NOTAS SOBRE EL "WAHRHEITSBEGRIFF", I*

MARIO GÓMEZ TORRENTE

1. Introducción

La teoría de la verdad de Alfred Tarski ha sido, sin lugar a dudas, una de las contribuciones más importantes a la lógica en el siglo XX. Solomon Feferman ha dicho de ella que es una de las "piedras angulares" de la lógica matemática (cf. Feferman (1988), p. 111). Además, ha sido un instrumento de gran importancia en otras áreas, especialmente la filosofía y la lingüística. Sus usos matemáticos y filosóficos han sido ampliamente estudiados, y se puede decir que hoy tenemos un buen conocimiento de sus implicaciones.

También hay una cantidad considerable de bibliografía acerca de aspectos históricos relacionados con la teoría de Tarski. El presente estudio es sobre todo una adición a esta bibliografía historiográfica, pero pretende ocuparse de algunos aspectos de la obra de Tarski a los que apenas se ha concedido atención, y proyectar una luz indirecta sobre ciertas cuestiones lógicas y filosóficas que surgen al considerarlos. Uno de los objetivos principales del estudio es explorar un grupo de cuestiones históricas, lógicas y filosóficas relacionadas con el teorema de la indefinibilidad de la verdad, que Tarski enunció en el mismo trabajo famoso en el que presentó por primera vez de una forma detallada su teoría de la ver-

* Partes de este trabajo fueron presentadas en conferencias en la Facultad de Filosofía y Letras de la Universidad de Valencia (1997), la Facultad de Filosofía y Letras de la Universidad de Buenos Aires (2000) y el Institut d'Histoire et Philosophie des Sciences et des Techniques, Centre National de la Recherche Scientifique, Universidad de París I (2001), y en ponencias en el Congreso del Centenario de Alfred Tarski en Varsovia (2001) y el Simposio Internacional sobre el Círculo de Viena y el Empirismo Lógico en Viena (2001). Agradezco a los auditorios de estas pláticas sus útiles comentarios.

La segunda parte de este trabajo se publicará en el nº 2 del Volumen XXI (Nota de edición).

dad. (Este trabajo es Tarski (1933), que fue traducido al alemán, con un importante nuevo *Nachwort*, como Tarski (1935); es común referirse a él, por el título de la versión alemana, como el "Wahrheitsbegriff"). Pero también es parte de los objetivos del presente estudio examinar otras varias cuestiones importantes histórica, lógica y filosóficamente que suscita la lectura del trabajo de Tarski.

La estructura del presente artículo es la siguiente. Está dividido en dos partes. La primera parte consta de las secciones 1 a 5. En la sección 2 explicamos de forma muy somera algunos aspectos del contexto histórico en el que Tarski propuso su teoría, prestando especial atención a las razones por las que el propio Tarski pensaba que era necesaria una teoría de la verdad como la que él finalmente propondría.

En las secciones 3, 4 y 5 exponemos la teoría de la verdad de Tarski. La teoría en la forma en que se la presenta hoy en día en manuales de lógica y filosofía del lenguaje es muy conocida. Sin embargo, en nuestra exposición pondremos énfasis en ciertos aspectos de la forma particular que adopta en el trabajo original de Tarski. A nosotros este trabajo previo nos será útil cuando luego comentemos las cuestiones históricas que nos interesan, especialmente las relacionadas con el teorema de la indefinibilidad de la verdad. Una parte del material en estas secciones y alguna otra es necesariamente expositivo; pero se acompaña de varias observaciones destinadas a rectificar algunas ideas que percibo como erróneas en la bibliografía secundaria, en particular la idea de que Tarski no definió o siquiera consideró en su obra de este período la noción de "verdad en una estructura". Desafortunadamente, la enunciación de la mayoría de las observaciones substantivas requiere una exposición previa de la forma de presentación que Tarski da a sus ideas. No parece haber una exposición semejante en la bibliografía sobre Tarski, que a menudo se basa en presentaciones actuales de sus ideas, inutilizables para propósitos historiográficos. Espero que la exposición del presente estudio pueda tener algún valor en sí misma como guía para la lectura del trabajo clásico de Tarski.

La segunda parte del artículo consta de las secciones 6 a 10. En las secciones 6, 7 y 8 explicamos cuál es el enunciado de Tarski del teorema sobre la indefinibilidad de la verdad y damos comentarios sobre su demostración. Aparecen cuestiones interesantes al comparar la forma que el enunciado del teorema adopta con la forma que adopta en otras presentaciones. En la sección 9 estudiamos una serie de cuestiones exegéticas relacionadas con el material del *Nachwort* de Tarski (1935), y en particular la cuestión de la exégesis correcta de la afirmación de Tarski de que una definición de verdad no puede construirse en el metalenguaje si el orden de éste no es superior al del lenguaje objeto.

En la sección 10 comparamos el enunciado de la versión del teorema de la indefinibilidad por Tarski con el enunciado de un resultado que es en cierto sentido el mismo y que fue demostrado con anterioridad por Gödel, e independientemente de Tarski. Lo interesante es discernir en qué sentido el resultado de Gödel es el mismo que el de Tarski. Esta discusión aclara de forma que nos parece iluminadora una serie de afirmaciones hechas recientemente acerca de la prioridad (de Tarski o Gödel) en el enunciado del teorema sobre la indefinibilidad de la verdad, así como acerca de las motivaciones filosóficas que subyacen en los enunciados de distintas versiones del teorema. Dicho brevemente, lo que vendremos a defender es que si bien Gödel demostró antes que Tarski una versión semántica intuitiva del teorema de la indefinibilidad, Gödel no la publicó pues no tenía una forma que le pareciera satisfactoria de enunciarla; y que merced a su profundo estudio de la noción de verdad, Tarski encontró la primera forma matemáticamente satisfactoria, y además *no semántica*, de enunciar el teorema (y luego otra forma formulada en términos de su semántica teórica, más común hoy en día que la primera formulada por Tarski, pero que en mi opinión no influye demasiado en la cuestión de la prioridad, pues es esencialmente la misma que la anterior de Gödel).

2. El contexto histórico

En un seminario de lógica que dio en la Universidad de Varsovia entre 1927 y 1929, Tarski demostró varios resultados en los que se hacía referencia a distintas nociones que luego (precisamente siguiendo un uso inaugurado por Tarski) se agruparían bajo el apelativo de 'nociones semánticas'[1], en especial resultados en los que se hacía referencia a las nociones de definibilidad y verdad en una estructura[2].

Según nos informa Vaught, ciertas dificultades encontradas por Tarski para dar una forma matemáticamente satisfactoria a los resultados presentados en el seminario de 1927-1929 fueron la causa de que Tarski buscara una teoría precisa de las nociones semánticas (cf. Vaught (1974), pp. 160 ss. y (1986), pp. 870 ss.). No existía una teoría tal en aquel momento. En particular, no existían definiciones de estas nociones en términos de los conceptos aceptados en los distintos sistemas fundacionales diseñados para la reconstrucción de la matemática clásica (por ejemplo, la teoría de los tipos de Russell y Whitehead o la teoría de conjuntos de Zermelo); por ello, los resultados existentes en los que se hacía referencia a estas nociones no podían reconstruirse en los sistemas fundacionales aceptados. Pero tampoco existía una teoría axiomática rigurosa donde se tomara a las nociones semánticas como primitivas.

[1] Aunque hay otros usos anteriores, el uso del término 'semántica' para nombrar a una teoría que trata de las nociones de verdad, verdad en una estructura, satisfacibilidad, denotación, definibilidad, etc. se debe, por lo que yo sé, a Tarski (Church (1956), p. 64, n. 140, coincide). En el resumen Tarski (1932), Tarski había usado la palabra 'semasiología'. Puede tener interés observar que Ramsey, al agrupar todas esas nociones bajo un mismo nombre para distinguir el tipo de las paradojas a que dan lugar de las paradojas de la teoría de conjuntos, las llamó 'epistemológicas' (cf. Ramsey (1925), p. 184). Gödel recoge el uso de Ramsey (cf. Gödel (1931), pp. 148-149, n. 14).

[2] Robert Vaught nos informa sobre algunos de estos resultados, y sobre la importancia que el seminario de 1927-1929 tuvo en el desarrollo de la semántica por Tarski, en Vaught (1974) y (1986).

A pesar de todo, los resultados acerca de nociones semánticas eran importantes y aun moderadamente abundantes en el momento en que Tarski publica en 1933 la versión original, en polaco, de su trabajo clásico sobre la verdad. El propio Tarski pasa revista en su trabajo a algunos de los resultados más importantes en este terreno, mencionando, entre otros, el teorema de compleción de Gödel (1930) y distintas versiones del teorema de Löwenheim-Skolem (cf. Tarski (1933), pp. 240-241, esp. p. 240 n. 1 y p. 241 n. 2). Todos estos resultados usan la noción de ser verdadera una oración en una estructura, o nociones funcionalmente equivalentes. Y Tarski dice que:

> "es evidente que todos estos resultados sólo reciben un contenido claro y pueden ser demostrados de una forma exacta si se acepta como base de la investigación una definición concreta y formulada de forma precisa de [la noción de] oración verdadera." (Tarski (1933), p. 241).

El objetivo de Tarski en su serie de trabajos sobre nociones semánticas es precisamente proponer definiciones formuladas de forma precisa de esas nociones, definiciones formuladas en términos de nociones que sean matemáticamente aceptables, y en particular aceptables en alguno de los sistemas fundacionales imperantes (el sistema escogido por Tarski variará con el tiempo, como veremos). Si es posible dar definiciones apropiadas de esas nociones, los teoremas que hacen referencia a ellas podrán ser formulados haciendo uso de las nociones definidas, y la inquietud descrita por Tarski en el texto citado habrá quedado aliviada. La alternativa de tomar a las nociones semánticas como primitivas es también considerada por Tarski en varios lugares, pero claramente pretende evitarla si es posible. La razón es, sin duda, que en la alternativa axiomática el riesgo de que subsistan formas de generar las antinomias semánticas para los primitivos semánticos del sistema no es desdeñable (cf. Tarski (1933), p. 255); en cambio, con el procedimiento definicional la consistencia de la definición dependerá exclusivamente de la consistencia de la teoría en la que se formule, que por hipótesis habrá de ser una teoría fundacional cuya consistencia tengamos razones para aceptar.

El primer trabajo donde se muestra cómo definir una de las nociones semánticas es Tarski (1931), donde Tarski examina un lenguaje en el que es posible formalizar la aritmética de los números reales, y da una definición recursiva de la noción de "conjunto definible de números reales" (definible en ese lenguaje) (cf. la Def. 9, en Tarski (1931), p. 128). La definición depende de otras que son un poco alambicadas; hoy definiríamos una noción similar de forma más simple con ayuda de la definición de verdad que el propio Tarski publicaría dos años después. Pero la esencia de la definición tarskiana de la noción de verdad se halla ya aquí. En este trabajo encontramos también, como era de esperar, la preocupación tarskiana por las cuestiones de precisión y rigor fundacional, y hallamos asimismo una reveladora afirmación acerca de la actitud de los matemáticos ante la noción de definibilidad, que podría extenderse a otras nociones semánticas:

> "la desconfianza de los matemáticos hacia la noción en cuestión la refuerza la opinión actual según la cual esta noción se encuentra fuera por completo de los límites de la matemática. Los problemas de hacer su significado más preciso, de eliminar las confusiones y malentendidos conectados con ella, y de determinar sus propiedades fundamentales pertenecen a otra rama de la ciencia —la metamatemática." (Tarski (1931), p. 110).

En su trabajo clásico sobre el concepto de verdad Tarski presentará un método de aplicación muy general que permite definir un predicado intuitivamente adecuado de verdad para oraciones de un gran número de lenguajes formalizados, y que lo hace usando sólo conceptos matemáticos no sospechosos en los *definientia*. Al mismo tiempo, Tarski muestra cómo en términos de la noción definida de verdad pueden darse definiciones intuitivamente adecuadas de las nociones semánticas de definibilidad y denotación, e indica cómo de una forma análoga a la usada para definir verdad es posible definir la noción de verdad en una estructura. Ello le permite concluir que la aceptabilidad y el rigor de los "recientes estudios metodológicos" (de Löwenheim, Skolem, Gödel y el propio Tarski, entre otros) han sido vindicados (cf. Tarski (1933), p. 266).

3. La estructura del metalenguaje y la convención T

En esta sección y la siguiente veremos cuál es el método de Tarski para definir el predicado de verdad para varios lenguajes, tal como lo presenta en Tarski (1933) y Tarski (1935). Al tiempo que lo hacemos, pondremos énfasis en ciertos aspectos de esta presentación que, en las secciones siguientes, nos será útil haber considerado.

Lo primero que hay que subrayar, como hace el propio Tarski, es que su método es eso: un *método* para dar definiciones de predicados de verdad para lenguajes particulares. No es una definición general de un predicado único de verdad o de un predicado relacional de la forma "*O* es verdadera en *L*" que es satisfecho por pares oración-lenguaje. Dado un lenguaje particular *L*, el método de Tarski nos indica cómo dar una definición de un predicado monádico "es *OV* en *L*" que es satisfecho intuitivamente precisamente por las oraciones verdaderas de *L*. Es importante notar que un lenguaje formal es para Tarski no meramente una sintaxis interpretada, sino una sintaxis interpretada para la que se enuncia también un conjunto de axiomas y reglas de inferencia especificados de manera efectiva (cf. Tarski (1933), p. 166). El término 'lenguaje' suele usarse actualmente en lógica para referirse a sintaxis interpretadas, o incluso meramente para ciertos tipos de sintaxis. En este trabajo, cuando expongamos a Tarski seguiremos su uso del término, sin entrar en la cuestión de si es más o menos afortunado que el uso actual.

La definición de "es *OV* en *L*" se formula en un metalenguaje apropiado para *L*. En un metalenguaje tarskiano para *L*, y por tanto en el *definiens* de esa definición, aparecerán sólo tres tipos de términos: (1) términos de la sintaxis de *L*, en particular expresiones para nombrar a todas las expresiones de *L*, y expresiones que denoten ciertas operaciones definidas sobre expresiones de *L*, como la operación de concatenación de expresiones de *L* (cf. Tarski (1933), p. 211); (2) términos de "*un carácter lógico general*, sacados de un sistema lo suficientemente desarrollado de lógica matemática" (Tarski (1933), p. 170; cf. también p. 211), es decir, aparato lógico-matemático que hoy veríamos como de naturaleza esencialmente conjuntista, pero que para Tarski ha de poder sacarse de uno de los sistemas fundacionales existentes; su preferencia particular en (1933) es un sistema que es una versión de la teoría simple de los tipos finitos, y que describiremos más adelante; su preferencia —y sus necesidades— cambian en (1935); en cualquier caso, ese aparato lógico-matemático debe contener símbolos para conectivas y cuantificadores y recursos para hablar de nociones como pertenencia, inclusión, funciones, secuencias finitas e infinitas, cardinalidad de conjuntos, etc.; (3) por último, deberán aparecer también los términos constantes específicos de *L* o traducciones suyas al metalenguaje (cf. Tarski (1933), pp. 210-211); a menudo, puesto que *L* será un lenguaje diseñado para formalizar alguna disciplina matemática que podrá ser reconstruida en el sistema fundacional cuyos términos aparecen en el grupo (2), no será necesario añadir términos especiales en el grupo (3). Todas las definiciones técnicas de Tarski, y en concreto las de nociones semánticas, contendrán exclusivamente términos de los grupos (1), (2) y (3). Esto es esencial para los propósitos fundacionales de Tarski, pues el uso de otros tipos de términos podría llevar a contradicciones, problemas de claridad u otras dificultades.

Correspondientes a los grupos (1) y (2) se especificarán axiomas sobre la sintaxis de *L* y axiomas y reglas de inferencia de carácter lógico-matemático "que basten para un sistema suficientemente amplio de lógica matemática" (Tarski (1933), p. 173; cf. también p. 211). Estos axiomas serán la base de la *metateoría* para *L*. Axiomas sobre la sintaxis de *L* serán, por ejemplo, que el cuantificador universal es una expresión de *L* (si el cuantificador universal es una expresión de *L*) o que la concatenación de dos expresiones cualesquiera de *L* es una expresión de *L*. Axiomas y reglas de carácter lógico-matemático serán, por ejemplo, axiomas y reglas de la lógica de enunciados y de la lógica cuantificacional, y axiomas apropiados para la teoría fundacional que se esté usando, por ejemplo, en el caso de la teoría de los tipos finitos, axiomas apropiados de extensionalidad, comprensión e infinitud. Por último, correspondientes al grupo (3) de términos, se especificarán axiomas y reglas que reproducen o traducen los axiomas y reglas del lenguaje objeto *L* (cf. Tarski (1933), p. 211).

Dar una formulación general y abstracta de su método para definir el predicado *OV* en el metalenguaje apropiado para un lenguaje *L* cualquiera de los que Tarski tiene en mente sería una tarea algo ardua y, lo que es peor, la descripción resultante no sería muy iluminadora. Por ello Tarski decide ilustrar su método describiendo su aplicación a un lenguaje particular especialmente simple, que él llama 'el lenguaje del cálculo de clases' (LCC). Nosotros presentaremos una versión "modernizada" de LCC, que sólo difiere de la versión de Tarski en que usa una notación habitual hoy en día para la formación de fórmulas lógicamente complejas en lugar de la notación polaca (de Lukasiewicz) usada por Tarski; esto hará las cosas más fáciles sin afectar en nada a la sustancia.

Los signos primitivos de LCC son el cuantificador universal ('∀'), el signo de disyunción ('v'), el signo de negación ('¬'), paréntesis ('(', ')'), un predicado diádico ('I'), la letra 'x', y un acento subíndice ('͵') para generar ω variables por posposición a 'x' (para abreviar, usaremos la notación 'x_n' para la variable en que la letra 'x' va seguida de un número positivo *n* de acentos subíndices). El recorrido de las variables en la interpretación deseada de LCC es el conjunto de todas las subclases de la clase de todos los individuos (del universo). El significado deseado de 'I' es la relación de inclusión entre subclases de esa clase. Las intepretaciones de los otros signos lógicos[3] son las

[3] 'I' es también un signo lógico, como subrayaremos luego.

normales. Las fórmulas atómicas son las de la forma Ix_kx_l. Las fórmulas complejas se obtienen por las operaciones de negación, disyunción (rodeada por paréntesis) y cuantificación universal con respecto a las variables.

En el metalenguaje para LCC hay, como dijimos, nombres para todas las expresiones de LCC (es decir, para las series finitas formadas por signos primitivos de LCC). Esto se consigue incluyendo en ese metalenguaje nombres para los signos primitivos y obteniendo nombres para las otras expresiones por aplicaciones sucesivas de la función binaria de concatenación. A estos nombres de expresiones en el metalenguaje Tarski los llama 'nombres descriptivo-estructurales' (y los distingue de los 'nombres por entrecomillado' disponibles en el lenguaje escrito usual, así como de otros nombres para expresiones). Tarski también tiene un sistema para abreviar en el metalenguaje nombres descriptivo-estructurales, y en general expresiones de naturaleza descriptivo-estructural que contengan variables metalingüísticas que varíen sobre expresiones del lenguaje objeto. Este sistema es de difícil legibilidad y no lo explicaremos. De nuevo sin afectar a la sustancia de lo que nos interesa en este trabajo, usaremos aquí el conocido artificio de las semicomillas (debido a Quine (1951), § 6) para abreviar expresiones de naturaleza descriptivo-estructural. Con este artificio podemos abreviar, por ejemplo, la expresión 'el cuantificador universal' por medio de la expresión '⌜$\forall$⌝'; la expresión 'la expresión formada concatenando el cuantificador universal con la primera variable con la letra i mayúscula con la primera variable con la primera variable, en ese orden', por medio de la expresión '⌜$\forall x_1 I x_1 x_1$⌝', y la expresión 'la expresión formada concatenando el cuantificador universal con la n-ésima variable con la letra i mayúscula con la n-ésima variable con la n-ésima variable, en ese orden', por medio de la expresión '⌜$\forall x_n I x_n x_n$⌝'.

La comodidad de estas abreviaturas es cosa sabida. Sin embargo, no parece haber una conciencia general de que no todo lo que se puede hacer con la notación de Tarski puede

hacerse con la notación de Quine. En particular, la notación de Tarski se puede usar para abreviar expresiones que hacen referencia a disyunciones (u otras operaciones proposicionales similares) con un número variable de miembros, así como expresiones que hacen referencia a ristras de un número variable de cuantificadores (cf. Tarski (1933), pp. 175-176), mientras que la notación de Quine debe como mínimo modificarse para esos propósitos. Así, por ejemplo, con la notación de Tarski puede decirse fácilmente y sin recurrir a "puntos suspensivos" que para cualquier entero positivo *n* y cualquier fórmula *F* la expresión $\forall x_1 \ldots \forall x_n F$ es una fórmula, pero no en la notación de Quine. (Nótese que '⌜$\forall x_1 \ldots \forall x_n F$⌝' no es una abreviatura quineana con sentido).

Usando sus abreviaturas Tarski da definiciones intachablemente rigurosas de 'fórmula' (Tarski usa 'función oracional' como sinónimo de 'fórmula'), 'aparecer libre una variable x_n en una fórmula *F*' y 'oración', siempre referidas a LCC. Esas definiciones son similares a las que encontramos en cualquier libro de lógica reciente para lenguajes como LCC (aquéllas son más rigurosas que la mayoría de éstas, pues Tarski fue uno de los que nos mostraron la forma de "relajarnos"): las dos primeras son definiciones recursivas que se pueden hacer explícitas a la manera de Frege y Dedekind cuantificando sobre conjuntos de expresiones de LCC (lo cual es posible hacer en el metalenguaje para LCC); la tercera define una oración como una fórmula en la que ninguna variable aparece libre.

Siguiendo sus propios requisitos, Tarski también da un cálculo deductivo para LCC. Este cálculo tiene como base un sistema de axiomas y reglas de inferencia lógicas y añade como axiomas acerca de la noción de inclusión cinco postulados que conforman un sistema equivalente a uno ofrecido por Huntington en 1904. Luego define en el metalenguaje los predicados '*O* es una oración demostrable a partir de un conjunto de oraciones *X*' y '*O* es una oración demostrable' (en el cálculo deductivo de LCC).

Después Tarski pasa a indicar cómo es posible definir un predicado 'es *OV*' que intuitivamente será coextensional con el predicado natural de verdad para oraciones de LCC. En primer lugar, Tarski enuncia su famosa 'convención T'[4]. La convención T es una definición enunciable en el metametalenguaje de LCC (similares definiciones, o "convenciones T" se enunciarán en los metametalenguajes de otros lenguajes para los que queramos definir un predicado coextensional con el predicado intuitivo de verdad), y que intuitivamente da condiciones necesarias y suficientes para que un predicado 'es *OV*' sea coextensional con el predicado natural de verdad para oraciones de LCC. La convención T para LCC, por ejemplo, dirá:

> una definición formalmente correcta de un símbolo *OV* es una *definición adecuada de verdad* si y sólo si tiene como consecuencias (módulo el sistema deductivo de la metateoría):
>
> (α) todas las oraciones que se obtienen a partir de la expresión '*OV*(*x*) si y sólo si *p*' reemplazando el símbolo '*x*' por un nombre descriptivo-estructural de una oración cualquiera de LCC y el símbolo '*p*' por la expresión que es la traducción de aquella oración al metalenguaje;
>
> (β) la oración 'para todo *x*, si *OV*(*x*) entonces x es una oración de LCC'.

Con lo de que una definición sea formalmente correcta Tarski quiere decir que se ha de construir usando reglas no controvertidas para dar definiciones de nuevas expresiones en términos de expresiones ya disponibles y que sólo ha de contener expresiones de los grupos (1), (2) y (3) del metalenguaje mencionadas más arriba.

[4] La 'T' es porque '*truth*' empieza con 't', y el inglés es el idioma en que más se ha hablado de la convención. En el original polaco la convención lleva el nombre 'P', en el alemán el nombre 'W' y en castellano podría llevar el nombre 'V'. Como a estas alturas podría ser confuso usar otro nombre que 'T', ese será el nombre usado aquí. Podemos, si lo deseamos, fingir que la 'T' está ahí por Tarski.

Como Tarski observa en Tarski (1933), p. 188, n. 1, una vez sometida la metateoría al proceso de formalización sería fácil dar definiciones precisas en la *metameta*teoría de nociones como "definición formalmente correcta", "nombre descriptivo-estructural", "traducción de una oración de LCC al metalenguaje de LCC", etc.

¿Cuál es el motivo por el que podemos pensar que la convención T enuncia una condición necesaria y suficiente para que '*OV*' sea coextensional con el predicado natural de verdad para las oraciones de LCC? El motivo lo da un argumento intuitivo como el siguiente. Sea '*OV*' un predicado definido cuya definición satisface la convención T. Supongamos primero que '*p*' es una oración de LCC que satisface el predicado '*OV*'[5] (independientemente de si la oración metalingüística "*OV*('*p*')" es demostrable en la metateoría); en otras palabras, supongamos que *OV*('*p*'). Sabemos que "*OV*('*p*') si y sólo si *p*" es demostrable en la metateoría, y por tanto que es verdadera (pues los axiomas de la metateoría son intuitivamente verdaderos y sus reglas de inferencia sólo llevan de verdades a verdades); así, por nuestro supuesto, podemos concluir que *p*, y por tanto que '*p*' es verdadera en el sentido intuitivo. Por otro lado, supongamos que '*p*' es verdadera en el sentido intuitivo; entonces *p*, y así, por el mismo razonamiento que antes, *OV*('*p*').

Un argumento relacionado con este es ofrecido en Coffa (1991), pero contiene un extraño error acerca de la naturaleza del razonamiento requerido. Coffa dice:

> "Si un predicado '*Tr**' que se aplica sólo a oraciones de OL [el lenguaje objeto] también satisface la convención (T), entonces para cada oración *X* de OL seremos capaces de probar en ML [el metalenguaje] que *Tr**(*X*)=p [*sic*; '=' parece haber de substituirse por un bicondicional]; por tanto, *Tr**=*Tr* ['*Tr*' es el predicado definido de verdad]. Dado el supuesto (obvio para

[5] No somos completamente rigurosos en el uso del entrecomillado en la exposición de argumentos que, como en este caso, ganan claridad como resultado.

> Tarski) de que los teoremas metalingüísticos son verdaderos, tenemos pues que una oración X de OL será un caso particular [('instance')] de '*Tr*' precisamente cuando su traducción *p* (y por tanto *X* misma) sea verdadera." (Coffa (1991), p. 296).

El error contenido en este pasaje es, desde luego, que una prueba metalingüística de que todo predicado de oraciones del lenguaje objeto que satisfaga la convención T es coextensional con '*Tr*' no implica nada acerca de la relación entre '*Tr*' y el predicado intuitivo de verdad. La razón es que el predicado intuitivo de verdad no es un predicado del que se pueda hablar, o que pueda aparecer como valor de las variables metalingüísticas para expresiones, en un metalenguaje tarskiano (so pena de introducir dudas acerca de su consistencia); por tanto, el argumento de Coffa no establece nada acerca de él. Sólo un argumento intuitivo no formalizable en un metalenguaje tarskiano puede establecer la coextensionalidad de un predicado definido de verdad con uno intuitivo.

El razonamiento intuitivo no aparece en Tarski, y sin duda no lo habría considerado un razonamiento matemáticamente satisfactorio (aunque se aproxima mucho a él en Tarski (1933), p. 187, y aún más en Tarski (1944), § 4). Pero da una idea del tipo de consideraciones informales que sin duda justificaron para él el *desideratum* de que una definición del predicado deseado '*OV*' satisfaga la convención T. Lo que busca Tarski puede resumirse, pues, así: busca un predicado definible con ayuda exclusivamente de los términos de los grupos (1), (2) y (3) del metalenguaje y que satisfaga la convención T (y del que pueda demostrarse que lo hace —en la metametateoría).

4. La definición de conceptos semánticos

La forma en que Tarski logra definir un predicado semejante es conocida. Tarski observa que aunque surge la idea de dar una definición recursiva directa del predicado '*OV*',

> "en general las oraciones compuestas no son de ningún modo compuestos de *oraciones* simples. Las funciones oracionales sí surgen de este modo de funciones elementales, es decir, de inclusiones; las oraciones son por el contrario ciertos casos especiales de funciones oracionales. En vista de este hecho, no puede darse ningún método que nos permita definir directamente por medios recursivos el concepto requerido." (Tarski (1933), p. 189).

(Nótese que esta observación no excluye la posibilidad de que para *algunos* lenguajes sea posible dar una definición recursiva directa de '*OV*'. Pero en general será imposible especificar un conjunto de oraciones que pueda servir como base para una recursión apropiada.) Sin embargo, continúa,

> "se presenta la posibilidad de introducir un concepto más general que sea aplicable a cualquier función oracional, pueda ser definido recursivamente, y que, cuando se aplique a oraciones, nos lleve directamente al concepto de verdad. Estos requisitos los cumple la noción de la *satisfacción de una función oracional dada por objetos dados*, y en el caso presente por clases dadas de individuos." (Tarski (1933), p. 189).

Tarski define entonces recursivamente la noción de satisfacción de una fórmula cualquiera *F* por una secuencia infinita *f* (con dominio los enteros positivos y recorrido incluido en la clase de las subclases del dominio de todos los individuos). Intuitivamente, *F* es satisfecha por *f* cuando *F* es "verdadera" si interpretamos cada variable libre x_n que aparezca en *F* como si fuera un nombre de la clase f_n. La definición es la siguiente (usando nuestra notación):

> La secuencia infinita de clases *f* satisface la fórmula *F* si y sólo si *f* y *F* son tales que o bien (α) existen números naturales k y l tales que $F=\ulcorner Ix_kx_l\urcorner$ y $f_k \subseteq f_l$; o (β) hay una fórmula *G* tal que $F=\ulcorner\neg G\urcorner$ y *f* no satisface la fórmula *G*; o (γ) hay fórmulas *G* y *H* tales que $F=\ulcorner(G\vee H)\urcorner$ y *f* satisface *G* o *f* satisface *H*; o finalmente (δ) hay un número natural *k* y una fórmula *G* tales que $F=\ulcorner\forall x_k G\urcorner$ y toda secuencia infinita de clases que difiera de *f* a lo sumo en el lugar *k*-ésimo satisface la fórmula *G*. (cf. Tarski (1933), p. 193).

De nuevo Tarski subraya que esta definición recursiva puede ponerse en forma explícita o normal (cf. Tarski (1933), p. 193, n. 1). La definición explícita es la siguiente:

> La secuencia infinita de clases f satisface la fórmula F si y sólo si el par $<f, F>$ pertenece a toda relación R entre secuencias y fórmulas tal que (α) si existen números naturales k y l tales que $G=\ulcorner \mathrm{I}x_k x_l \urcorner$ y $g_k \subseteq g_l$, entonces $<g, G>$ pertenece a R; (β) si $<g, G>$ no pertenece a R, entonces $<g, \ulcorner\neg G\urcorner>$ pertenece a R; (γ) si $<g, G>$ pertenece a R o $<g, H>$ pertenece a R, entonces $<g, \ulcorner(G \vee H)\urcorner>$ pertenece a R; (δ) para todo número natural k y fórmula G, si toda secuencia infinita de clases h que difiera de una secuencia g a lo sumo en el lugar k-ésimo es tal que $<h, G>$ pertenece a R, entonces $<g, \ulcorner\forall x_k G\urcorner>$ pertenece a R.

A continuación Tarski observa que la definición de satisfacción tiene como consecuencia que el que una secuencia f satisfaga una fórmula F o no sólo depende de lo que f asigne a (los subíndices de) las variables que aparezcan libres en F. (Esta observación se pone más adelante en forma más precisa, en el Lema A, según el cual, "Si la secuencia f satisface la función oracional F, y la secuencia infinita de clases g es tal que para todo k, $f_k = g_k$ si v_k es una variable libre de F, entonces la secuencia g también satisface la función F" (Tarski (1933), p. 198).) Esto quiere decir que si F es una oración —una fórmula sin variables libres— entonces o toda secuencia satisface F o ninguna lo hace. Además, la definición de satisfacción se ha construido de forma que las oraciones intuitivamente verdaderas son las del primer tipo, las satisfechas por toda secuencia, y por tanto es posible definir de esta forma el predicado '*OV*':

> *O* es *OV* si y sólo si *O* es una oración de LCC y toda secuencia infinita de clases satisface *O*. (cf. Tarski (1933), p. 195).

La comprobación de que esta definición es apropiada se haría demostrando en la metametateoría que es una "definición adecuada de verdad" en el sentido de la convención T. Ello requeriría la formalización de la metateoría, tarea ardua

donde las haya que Tarski no está dispuesto a realizar. Sin embargo, el hecho es intuitivamente claro y Tarski se conforma con mostrar cómo se demostrarían en el metalenguaje algunos bicondicionales de los mencionados en la parte (α) de la convención T.

Usando la definición de satisfacción, o construyendo definiciones análogas, Tarski da definiciones de otros conceptos semánticos, como definibilidad y verdad en un dominio (siempre para LCC). La noción de definibilidad la obtiene primero definiendo "una fórmula F define un conjunto P de clases (de individuos)": F define a P si y sólo si hay una variable x_k que es la única variable libre de F y tal que para toda secuencia f, f satisface F si y sólo si f_k pertenece a P. Luego se define definibilidad así: un conjunto P (de clases de individuos) es definible (en LCC) si y sólo si hay alguna fórmula (de LCC) que lo define (cf. Tarski (1933), p. 194, n. 1).

La definición de verdad en un dominio se da de forma muy similar a la definición de verdad. Primero Tarski define la noción de satisfacción de una fórmula por un dominio con respecto a una secuencia, como sigue (usando nuestra notación):

> El dominio d satisface la fórmula F con respecto a la secuencia infinita f de subclases de d si y sólo si f y F son tales que o bien (α) existen números naturales k y l tales que $F=\ulcorner \mathrm{I}x_k x_l\urcorner$ y $f_k \subseteq f_l$; o (β) hay una fórmula G tal que $F=\ulcorner\neg G\urcorner$ y d no satisface la fórmula G con respecto a f; o (γ) hay fórmulas G y H tales que $F=\ulcorner(G \vee H)\urcorner$ y d satisface G con respecto a f o d satisface H con respecto a f; o finalmente (δ) hay un número natural k y una fórmula G tales que $F=\ulcorner\forall x_k G\urcorner$ y para toda secuencia infinita de clases g que difiere de f a lo sumo en el lugar k-ésimo, d satisface la fórmula G con respecto a g. (Cf. Tarski (1933), p. 200).

Luego verdad en un dominio se define así: *O* es una oración verdadera en un dominio d si y sólo si *O* es una oración y *O* es satisfecha por d con respecto a toda secuencia f de subclases de d (cf. Tarski (1933), p. 200).

El procedimiento de Tarski de explicar su método para la definición de conceptos semánticos por medio del caso particular de LCC ha dado lugar en este caso a lo que en mi opinión son algunos errores de apreciación. Varios autores han sostenido, en particular, que con base en Tarski (1933) y en su obra de este período, no es posible concluir que Tarski definiera entonces la noción de verdad en una estructura, o, de manera algo más fuerte, que se puede concluir que no la definió –incluso si tenía a su disposición los instrumentos teóricos para hacerlo. Veamos las razones de esos autores, para después argumentar que no son concluyentes y que son en realidad demostrablemente incorrectas.

Wilfrid Hodges relata lo que llama

> "una experiencia desconcertante. Leí la famosa monografía de Tarski 'El concepto de verdad en los lenguajes formalizados' para ver lo que él mismo dice acerca de la noción de verdad en una estructura. La noción simplemente no estaba allí. Esto parecía curioso, de manera que miré en otros artículos de Tarski. Por lo que yo pude descubrir, la noción aparece por primera vez en la alocución Tarski (1952) al Congreso Internacional de Matemáticos de 1950 y en su artículo Tarski (1954). Pero incluso en estos artículos no la define. (...) Creo que la primera vez que Tarski presentó explícitamente su definición matemática de verdad en una estructura fue en Tarski y Vaught (1957)." (Hodges (1985/6), pp. 137-138).

Hodges concluye que Tarski debió ver algún tipo de dificultad entrañada por la noción de estructura, al menos en sus artículos tempranos, y que sólo cuando se convenció de que no la había empezó a hablar de la noción y a definirla por medio de su aparato semántico. Hodges no hace muchas o muy claras hipótesis sobre cuál pudo haber sido esa dificultad, pero identifica claramente tres "partes de la definición de verdad en una estructura que están ausentes de los trabajos tempranos de Tarski" (Hodges (1985/6), p. 138), y cuya ausencia cabe suponer relacionada con la dificultad vista por Tarski.

La primera de esas partes es la noción misma de estructura, en cuanto que distinta de la noción de secuencia infinita de asignaciones posibles a las variables. "Una secuencia infinita de objetos no es en absoluto lo mismo que una estructura. Por ejemplo, los grupos finitos no son secuencias infinitas de objetos, y no hay una manera natural de convertirlos en secuencias infinitas de objetos" (Hodges (1985/6), p. 138). La segunda noción ausente en el Tarski temprano es "la distinción entre variables y constantes no-lógicas" (Hodges (1985/6), p. 42). Las constantes no-lógicas de un lenguaje formal de primer orden son susceptibles de recibir interpretaciones distintas en distintas estructuras, pero no por ser variables de esos lenguajes, que son susceptibles de recibir interpretaciones distintas dentro de una misma estructura. "Cuando nos preguntamos si la ley conmutativa es verdadera en un grupo, hemos de mantener un nombre fijo para la operación de producto del grupo, pero no tendría caso asignar elementos del grupo a todas las variables libres del lenguaje formal" (Hodges (1985/6), p. 146). La tercera noción ausente es "la noción de un símbolo constante no interpretado que recibe una interpretación al ser aplicado a una estructura particular" (Hodges (1985/6), p. 147). Esta es una noción habitual en álgebra[6], pero según Hodges "Tarski parece haber tenido extremado cuidado de no mencionar esta noción, tanto en los años 30 como más tarde" (Hodges (1985/6), pp. 147-148).

Aunque Hodges no menciona la definición por Tarski de la noción de verdad en un dominio en Tarski (1933), cabe suponer que esta noción le pareció en su lectura una noción específica que Tarski no veía o no pudo ver como ilustración de un método general para definir la noción de verdad en una estructura, seguramente por razones relacionadas con la ausencia en Tarski de las partes de la noción de estructura que aísla Hodges, o por otras razones. Un autor inspirado en Hod-

[6] Habitual, aunque no interviene necesariamente en el caso de la noción de verdad en una estructura para todos los lenguajes de primer orden –muchos tienen interpretaciones deseadas únicas.

ges, Milne (1999), ofrece algunos argumentos en este sentido, dos de los cuales explicaremos someramente.

El primero de estos argumentos consiste en señalar que los dominios de las variables de LCC no son los mismos que los dominios con respecto a los cuales puede ser verdadera o falsa una oración de LCC: "en Tarski (1935) el dominio *no* tiene por qué ser el recorrido de los cuantificadores de primer orden. En el caso del cálculo de clases, el dominio está un orden por debajo del recorrido de los cuantificadores de primer orden" (Milne (1999), p. 154). Por tanto, concluye Milne, Tarski no *podía* estar viendo su noción de verdad en un dominio para LCC como ilustración de un método para definir la noción de verdad en una estructura de primer orden. La razón de Milne parece ser que esta noción presupone que no tiene sentido que el dominio de cuantificación sea otro que el de la interpretación, y en Tarski esto tiene sentido para LCC. Si LCC es un lenguaje de primer orden Tarski debería haber rechazado que el dominio de una interpretación para el lenguaje fuera otro que su dominio de cuantificación, pero no lo hizo.

El segundo de estos argumentos se basa en la idea de que cuando Tarski considera la cuestión de si una oración es consecuencia lógica de otras o independiente de otras, no considera interpretaciones alternativas del lenguaje en las que varíe el dominio de unas a otras. Según Milne, para Tarski una oración conclusión se sigue de otras cuando, una vez *fijado* el dominio de su lenguaje común, la oración conclusión resulta ser verdadera en todas las interpretaciones *sacadas de ese dominio* en las que las otras oraciones son verdaderas: "la consecuencia lógica queda determinada por el cambio de interpretación del vocabulario no-lógico exclusivamente" (Milne (1999), p. 158). Milne se basa en esto para pensar que Tarski no disponía en esta época de la noción de estructura, pues es esencial a esta noción que un mismo lenguaje tenga estructuras con diferentes dominios. Milne busca apoyo para su tesis en la observación de que Tarski (1936) no menciona explícitamente el cambio de dominio en su esbozo de definición de la noción de consecuencia lógica. Milne comparte aquí

una (en mi opinión sorprendentemente) popular exégesis de Tarski (1936) que parece haberse originado en Corcoran (1972), p. 43, y que aparece en varios escritos de Etchemendy (véase por ejemplo, Etchemendy (1988), p. 73)[7].

Examinemos los argumentos de Hodges y Milne. Quizá el argumento más claramente falaz sea el primero de Milne. Sin duda en una interpretación para un lenguaje de primer orden el dominio ha de coincidir con el dominio de cuantificación. Pero eso sólo apoya la tesis de Milne dado el supuesto de que LCC sea un lenguaje de primer orden. La realidad es que LCC es un lenguaje de *segundo* orden, tanto en el sentido que tiene esta expresión hoy en día como en el que tiene en Tarski (1933) (cf. p. 220)[8]: es de segundo orden porque las variables de tipo

[7] Un tercer argumento de Milne no es expresado de manera muy precisa por este autor, pero parece basarse en la idea de que Tarski no pudo haber tenido en mente la noción de verdad en una estructura porque no usa la *expresión* 'verdadero en', mientras que según Milne es "la obvia expresión que usar" (Milne (1999), p. 144) si uno tiene en mente la noción de verdad en una estructura –según Milne, "Tarski (...) evitó escrupulosamente el uso de la frase 'verdadero en un modelo' y términos similares" (Milne (1999), p. 144). En realidad, ni siquiera el dato empírico es correcto, pues como vimos Tarski usa la expresión 'verdad en un dominio' para la noción que define para LCC. (Milne es curiosamente consciente de ello, pero en mi opinión bordea el absurdo al apoyar su idea diciendo "excepto cuando introduce por primera vez la noción (...), Tarski siempre tiene cuidado de hablar de *corrección* de oraciones con respecto a un dominio, no *verdad*" (Milne (1999), p. 153); confieso que se me escapa en virtud de qué sutileza el usar poco un nombre viene a ser lo mismo que no usarlo. Un dato que sí es correcto, pero no señalado por Milne, es que el Tarski de esta época reserva la expresión 'verdad', ya sea al referirse a la noción absoluta o a la noción relativa de verdad en una estructura (dominio en el caso de LCC), para *oraciones* y nunca la usa para nociones en que la propiedad semántica pertinente se dice de fórmulas abiertas (funciones oracionales abiertas). Pero incluso si Tarski no hubiera usado la *expresión* 'verdad en un modelo' o expresiones similares, quedaría en pie la cuestión de si usó la *noción*. Esta es la cuestión que tratamos en el texto principal, y que en mi opinión se contesta afirmativamente.

[8] En la versión inglesa de Tarski (1933), p. 221, se alude a LCC como un lenguaje de primer orden ("the method of construction sketched there [en la §3, donde se dan las definiciones semánticas para LCC] can be applied as a whole to other languages *of the 1st order*"). Lo que hemos pues-

más alto varían sobre subclases de los dominios de sus interpretaciones posibles (en el caso de LCC no hay variables de primer orden). Y naturalmente, tanto hoy en día como en el trabajo de Tarski, una estructura (estándar) de segundo orden sólo especifica un dominio de individuos y no uno de clases[9].

En cuanto a la razón de que una interpretación en el caso particular de LCC pueda venir dada *simplemente* por un dominio de objetos, esta es que el símbolo 'I' de LCC es una *constante lógica* (una constante de tercer orden) no susceptible de reinterpretación en distintas interpretaciones. Así, Tarski dice que el grupo de términos lógicos del metalenguaje y el grupo de términos constantes del metalenguaje que corresponden a, o traducen términos constantes del lenguaje objeto "se solapan parcialmente, y en aquellos casos en que la lógica matemática, o un fragmento de ella, es el objeto de la investigación (como en el caso del cálculo de clases), incluso se combinan para formar *un* grupo" (Tarski (1933), pp. 211-212). El haber visto a 'I' como una constante no-lógica posiblemente explique en gran medida la resistencia de Hodges y Milne a ver la definición de la noción de verdad en un dominio (para LCC) dada en Tarski (1933) como una ilustración de un método que sirve también para definir la noción de verdad en una estructura. Lo cierto es que LCC no contiene constantes no-lógicas, y el argumento de Milne resulta inconcluyente[10].

to en cursiva es una clara errata, como se comprueba en la versión alemana. (Quizá la errata haya contribuido al error de Milne.) Debería decir "de la primera clase" ("of the first kind") –para las clases de lenguajes distinguidas por Tarski, en cuya discusión no podemos entrar aquí, véase Tarski (1933), p. 220.

[9] Estrictamente hablando, la sintaxis de LCC es un fragmento de lo que dadas algunas convenciones actuales se llamaría un lenguaje de segundo orden, en el que aparece un predicado (de tercer orden) que se predica de pares de variables de segundo orden.

[10] Tarski (1933), p. 239, subraya que la noción de verdad en un dominio puede definirse análogamente también para cualquier otro lenguaje sin constantes no-lógicas, así como para lenguajes con constantes no-lógicas. Aunque Tarski no es muy explícito, esta noción para lenguajes con constantes no lógicas es quizá diferente de la noción actual de verdad en una estructura, pues tal vez Tarski no contempla cambios de interpreta-

El segundo argumento de Milne se basa en una exégesis en mi opinión obviamente incorrecta de Tarski (1936). Si bien es cierto que Tarski no menciona explícitamente cambios de dominio y sólo menciona cambios de interpretación de las constantes no-lógicas en la definición de consecuencia lógica, es claro que no lo hace porque tácitamente está suponiendo que el dominio de la interpretación de partida está nombrado por una constante no-lógica (una práctica inusual hoy en día, pero no esencialmente diferente de la usual). La evidencia textual en los trabajos tempranos de Tarski es considerable. Examinemos un texto especialmente claro, el manual Tarski (1937), pero antes notemos que en el mismo texto aparece también la noción de estructura, en el sentido de un sistema, para algunos lenguajes finitos, que consta de un dominio y de interpretaciones de las constantes no-lógicas extraídas de ese dominio[11]. Veremos pues, de paso, que la primera parte de la noción de verdad en una estructura, que a Hodges le parecía ausente de la obra temprana de Tarski, aparece sin embargo en ella, y muy claramente.

Tarski (1937) formula una teoría de primer orden muy simple para propósitos pedagógicos, cuyo lenguaje tiene los símbolos 'S' y '≅' como primitivos. "El primero es una abreviatura del término 'el conjunto de todos los segmentos'; el últi-

ción de las constantes no lógicas de dominio a dominio. Pero esa posibilidad no excluye que viera la definición de la noción de verdad en un dominio para lenguajes sin constantes no lógicas como una ilustración de un método general aplicable a la definición de la noción de verdad en una estructura, y quizá incluso como un caso particular de la noción de verdad en una estructura. Esta noción aparece más claramente en otros textos que comentaremos después, pero también aparece implícitamente en Tarski (1933), p. 199, n. 3, donde se la relaciona claramente con la noción de verdad en un dominio. La n. 3 de Tarski (1933), p. 199, se comenta más adelante en el texto principal.

[11] Los comentarios de este párrafo y el siguiente son paralelos a los de Gómez Torrente (1996), pp. 143-144, donde hay también una discusión más detallada de estos asuntos. Para trabajos que apoyan puntos de vista similares (aunque con otras consideraciones que las nuestras) véanse Ray (1996), sec. 3, Sher (1991), p. 41, y Soames (1999), p. 118.

mo designa la relación de congruencia" (cf. Tarski (1937), pp. 84-85). Los axiomas de la teoría son "Axioma 1. Para cada elemento x del conjunto S, $x \cong x$. (...) Axioma 2. Para cualesquiera elementos x, y, z del conjunto S, si $x \cong z$ y $y \cong z$, entonces $x \cong y$" (Tarski (1937), p. 85). Tarski describe a continuación, con palabras casi completamente idénticas a las de Tarski (1936), el primer paso en el camino hacia la noción de modelo para esa teoría: convertir sus axiomas y teoremas en funciones oracionales:

> "Reemplacemos los términos primitivos en todos los axiomas y teoremas de nuestra teoría por variables apropiadas, por ejemplo, el símbolo 'S' por la variable '*K*' que denota clases, y el símbolo '≅' por la variable '*R*' que denota relaciones (...). Los enunciados de nuestra teoría no son ya oraciones, sino que se convierten en funciones oracionales que contienen dos variables libres, '*K*' y '*R*', y que expresan, en general, el hecho de que la relación *R* tiene tal o cual propiedad en la clase *K*." (Tarski (1937), p. 86).

Entonces Tarski pasa a introducir la noción de modelo de una teoría, no como una noción general, sino usando su pequeña teoría como ejemplo:

> "Si una relación *R* es reflexiva y tiene la propiedad P en una clase *K* [tener P consiste en que para cualesquiera elementos x, y, z de la clase *K*, si x*R*z y y*R*z, entonces x*R*y], decimos que *K* y *R* juntos forman un *modelo* o *realización del sistema de axiomas* de nuestra teoría, o simplemente, que satisfacen los axiomas." (Tarski (1937), p. 87).

Tarski (1936), p. 417 habla de los modelos, que ha introducido por el mismo método y casi con las mismas palabras que en Tarski (1937), como "secuencias de objetos". Es pues muy claro que Tarski usa sin problema la noción de una secuencia (posiblemente finita, como en el caso citado) de objetos que puede proporcionar una interpretación verdadera de un conjunto de oraciones (axiomas, en el texto citado), y que

la noción que usa es claramente distinta de la de secuencia infinita de asignaciones a las variables del lenguaje original de la teoría –estas secuencias infinitas de asignaciones a las variables originales no forman parte de lo que Tarski llama modelos de los axiomas del lenguaje original.

Es importante observar que Tarski (1933), p. 199, n. 3, señala que la escuela de Hilbert emplea el concepto de verdad en un dominio pero aplicándolo a fórmulas con variables libres de segundo orden (en los lugares donde hoy pondríamos predicados no lógicos) y con todas las variables de primer orden ligadas. (Estas fórmulas son funcionalmente equivalentes a las de los lenguajes de primer orden.). De una fórmula de este tipo se dice que es válida en un dominio cuando todas las secuencias de asignaciones a las variables de segundo orden (sacadas de ese dominio) la satisfacen y, generalmente válida cuando es válida en todo dominio. Nótese que la definición básica para todas estas nociones es la de satisfacción de una fórmula por una *secuencia de objetos* que interpreta las variables *de segundo orden*, y que Tarski está obviamente dando a entender que esta noción se puede definir usando sus métodos, como vimos al señalar que decía que los resultados acerca de las nociones hilbertianas, entre otras, "sólo reciben un contenido claro" (Tarski (1933), p. 241) aplicando sus métodos u otros similares. Esta es la noción que está en juego también en la definición de la noción de modelo.

Seamos un poco más precisos. Cuando Tarski quiere contemplar la posibilidad de constantes no-lógicas en ciertos lenguajes, las sustituye por variables libres *nuevas* y del mismo tipo, susceptibles de recibir asignaciones diferentes en distintas "secuencias de objetos" apropiadas para esos lenguajes. Por ejemplo, digamos que $\langle A, T \rangle$ es una "secuencia de objetos" para el lenguaje de primer orden *L* cuyo único símbolo no-lógico es el predicado binario '≅' (por simplicidad, omitamos el símbolo para el dominio). Tarski procedería así. Primero sustituiría el símbolo no lógico por la variable nueva '*R*' en cada fórmula *F* de *L* obteniendo una fórmula *F*', y por tanto un conjunto L' de funciones oracionales. Después definiría la siguiente noción:

> una fórmula F de L' es satisfecha por la secuencia $<A, T>$ con respecto a una secuencia f de objetos de A, si y sólo si o bien (α) existen números naturales k y l tales que $F = \ulcorner Rx_k x_l \urcorner$ y $<f_k, f_l> \in T$; o (β) hay una fórmula G tal que $F = \ulcorner \neg G \urcorner$ y $<A, T>$ no satisface la fórmula G con respecto a f; o (γ) hay fórmulas G y H tales que $F = \ulcorner (G \vee H) \urcorner$ y $<A, T>$ satisface G con respecto a f o $<A, T>$ satisface H con respecto a f; o finalmente (δ) hay un número natural k y una fórmula G tales que $F = \ulcorner \forall x_k G \urcorner$ y para toda secuencia infinita g de objetos de A que difiere de f a lo sumo en el lugar k-ésimo, $<A, T>$ satisface la fórmula G con respecto a g.

Y luego definiría la noción básica del siguiente modo: una oración O de L tiene como modelo a (o es satisfecha por) una secuencia de objetos $<A, T>$ si y sólo si para toda secuencia f de objetos de A, O' es satisfecha por la secuencia $<A, T>$ con respecto a f. Obsérvese que O es verdadera en la estructura $<A, T>$ (en nuestro sentido) si y sólo si O tiene como modelo a la secuencia $<A, T>$ (en el sentido recién explicado).

En contra de la idea de Milne y otros acerca de la invariabilidad del dominio en la definición de consecuencia lógica, en las páginas inmediatamente siguientes de su manual Tarski muestra que los Axiomas 1 y 2 tienen diferentes modelos, con diferentes dominios (un modelo tiene como dominio el conjunto de todos los segmentos, otro el conjunto de los números naturales, otro el de los enteros), y cómo usando uno de estos modelos (uno que tiene como dominio el conjunto de los enteros) puede uno mostrar que una cierta oración, que es falsa en ese modelo de los axiomas, no se sigue de éstos.

Mencionemos que justo después de definir la noción de un modelo de un conjunto de oraciones, Tarski dice, de nuevo usando casi las mismas palabras que Tarski (1937), que "precisamente en este sentido habla uno usualmente de modelos de un sistema de axiomas de una teoría deductiva" (Tarski (1936), p. 417). Este "sentido usual" de 'modelo' es el sentido usual según el cual uno especifica un modelo para una teoría cuando uno especifica tanto *un dominio* como ciertos objetos y relaciones *en el dominio* como los significados de los términos primitivos de la teoría (y que hacen verdaderos sus axiomas), y es también, como acabamos de ver, la noción de modelo introducida en Tarski (1937).

Pasemos a las otras dos nociones de que habla Hodges. En mi opinión, es bastante claro que se hallan presentes en la obra temprana de Tarski. La segunda noción de Hodges era "la distinción entre variables y constantes no-lógicas". Estrictamente hablando, la distinción entre variables del lenguaje original y constantes no-lógicas es obvia e indispensable para la definición de la noción de secuencia de objetos modelo de una oración, como es palpable en la definición de más arriba. Pero lo que Hodges quiere decir es que la naturaleza de las constantes no-lógicas de un lenguaje formal de primer orden es tal que son susceptibles, "directamente" por así decir, de recibir interpretaciones distintas en distintas estructuras, no por ser variables de esos lenguajes o de otros ligeramente distintos, sino por ser constantes de un cierto tipo, no interpretadas. Sin embargo, parecería que el procedimiento de Tarski las condena a no ejercer "directamente" su potestad. En la medida en que la objeción es clara, es sólo muy ligeramente diferente de la basada en la tercera noción de Hodges, la de "un símbolo constante no interpretado que recibe una interpretación al ser aplicado a una estructura particular"; esta tercera noción incluye a la de interpretación, mientras que la segunda no. Simplificando, la segunda objeción es que Tarski trata a las constantes como variables, y la tercera es que no concibe que una constante pueda no estar interpretada.

La cuestión es tremendamente sutil, pues a mi manera de ver no se trata de que estas objeciones impliquen que Tarski no definió una noción funcional o extensionalmente equivalente a la de verdad en una estructura común hoy en día, al menos para lenguajes con constantes "interpretadas". Lo que hacen estas objeciones es más bien proponer que Tarski no concibió la posibilidad de *aplicar* su noción funcionalmente equivalente (si es que la tenía, como nosotros hemos defendido que la tenía) *a casos en los que las constantes son no interpretadas* (la tercera objeción) y *sin pasar indirectamente por el proceso de transformar las oraciones del lenguaje original en nuevas funciones proposicionales* (la segunda objeción).

Sin embargo, incluso estas tesis acerca de las intenciones de Tarski y acerca de lo que le parecían aplicaciones concebibles de sus métodos son refutadas por la evidencia histórica. La mera práctica de Tarski de usar constantes no-lógicas en sus artículos sobre sistemas formales algebraicos (cf. por ejemplo, Tarski (1935a), un artículo sobre álgebras de Boole), de los que se dice que reciben diferentes "importantes interpretaciones" (Tarski (1935a), p. 320) indica que la conjetura de Hodges es implausible, pues la idea natural al estudiar estos sistemas es tomarlos como no interpretados y definir la noción de modelo para ellos sin pasar por el proceso de transformación en funciones oracionales. En cualquier caso, encuentro que un pasaje de Tarski (1937a) resuelve definitivamente la cuestión en contra de las conjeturas de Hodges. En este Tarski pasa revista a varias nociones de la lógica moderna, incluida la noción de modelo de un sistema de axiomas.

En primer lugar Tarski introduce la noción de modelo de una manera estrictamente idéntica a la manera como se introduce en Tarski (1936) y Tarski (1937); es decir, comenzando por la sustitución de los términos primitivos (no-lógicos) por variables nuevas y luego esbozando la definición de satisfacción de una función proposicional de este tipo por un "sistema de objetos" (es la palabra que emplea a menudo en Tarski (1937a)) para llegar finalmente a la noción de modelo. Estas consideraciones se aplican, sobre todo, a sistemas de axiomas cuyos términos primitivos tienen una interpretación concreta, la proporcionada por "los objetos designados por los términos primitivos" (Tarski (1937), p. 102). Ahora bien, Tarski es consciente de que

> "ocurre a veces que en el momento mismo en que nos ponemos a construir una teoría deductiva pensamos en muchas interpretaciones diferentes del sistema de axiomas admitido dentro de la teoría, sin querer comprometernos con ninguna de las interpretaciones posibles. Parece lo más conveniente, en este caso, tratar dicha teoría como un *sistema formal*; se considerarán, desde el principio, los términos primitivos de la teoría como símbolos variables, desprovistos de significación concreta, y los axiomas y teoremas serán considerados, no como proposiciones, sino como funciones proposicionales." (Tarski (1937a), pp. 102-103).

Tarski está describiendo casos como los que preocupan especialmente a Hodges, a saber los de teorías axiomáticas algebraicas, con términos primitivos sin ninguna interpretación privilegiada. Así pues, Tarski considera estos casos, en contra de la tercera objeción de Hodges. Pero lo que es más, señala que el estudio de los modelos de los axiomas de estas teorías no tiene que pasar por un proceso de substitución de constantes (los términos primitivos) por nuevas variables (en contra de la primera objeción de Hodges). Al estudiar estos sistemas, nos dice Tarski, lo más *conveniente* es tratarlos *considerando* a *los términos primitivos* como si fueran variables y a los *axiomas y teoremas* como si fueran funciones proposicionales. Nótese que no dice que los términos primitivos *sean* variables o los axiomas *sean* funciones proposicionales. (Tarski no parece comprometerse con ninguna teoría acerca de la naturaleza exacta de los primitivos de las teorías algebraicas.) Tarski dice que se los puede considerar directamente como variables, y evitar en el estudio de sus modelos el paso por la transformación de oraciones en funciones proposicionales[12].

[12] Si las objeciones de Hodges que hemos examinado son tremendamente sutiles, más sutil aún sería una objeción basada en otra tesis de Hodges según la cual la naturaleza de los primitivos no-lógicos de las teorías algebraicas es la de los indéxicos del lenguaje natural. Esta tesis convierte a esos primitivos en "constantes", al menos en cuanto que tienen un significado fijo aunque no tengan una referencia independientemente del contexto de uso. En mi opinión, una objeción a Tarski no puede basarse en el hecho de que no adoptara una tesis de este tipo, que es no sólo disputable, sino prácticamente desconocida por los matemáticos que usan de hecho la noción de verdad en una estructura. Independientemente de esta consi-

En vista de los comentarios precedentes creo que es razonable e incluso obligatorio concluir que, aunque Tarski no menciona en (1933) una noción general de estructura como la habitual hoy en día, no deja de tener en mente una noción idéntica o al menos funcionalmente equivalente a esa noción (especialmente en (1933), p. 199, n. 3); que es consciente de que su método puede aplicarse a la definición de esa noción; que consideró la aplicación de esta noción al caso de oraciones de teorías sin interpretaciones privilegiadas (como las algebraicas); y que incluso consideró el procedimiento "directo" de aplicación común hoy en día en estos casos.

5. Lenguajes de orden infinito

Después de dar definiciones precisas de las principales nociones semánticas para LCC, Tarski ofrece una extensa serie de consideraciones sobre las modificaciones que sería preciso hacer de esas definiciones cuando se quisiera dar definiciones análogas para lenguajes formales más complicados. En la inmensa mayoría de estas consideraciones no nos podemos detener aquí. Pero es importante para nuestros propósitos explicar de modo simplificado algunas de ellas, que tienen que ver con la posibilidad de aplicar el método de Tarski a lo que él llama "lenguajes de orden infinito".

Primero expliquemos de manera general qué es el orden de un lenguaje según Tarski. En primer lugar, las variables individuales (que recorren un conjunto de individuos) y los

deración fundamental, hay en mi opinión serias razones para dudar de la tesis de Hodges. Una de estas razones es que no parece que los matemáticos usen (o al menos, usen siempre) los primitivos de las teorías algebraicas como indéxicos, sino como variables, y en concreto como variables con un dominio de variación *genérico* proporcionado por la teoría algebraica. Por ejemplo, en el curso de una investigación sobre álgebras de Boole, '≤' se usa como una variable que recorre los órdenes parciales de los retículos correspondientes. No es una variable para relaciones diádicas arbitrarias en modelos cualesquiera. Este hecho puede haber llevado a Hodges a pensar que no es una variable sino un indéxico.

nombres de individuos tienen orden 1. Las variables, predicados y functores que (si los toman) toman como argumentos signos de a lo sumo orden 1, y tales que al menos uno de sus argumentos es de orden 1, son de orden 2 (por ejemplo, las variables de LCC son de orden 2). Las variables, predicados y functores que (si los toman) toman argumentos de a lo sumo orden 2, con al menos un argumento de orden 2, son de orden 3 (por ejemplo, el signo 'I' de LCC es de orden 3; un predicado binario '∈' que tome como primer argumento un signo de orden 1 y como segundo argumento un signo de orden 2 es de orden 3). Etc. Para estos casos el orden de un lenguaje L es el mayor número entero positivo n tal que L tiene variables de orden n, si lo hay. Si no hay un mayor n, entonces L es de orden infinito. Así, como ya adelantamos, LCC es un lenguaje de segundo orden. Pronto veremos un lenguaje de orden infinito, no representado por un entero.

Tarski explica cómo su método para dar definiciones de nociones semánticas puede aplicarse a los lenguajes de orden finito en general. Es en estas explicaciones donde no nos podemos detener aquí. Sin embargo, en (1933), aunque no en el *Nachwort* de (1935), Tarski niega que su método pueda aplicarse a lenguajes de orden infinito. Es importante para nuestros propósitos explicar qué quiere decir con esto, y por qué en (1935) Tarski pasa a afirmar que su método puede aplicarse igualmente a lenguajes de orden infinito.

Primero veamos un ejemplo de lenguaje de orden infinito que presenta Tarski, y que será especialmente importante para nosotros. A este lenguaje lo llama 'el lenguaje de la teoría general de clases' (abreviado 'LTGC'). De nuevo lo presentaremos adaptando la notación polaca a la nuestra. Los signos primitivos de LTGC son el cuantificador universal ('∀'), el signo de disyunción ('v'), el signo de negación ('¬'), paréntesis ('(', ')'), la letra 'X', un acento subíndice ('ˌ') y un acento superíndice ('ˈ') para generar $\omega\times\omega$ variables por posposición a 'X' de n acentos subíndices y m superíndices (para abreviar, nos referiremos a la variable en que la letra 'X' va seguida de n acentos subíndices y *m* acentos superíndices como 'X_n^{m}'). El recorrido de las variables en la interpretación deseada de LTGC es intuitivamente como sigue. Una variable de la forma X_n^1 toma como valores individuos, una variable de la forma X_n^2 toma como valores clases de individuos, una variable de la forma X_n^3 toma como valores clases de clases de individuos, etc. Las interpretaciones de las constantes lógicas son las normales.

Las fórmulas atómicas de LTGC son de la forma $X_n^{m+1} X_p^m$. Las fórmulas complejas se obtienen por las operaciones de negación, disyunción y cuantificación universal con respecto a las variables de todos los órdenes. De nuevo Tarski da un cálculo deductivo para LTGC, que consta de axiomas y reglas para las conectivas y los cuantificadores, de axiomas de comprensión y extensionalidad para cada orden y de un axioma de infinitud, que dice que hay un número infinito de individuos. Tarski observa que LTGC basta para desarrollar la matemática que se puede desarrollar con la teoría simple de los tipos finitos, a pesar de contar con menos tipos de variables que esta última. Los "trucos" necesarios para mostrar que LTGC tiene la misma potencia matemática que la teoría simple de los tipos finitos son conocidos: tratar las relaciones como clases de tuplas, construidas a su vez como clases usando el método de Kuratowski; y homogeneizar las relaciones no-homogéneas de alguna de las formas posibles, por ejemplo, elevando el orden de los argumentos pertinentes por aplicación de la operación de formación de unitarios.

¿Cuál es el problema que presenta LTGC (y otros lenguajes de su clase)? ¿Por qué dice Tarski (en (1933)) que no se le puede aplicar su método para construir definiciones de las nociones semánticas? La razón es relativamente simple. Tarski, en (1933), adopta como aparato lógico-matemático de la metateoría (de los lenguajes para los que quiere definir las nociones semánticas) la teoría simple de los tipos finitos. En esta teoría Tarski tiene a su disposición todo el aparato necesario para dar, por ejemplo, su definición de satisfacción para LCC. Esta es una relación entre fórmulas (que se pueden

tomar como individuos, o como secuencias finitas de individuos) y secuencias de clases de individuos (estas secuencias son objetos de un tipo ligeramente superior al de los números naturales (que son clases de clases de individuos) y al de las clases de individuos). Por tanto, la relación de satisfacción para LCC es una relación que se halla en la jerarquía de tipos finitos. Y la forma en la que se define sólo hace uso de recursos de la teoría de tipos finitos. En particular, hace uso de la cuantificación sobre secuencias de clases (en la cláusula para la satisfacción de cuantificaciones universales) y sobre relaciones entre secuencias de clases y fórmulas (en la definición explícita de satisfacción) que son objetos de la jerarquía de los tipos finitos. La construcción tarskiana puede llevarse enteramente a cabo dentro de la teoría de tipos finitos.

Pero la construcción de definiciones tarskianas para LTGC no podría llevarse a cabo dentro de la teoría de los tipos finitos. La relación de satisfacción para LTGC es una relación entre fórmulas (individuos) y secuencias de individuos, secuencias de clases de individuos, secuencias de clases de clases de individuos, etc. O, en cualquier caso, una relación entre fórmulas y secuencias formadas por: una secuencia de individuos, una secuencia de clases de individuos, etc. Es una relación con términos formados por objetos de tipos finitos arbitrariamente grandes. Y no hay ninguna relación en la jerarquía de tipos finitos que esté constituida por objetos de tipos arbitrariamente grandes. Por tanto, la relación de satisfacción para LTGC no es una relación de la jerarquía de tipos finitos. Menos aún se puede cuantificar en la teoría de tipos finitos sobre relaciones de la misma clase entre fórmulas y secuencias formadas por... No es posible cuantificar en la teoría de tipos finitos sobre todas las secuencias necesarias para aplicar el método tarskiano. Ni siquiera una secuencia de secuencias apropiadas (una secuencia de individuos, una secuencia de clases de individuos, etc.) es un objeto de la jerarquía de tipos finitos[13].

[13] Estas someras observaciones no excluyen a priori la posibilidad de identificar la relación de satisfacción con alguna relación funcionalmen-

Este tipo de consideraciones acerca de su propio método llevan a Tarski a preguntarse:

> "si nuestro fracaso es accidental y está conectado de alguna manera con defectos en los métodos usados, o si hay obstáculos de un carácter fundamental que tienen la culpa y que están conectados con la naturaleza de los conceptos que queremos definir, o con la de aquellos con cuya ayuda hemos intentado construir las definiciones buscadas." (Tarski (1933), p. 246).

¿Cómo responder a esta pregunta? Dándole una forma precisa e intentando resolver la cuestión matemáticamente. La forma precisa que Tarski le da es la siguiente:

> "es cuestión de *si sobre la base del lenguaje que estamos considerando* [LTGC] *es en principio posible construir una definición correcta de verdad en el sentido de la convención T.* Como veremos, el problema en esta forma puede ser resuelto de manera definitiva, pero en un *sentido negativo.*" (Tarski (1933), p. 246).

Esa respuesta negativa la proporciona la versión del teorema de la indefinibilidad de la verdad que Tarski demuestra a continuación, y que nosotros examinaremos en la segunda parte de este trabajo.

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS, UNAM
México, DF 04510, México
mariogt@servidor.unam.mx

te equivalente de la jerarquía de tipos finitos (como cuando homogeneizamos relaciones no-homogéneas, o usamos trucos similares). Tarski examina trucos parecidos que usa para indicar cómo su método se aplicaría a lenguajes de orden finito pero con variables de infinitos tipos. Pero es claro que esos trucos no pueden usarse para lenguajes de orden infinito. La idea de que ningún truco tal existe quedará justificada por medio de la versión del teorema de la indefinibilidad de la verdad ofrecida por Tarski y que comentaremos en las siguientes secciones.

BIBLIOGRAFIA

Church, A. (1956), *Introduction to Mathematical Logic I.* Princeton, Princeton University Press.

Coffa, J. A. (1991), *The Semantic Tradition from Kant to Carnap.* Cambridge, Cambridge University Press.

Corcoran, J. (1972), "Conceptual Structure of Classical Logic". *Philosophy and Phenomenological Research*, vol. 33, pp. 25-47.

Etchemendy, J. (1988), "Tarski on Truth and Logical Consequence". *Journal of Symbolic Logic*, vol. 53, pp. 51-79.

Feferman, S. (1988), "Kurt Gödel: Conviction and Caution", en S. Shanker (comp.), *Gödel's Theorem in Focus.* Nueva York. Croom Helm, pp. 96-114.

Gödel, K. (1930), "Die Vollständigkeit der Axiome des logischen Funktionenkalküls", *Monatshefte für Mathematik und Physik*, vol. 37, pp. 349-360; también en Gödel (1986), pp. 102-122. Traducido al inglés como "The Completeness of the Axioms of the Functional Calculus of Logic", en Gödel (1986), pp. 103-123.

—— (1931), "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik*, vol. 38, pp. 173-198; también en Gödel (1986), pp. 144-194. Traducido al inglés como "On Formally Undecidable Propositions of Principia Mathematica and Related Systems I", en Gödel (1986), pp. 145-195. (Referencias a la reimpresión y a la traducción).

—— (1986), *Collected Works*, vol. I. Nueva York, Oxford University Press.

Gómez Torrente, M. (1996), "Tarski on Logical Consequence", *Notre Dame Journal of Formal Logic*, vol. 37, pp. 125-151.

Hodges, W. (1985/6), "Truth in a Structure", *Proceedings of the Aristotelian Society*, n.s., vol. 86, pp. 135-151.

Milne, P. (1999), "Tarski, Truth and Model Theory", *Proceedings of the Aristotelian Society*, n.s., vol. 99, pp. 141-167.

Quine, W. V. (1951), *Mathematical Logic*, edición revisada, Cambridge (Massachussets), Harvard University Press.

Ramsey, F. P. (1925), "The Foundations of Mathematics", *Proceedings of the London Mathematical Society*, 2a serie, vol. 25, pp. 338-384. Reimpreso en F. P. Ramsey. *Philosophical Papers.* Cambridge, Cambridge University Press, 1991, pp. 164-224. (Referencias a esta reimpresión).

Ray, G. (1996), "Logical Consequence: a Defense of Tarski". *Journal of Philosophical Logic*, vol. 25, pp. 617-677.
Sher, G. (1991), *The Bounds of Logic*, Cambridge (Massachusetts). M.I.T. Press.
Soames, S. (1999), *Understanding Truth*. Nueva York. Oxford University Press.
Tarski, A. (1931), "Sur les ensembles définissables de nombres réels. I.". *Fundamenta Mathematicae*, vol. 17, pp. 210-239. Traducido como "On Definable Sets of Real Numbers", en Tarski (1956), pp. 110-142 (referencias a la traducción).
—— (1932), "Mitteilung: Der Wahrheitsbegriff in den Sprachen der deduktiven Disciplinen", *Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse, Akademischer Anzeiger*, vol. 39, pp. 23-25.
—— (1933), *Pojecie prawdy w jezykach nauk dedukcyjnych* (Sobre el concepto de verdad en lenguajes de las ciencias deductivas), *Travaux de la société des sciences et des lettres de Varsovie, Classe III*, no. 34. Varsovia. Traducido al inglés como parte de "The Concept of Truth in Formalized Languages", en Tarski (1956), pp. 152-267. (Referencias a la traducción inglesa.)
—— (1935), "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica*, vol. 1, pp. 261-405. Una traducción al alemán de Tarski (1933), con un nuevo "Nachwort" añadido. Traducido al inglés como "The Concept of Truth in Formalized Languages", en Tarski (1956), pp. 152-278. (Referencias a la traducción inglesa.)
—— (1935a), "Zur Grundlegung der Booleschen Algebra. I", *Fundamenta Mathematicae* vol. 24, pp. 177-198. Traducido al inglés como "On the Foundations of Boolean Algebra", en Tarski (1956), pp. 320-341. (Referencias a la traducción inglesa.)
—— (1936), "Über den Begriff der logischen Folgerung", en *Actes du Congrès International de Philosophie Scientifique*, fasc. 7 (Actualités Scientifiques et Industrielles, vol. 394), París, Hermann et Cie, pp. 1-11. Traducido al inglés como "On the Concept of Logical Consequence", en Tarski (1956), pp. 409-420. (Referencias a la traducción inglesa.)
—— (1937), *Einführung in die mathematische Logik und in die Methodologie der Mathematik*, Viena. Julius Springer.
—— (1937a), "Sur la méthode déductive", en *Travaux du IXe Congrès International de Philosophie*, tomo 6 (Actualités Scientifiques et Industrielles, vol. 535), París, Hermann et Cie, pp. 95-103.

—— (1944), "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research*, vol. 4, pp. 341-376.
—— (1952), "Some Notions and Methods on the Borderline of Algebra and Metamathematics", en *Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950, vol. I*, Providence (Rhode Island), American Mathematical Society, pp. 705-720.
—— (1954), "Contributions to the Theory of Models I, II", en *Indagationes Mathematicae*, vol. 16, pp. 572-588.
—— (1956), *Logic, Semantics, Metamathematics*, Oxford, Oxford University Press, 2a ed.: Indianapolis, Hackett, 1983.
—— y R. L. Vaught (1957), "Arithmetical Extensions of Relational Systems", *Compositio Mathematica*, vol. 13, pp. 81-102.
Vaught, R. L. (1974), "Model Theory before 1945", en L. Henkin et al. (comps.), *Proceedings of the Tarski Symposium*, Proceedings of Symposia in Pure Mathematics, vol. XXV. Providence (Rhode Island), American Mathematical Society, pp. 153-172.
—— (1986), "Tarski's Work in Model Theory", *Journal of Symbolic Logic*, vol. 51, pp. 869-882.

ABSTRACT

This is the first part of a two-part devoted to a study of some historical, logical and philosophical issues arising from a reading of Tarski's celebrated "Wahrheitsbegriff" monograph. Thi first part starts with an exposition of Tarski's "Wahrheitsbegriff" theory of truth which is historically accurate but more easily intelligible to the modern reader than Tarski's "Wahrheitsbegriff" presentation. This exposition is then used in the refutation of some historical, logical and philosophical ideas appearing in the literature on Tarski, and in particular of the idea that he did not define or even considered in his work of this period the notion of "truth in a structure". The second part of the paper will concentrate on issues related to Tarski's "Wahrheitsbegriff" version of the theorem on the indefinability of truth.