wishes to reduce semantic concepts to non-semantic concepts, this issue is very important. Any semantic theory which attempts to reduce semantic concepts to non-semantic concepts must be particularly

---

[11] It is worth noting, however, that Field's T1 - his own reformulation of Tarski's definition - does not face this problem because it explicitly recognizes an unreduced concept of "primitive denotation". The arguments presented in this paper lend further support in defence of Field's revised version.

[12] Field, p. 355.

careful to avoid circularity by presupposing the very concepts they wish to reduce.

BIBLIOGRAPHY

Burges, A. "Semantic Paradox" in: Recent Essays on Truth and the Liar Paradox. ed. R.L. Martin. Oxford: Clarendon Press, 1984.

Carnap, Rudolph. Introduction to Semantics. Cambridge, Mass.: Harvard University Press, 1942.

Church, Alonzo. "Comparison of Russell's Resolution of the Semantic Antinomies With That of Tarski" in: Recent Essays on Truth and the Liar Paradox. ed. R.L. Martin. Oxford: Clarendon Press, 1984.

Copi, Irving. Symbolic Logic. Fifth Edition. New York: Macmillan Publishing Co. Inc., 1979.

Edwards, Paul. ed. The Encyclopedia of Philosophy. New York: Macmillan Publishing Co. Inc., 1967.

Field, Hartry. "Tarski's Theory of Truth" in The Journal of Philosophy. vol. 69, 1972, pp.347-375.

Fitch, F. and R. Feys. Dictionary of Symbols of Mathematical Logic. Amsterdam: North-Holland Publishing Company, 1969.

Fitch, Frederic. "Review of 'The Semantic Conception of Truth and the Foundations of Semantics' " in Journal of Symbolic Logic. vol. 9, 1944, p.68.

Flew, A. ed. The Dictionary of Philosophy. Revised Second Edition. New York: St. Martin Press, 1984.

Goddard, L. and R. Routley, "Use, Mention, and Quotation" in: The Australian Journal of Philosophy. vol. 44, 1966. pp.1-49.

Hazen, Allen. "Predicative Logics" in: Handbook of Philosophical Logic vol.I. ed. D. Gabbay and F. Guenthner. Boston: D. Reidel Publishing Co., 1983.

Kneale, W. and M.Kneale. The Development of Logic. Oxford: Clarendon Press, 1962.

Kripke, Saul. "Outline of a Theory of Truth" in: Recent Essays on Truth and the Liar Paradox. ed. R.L. Martin. Oxford: Clarendon Press, 1984.

Marciszewski, Witold. ed. Dictionary of Logic. Boston: Martinus Nijhoff Publishers, 1981.

Martin, R.L. "Towards a solution to the Liar Paradox" in: Philosophical Review. LXXVI, 1967, p. 280.

Martin R. L. ed. Recent Essays on Truth and the Liar Paradox. Oxford: Clarendon Press, 1984.

Myhill John R. "A Complete Theory of Natural, Rational, and Real Numbers" Journal of Symbolic Logic, vol. 15, 1950, pp.185-196.

Parsons, Charles. "The Liar Paradox" in: Recent Essays on Truth and the Liar Paradox. ed. R.L. Martin. Oxford: Clarendon Press, 1984.

Quine, W.V.O. Philosophy of Logic. New Jersey: Prentice-Hall Inc. 1970.

Prior, A. N. "Russell, Bertrand Arthur William: Logic and Mathematics" in: The Encyclopedia of Philosophy. ed. Paul Edwards. vol. 7, New York: Macmillan Publishing Co. Inc., 1967.

Prior, A. N. "Correspondence Theory of Truth" in: The Encyclopedia of Philosophy. ed. Paul Edwards. vol. 7, New York: Macmillan Publishing Co. Inc., 1967.

Reach, K. "The Name Relation and the Logical Antinomies" in: The Journal of Symbolic Logic. vol.3, 1938, pp.97-111.

Rees, W. J. "What Achilles Said to the Tortoise" in Mind. vol. 60, 1951, pp.241-246.

Tarski, Alfred. "On Decidable Statements in Enlarged Systems of Logic and the Concept of Truth" in Journal of Symbolic Logic. vol. 4, number 3, 1939.

Tarski, Alfred. "The Semantic Conception of Truth and the Foundations of Semantics" in Readings in Philosophical Analysis. ed. H. Feigl and W. Sellars. New York: Appleton, Century, Crafts Inc. 1949.

Tarski, Alfred. "The Concept of Truth in Formalized Languages" in: Logic, Semantics, and Metamathematics. London: Oxford University Press. 1956.

Tarski, Alfred. "The Establishment of Scientific Semantics" in: Logic, Semantics, and Metamathematics. London: Oxford University Press. 1956.

Tarski, Alfred. "On Undecidable Statements In Enlarged Systems of Logic and the Concept of Truth" in: The Journal of Symbolic Logic vol. 4 Number 3, Sept. 1939.

Tarski, Alfred. "Truth and Proof" in: Scientific American. 1969, pp.63-72.

van Heijenoort, John. "Logical Paradoxes" in: The Encyclopedia of Philosophy. ed. Paul Edwards. vol. 5, New York: Macmillan Publishing Co. Inc., 1967.

Wittgenstein, Ludwig. Tractatus Logico Philosophicus. London: Routledge, Keegan Paul, Ltd. Reprinted 1983.