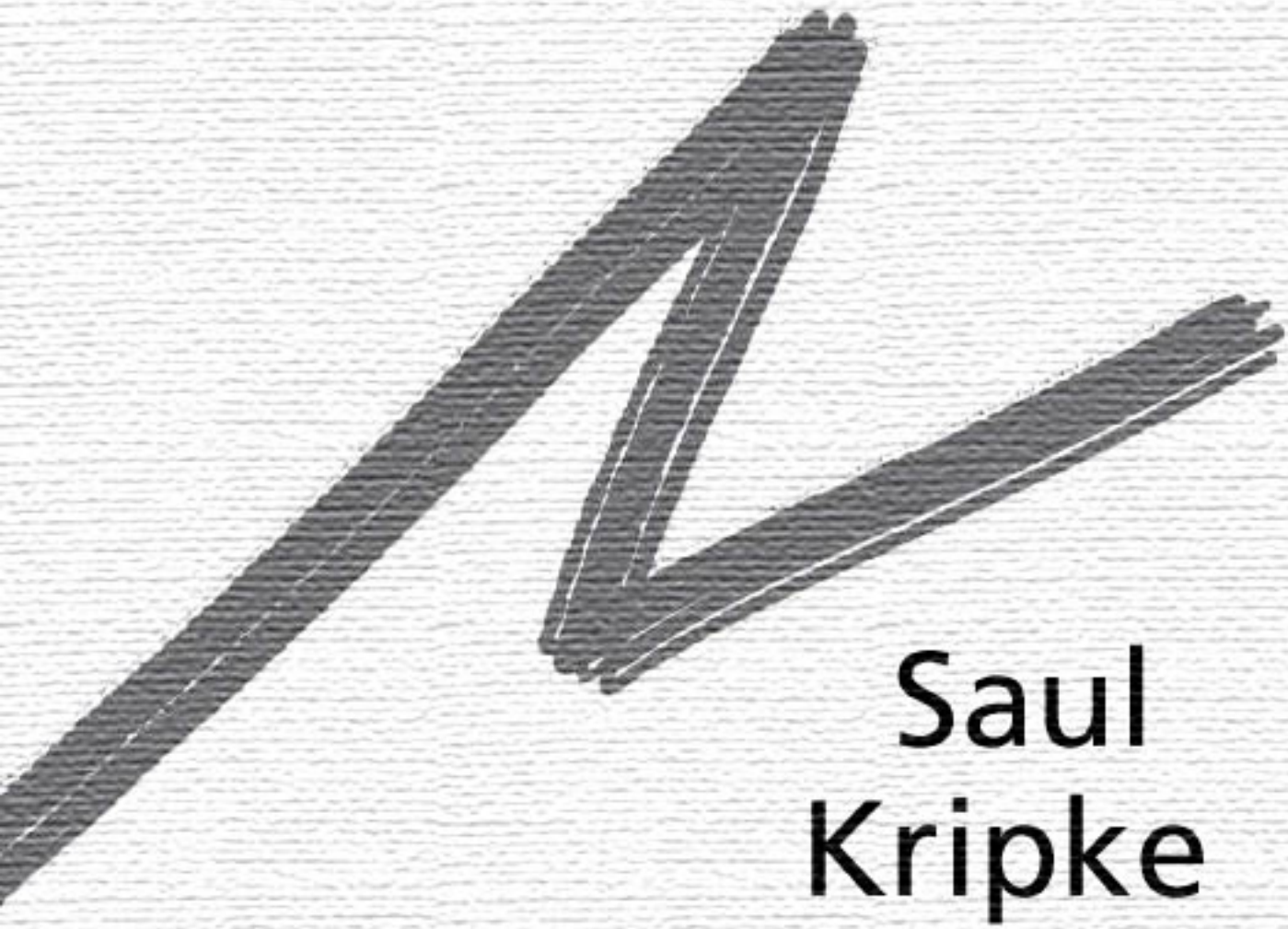
Esbozo de una teoría de la verdad



Saul Kripke

Esbozo de una teoría de la verdad

Saul Kripke

Traducido por M. M. Valdés
En Juan Antonio Nicolás y María José Frápoli (ed.), *Teorías de la verdad en el siglo XX*. Tecnos, Madrid, 1997

Edición original:
"Outline of a Theory of Truth", *Journal of Philosophy*, 72/19 (1975), pp. 690-715
Reeditado en R. L. Martin (ed.), *Truth and de Liar Paradox*, Clarendom Press, Oxford, 1984, pp. 53-81

Edición castellana:
Esbozo de una teoría de la verdad,
UNAM, México, 1984

Se han sustituido algunos términos de la traducción utilizada, para adaptarla a la nomenclatura comúnmente aceptada.

Los números entre corchetes corresponden a la paginación de la edición impresa.

letra e

[110]

Esbozo de una teoría de la verdad[1]

I. EL PROBLEMA

Desde que Pilatos preguntó: «¿Qué es la verdad?» (*San Juan*, XVIII, 38) la búsqueda subsecuente de una respuesta correcta se ha visto inhibida por otro problema que, como es bien sabido, surge también en el contexto del Nuevo Testamento. Si, como supone el autor de la Epístola a Tito (*Tito* 1, 12), un profeta cretense, «incluso un

[1] Presentado en el Simposio sobre la Verdad organizado por la *American Philosophical Association*, diciembre 28 de 1975.

Originalmente habíamos acordado que presentaría este trabajo oralmente sin entregar previamente un texto preparado. En una fecha relativamente tardía, los editores del *Journal of Philosophy* me pidieron que entregara por lo menos los «lineamientos generales» de mi trabajo por escrito. Estuve de acuerdo en que esto sería de utilidad. Recibí la solicitud cuando ya había aceptado otro compromiso y tuve que preparar la presente versión a toda prisa sin tener siquiera la oportunidad de revisar el primer borrador. Si hubiera tenido la oportunidad de hacer una revisión habría ampliado la presentación del modelo básico en la sección III con el fin de hacerlo más claro. El texto muestra que una buena parte del material formal y filosófico, así como las pruebas de los resultados, tuvieron que omitirse.

Breves resúmenes del presente trabajo se presentaron en la reunión de primavera de 1975 de la *Association for Symbolic Logic* que tuvo lugar en Chicago. Una versión más amplia se presentó en forma de tres conferencias en la Universidad de Princeton en junio de 1975. Espero publicar una versión más detallada en algún otro lugar. Dicha versión más amplia debería contener algunos planteamientos técnicos hechos aquí sin suministrar la prueba y una buena cantidad de material técnico y filosófico no mencionado o resumido en este esbozo.

profeta de ellos mismos», afirma que «los cretenses son siempre mentirosos» y si «este testimonio es verdadero» con respecto a todas las demás proferencias cretenses, parece entonces que las palabras del profeta cretense son verdaderas si y sólo si son falsas. Cualquier tratamiento del concepto de verdad tiene que evitar esta paradoja.

El ejemplo cretense ilustra una manera de lograr la autorreferencia. Sean P(x) y Q(x) predicados de oraciones. Entonces, en algunos casos, las pruebas empíricas establecen que la oración «(x)(P(x) $\supset$ Q(x))» [o «($\exists$x)(P(x) $\wedge$ Q(x))» u otras similares] satisface ella misma el predicado P(x); algunas veces las pruebas empíricas muestran que dicha oración es el único objeto que satisface P(x). En este último caso, la oración en cuestión «dice de sí misma» que satisface Q(x). Si Q(x) es el predicado[2] «es falso», el resultado es la paradoja del [111] Mentiroso.

[2] Sigo la convención usual de la teoría «semántica» de la verdad al considerar que la verdad y la falsedad son predicados que son verdaderos de las oraciones. Si los predicados de verdad y falsedad se aplican en primer lugar a las proposiciones o a otras entidades no lingüísticas, interprétese el predicado aplicado a oraciones como «expresa una verdad».

He elegido considerar a las oraciones como los vehículos primarios de la verdad no porque piense que la objeción que dice que la verdad es primariamente una propiedad de las proposiciones (o de los «enunciados») no es pertinente para el trabajo serio sobre la verdad o para las paradojas semánticas. Por el contrario, creo que en último término un tratamiento cuidadoso del problema bien puede hacer necesaria la separación entre el aspecto «expresa» (que relaciona las oraciones con las proposiciones) y el aspecto «verdad» (que putativamente se aplica a las proposiciones). No he investigado si las paradojas semánticas presentan problemas cuando se aplican directamente a las proposiciones. La razón principal por la que aplico el predicado verdad directamente a los objetos lingüísticos, es porque se ha desarrollado una teoría matemática de la autorreferencia para tales objetos. (Véase también la nota 32.)

A manera de ejemplo, digamos que P(x) abrevia el predicado «tiene instancias impresas en los ejemplares de *Teorías de la Verdad en el siglo XX*, artículo 5, sección I, párrafo 2.°». Entonces, la oración

$$(x)P(x) \supset Q(x)$$

conduce a la paradoja si interpretamos Q(x) como la falsedad.

Las versiones de la paradoja del Mentiroso que usan predicados empíricos señalan ya un aspecto importante del problema: *muchas de nuestras afirmaciones ordinarias sobre la verdad y la falsedad, probablemente la mayoría de ellas, son susceptibles de exhibir rasgos paradójicos cuando los hechos empíricos son extremadamente desfavorables.* Considérese el enunciado ordinario hecho por Juan:

(1) La mayor parte (es decir, una mayoría) de las afirmaciones de Nixon acerca de Watergate son falsas.

Además, una versión más desarrollada de la teoría admitiría a aquellos lenguajes que contienen demostrativos y ambigüedades y hablaría de las proferencias, las oraciones bajo una interpretación, y cosas similares, como aquello que tiene un valor de verdad. En la exposición informal este artículo no pretende ser preciso con respecto a estos asuntos. Las oraciones son los vehículos oficiales de la verdad pero informalmente hablaremos en ocasiones de las proferencias, los enunciados, las afirmaciones y otras cosas. Podemos hablar ocasionalmente como si cada una de las proferencias de una oración en un lenguaje constituyera un enunciado, aunque sugiramos más adelante que una oración puede no ser enunciado en el caso de ser paradójica o infundada. Trataremos de ser precisos sobre estos asuntos sólo cuando consideremos que la imprecisión puede dar lugar a confusión o malentendidos. Observaciones similares se aplican a las convenciones sobre el uso de comillas.

Evidentemente no hay nada intrínsecamente incorrecto con respecto a (1), tampoco es un enunciado mal formado. Comúnmente el valor de verdad de (1) podrá evaluarse mediante una enumeración de [112] las afirmaciones de Nixon relacionadas con Watergate y una evaluación de cada una de ellas con respecto a la verdad o la falsedad. Sin embargo, supongamos que las afirmaciones de Nixon sobre Watergate se encuentran repartidas por parejo entre la verdad y la falsedad, excepto por un caso problemático:

(2) Todo lo que dice Juan sobre Watergate es verdadero.

Supongamos, además, que (1) es la única afirmación que hace Juan sobre Watergate o, alternativamente, que todas sus afirmaciones relacionadas con Watergate son verdaderas excepto, tal vez, (1). No se requiere demasiada habilidad entonces para mostrar que tanto (1) como (2) son paradójicas: son verdaderas si y sólo si son falsas.

El ejemplo de (1) pone de relieve una lección importante: sería una tarea estéril buscar un criterio intrínseco que nos permitiera cribar —por carecer de significado o estar mal formadas— aquellas oraciones que conducen a paradojas. Ciertamente (1) es el paradigma de una afirmación común que contiene la noción de falsedad; justamente este tipo de afirmaciones caracterizaron nuestro reciente debate político. Sin embargo, ningún rasgo sintáctico o semántico de (1) garantiza que no sea paradójica. Bajo los supuestos del párrafo anterior (1) conduce a una paradoja[3]. Que se den o no dichos supuestos depende de los

[3] Tanto Nixon como Juan pueden haber hecho sus proferencias respectivas sin darse cuenta de que los hechos empíricos los hacen paradójicos.

hechos empíricos sobre las afirmaciones de Nixon (y del otro) y no de algo intrínseco a la sintaxis y a la semántica de (1). (Aun los expertos más sutiles pueden ser incapaces de evitar proferencias que conducen a paradojas. Se cuenta que Russell preguntó en una ocasión a Moore si siempre decía la verdad y que consideró la respuesta negativa de Moore como la única falsedad emitida por Moore. No hay duda de que nadie ha tenido un olfato más fino para las paradojas que Russell. Sin embargo, es obvio que no se percató de que si, como él pensaba, todas las otras proferencias de Moore eran verdaderas, la respuesta negativa de Moore no sólo era falsa, sino paradójica[4].) La moraleja: una teoría adecuada debe per-[113]mitir que sean riesgosos nuestros enunciados que contienen la noción de verdad; corren el riesgo de ser paradójicos si los hechos empíricos son extremadamente (e inesperadamente) desfavorables. No puede haber ninguna «criba» sintáctica o semántica que deseche los casos «malos» y conserve los casos «buenos».

En lo anterior me he concentrado en versiones de la paradoja que usan propiedades empíricas de las oraciones, tales como el ser proferidas por ciertas personas particulares. Gödel mostró esencialmente que dichas propiedades son dispensables en favor de propiedades puramente sintácticas: mostró que, para todo predicado Q(x), podía producirse un predicado sintáctico P(x) tal que la oración (x)(P(x) ⊃ Q(x)) es el único objeto que satisface P(x) y que esto es demostrable. Así, en un sentido, (x)(P(x) ⊃ Q(x)) «dice de sí misma» que satisface Q(x). Tam-

[4] Conforme a la manera ordinaria de entender esto (en tanto que opuesta a las convenciones de quienes enuncian paradojas del tipo del Mentiroso) el problema radica en la sinceridad de las proferencias de Moore y no en su verdad. Probablemente también podrían derivarse las paradojas bajo esta interpretación.

bién demostró que la sintaxis elemental puede interpretarse en la teoría del número. De esta manera, Gödel puso fuera de toda duda el asunto de la legitimidad de las oraciones autorreferenciales; demostró que son tan irreprochablemente legítimas como la aritmética misma. Pero los ejemplos que usan predicados empíricos preservan su importancia: ponen de relieve la moraleja acerca del carácter riesgoso al que apunté antes.

Una forma más simple, y más directa, de autorreferencia usa los demostrativos o los nombres propios: Sea «Jack» un nombre de la oración «Jack es breve» y tenemos una oración que dice de sí misma que es breve. No veo que haya nada incorrecto en la autorreferencia «directa» de este tipo. Si «Jack» no había sido introducido previamente como un nombre en el lenguaje[5], ¿por qué no hemos de poderlo introducir como un nombre de cualquier entidad que nos plazca? En particular, ¿por qué no puede ser el nombre de la secuencia finita (no interpretada) de signos «Jack es breve»? (¿Se permitiría llamar a esta secuencia de signos «Harry», pero no «Jack»? Sin duda alguna las prohibiciones acerca de dar nombres son arbitrarias en este caso.) No hay ningún círculo vicioso en esta manera de proceder, ya que no tenemos que interpretar la secuencia de signos «Jack es breve» antes de nombrarla. No obstante, si le damos el nombre «Jack», de inmediato se convierte en significativa y verdadera. (Nótese que estoy hablando de oraciones autorreferenciales, no de proposiciones autorreferenciales[6].) [114]

[5] Asumimos que «es breve» está ya en el lenguaje.

[6] No es obviamente posible aplicar esta técnica para obtener proposiciones «directamente» autorreferenciales.

En una versión más extensa, apuntalaría la conclusión anterior no sólo mediante una formulación filosófica más detallada, sino también mediante una demostración matemática de que la clase sencilla de autorreferencia ejemplificada mediante el caso de «Jack es breve» podría de hecho usarse para probar el teorema mismo de incompletud de Gödel (y también el teorema de Gödel y Tarski sobre la indefinibilidad de la verdad). Tal presentación de la prueba del teorema de Gödel podría ser más perspicua para el principiante que la prueba usual. También despeja la impresión de que Gödel estaba forzado a reemplazar la autorreferencia directa por otro artificio más circunlocutorio. Tengo que omitir el argumento en este esbozo[7].

Desde hace mucho tiempo se ha reconocido que parte del problema intuitivo que tenemos con oraciones del tipo del Mentiroso también se encuentra en oraciones como:

(3) (3) es verdadera

las cuales, aunque no son paradójicas, tampoco dan lugar a condiciones de verdad determinadas. Entre los ejemplos más complicados se encuentran, por ejemplo, el de un par de oraciones cada una de las cuales dice de la otra que es verdadera y el de una secuencia infinita de oraciones P_i en donde P_i dice que P_{i+1} es verdadera. En general, si una oración como (1) afirma que (todas, la mayoría de, algunas de, etcétera) las oraciones de cierta clase C son verdaderas, su valor de verdad

[7] Hay varias maneras de hacer esto, usando una numeración de Gödel no estándar en la que los enunciados pueden contener numerales que designan sus propios números de Gödel, o usando una numeración de Gödel estándar añadiendo además constantes del tipo de «Jack».

puede evaluarse si el valor de verdad de las oraciones de la clase C puede evaluarse. Si algunas de estas oraciones contienen la noción de verdad, su valor de verdad debe a su vez evaluarse considerando otras oraciones y así sucesivamente. Si este proceso finaliza en último término en oraciones que no contienen el concepto de verdad, de manera que el valor de verdad del enunciado original puede establecerse, decimos que la oración original es fundada [*grounded*], de otra manera será *infundada* [*ungrounded*][8]. Como lo indica el ejemplo (1), el que una oración sea, o no, fundada, no es en general [115] una propiedad intrínseca (sintáctica o semántica) de la oración, sino que generalmente depende de los hechos empíricos. Hacemos proferencias con la esperanza de que resulten fundadas. Las oraciones como (3), aunque no son paradójicas, son infundadas. Lo anterior es un tosco bosquejo de la noción común de fundamentación y no pretende suministrar una definición formal: el hecho de que pueda suministrar una definición formal será una de las virtudes principales de la teoría formal sugerida en lo que sigue[9].

[8] Si una oración afirma, por ejemplo, que todas las oraciones de la clase C son verdaderas, dejaremos que sea falsa y fundada si hay una oración en C que sea falsa, sin importamos si son fundadas las otras oraciones en C.

[9] La fundamentación *[groundedness]* parece haber sido explícitamente introducida, con ese nombre, en la literatura filosófica en el articulo de Hans Hertzberger, «Paradoxes of Grounding in Semantics», *The Journal of Philosophy*, XVII, 6, marzo 26 de 1970, pp. 145–167. El artículo de Hertzberger se basa en un trabajo no publicado sobre un enfoque de las paradojas semánticas desde el punto de vista de la «fundamentación» *[«groundedness» approach]* elaborado conjuntamente con Jerrold J. Katz. En semántica, la noción intuitiva de «estar fundado» formaba parte del folklore del asunto ciertamente desde mucho antes. Hasta donde yo sé, el presente trabajo proporciona la primera definición rigurosa.

II. PROPUESTAS ANTERIORES

Hasta el momento, el único enfoque de las paradojas semánticas que se ha elaborado con algún detalle, es el que llamaré «el enfoque ortodoxo» que conduce a la célebre jerarquía de lenguajes de Tarski[10]. Sea L_0 un lenguaje formal construido mediante las operaciones comunes del cálculo de predicados de primer orden a partir de un elenco de predicados primitivos (completamente definidos) y ade-[116]cuado para discutir su propia sintaxis (usando tal vez la aritmetización). (Omito una caracterización exacta.) Un lenguaje así, no puede contener su propio predicado de verdad (en realidad, de satisfacción) $T_1(x)$ para L_0. (De hecho, Tarski muestra cómo definir dicho predicado en un

[10] Entiendo por «enfoque ortodoxo» cualquier enfoque que trabaje dentro de la teoría de la cuantificación clásica y exija que todos los predicados sean totalmente definidos sobre el recorrido de las variables. Varios escritores hablan como si la «jerarquía de lenguajes», o el enfoque tarskiano, le prohibiera a uno formar, por ejemplo, lenguajes con cierto tipo de autorreferencia, o lenguajes que contienen sus propios predicados de verdad. De acuerdo a mi interpretación, no hay ninguna prohibición; hay solamente teoremas sobre lo que se puede y no se puede hacer dentro del marco de la teoría clásica ordinaria de la cuantificación. Así Gödel demostré que un lenguaje clásico puede hablar de su propia sintaxis; usando definiciones restringidas de la verdad y otros artificios, dicho lenguaje puede decir muchas cosas sobre su propia semántica. Por otro lado, Tarski probó que un lenguaje clásico no puede contener su propio predicado de verdad y que un lenguaje de un orden superior puede definir un predicado de verdad para un lenguaje de orden inferior. Nada de esto surgió a partir de ningunas restricciones *a priori* sobre la autorreferencia distintas de aquellas que se derivan de la restricción para un lenguaje clásico en el que todos los predicados están totalmente definidos.

lenguaje de orden superior.) El proceso puede repetirse, conduciendo a una secuencia L_0, L_1, L_2, L_3,... de lenguajes, cada uno de los cuales con su predicado de verdad para el anterior.

Los filósofos han tenido suspicacias con respecto al enfoque ortodoxo en tanto que análisis de nuestras intuiciones. Sin lugar a dudas nuestro lenguaje contiene una sola palabra «verdad», y no una secuencia de expresiones distintas «verdad$_n$», la cual se aplica a oraciones de niveles más y más altos. Un defensor de la posición ortodoxa puede responder en contra de esta objeción (en el caso de que no mande a volar de una vez por todas al lenguaje natural, como Tarski se inclinaba a hacerlo) que la noción ordinaria de verdad es sistemáticamente ambigua: su «nivel» en una figuración particular se determina por el contexto de la proferencia y por las intenciones del que habla. La noción de predicados de verdad que difieren, cada uno de ellos con su propio nivel, parece corresponder a la idea intuitiva siguiente, implícita en la discusión anterior sobre el «ser fundado»: Primeramente hacemos varias proferencias, tales como «la nieve es blanca», que no contienen la noción de verdad. Luego, les atribuimos a dichas proferencias el predicado «verdadero$_1$». («Verdadero$_1$» significa —toscamente— «es un enunciado verdadero que no contiene en sí mismo la noción de verdad u otras semejantes».) Podemos entonces formar el predicado «verdadero$_2$» que se aplica a oraciones que contienen «verdadero$_1$» y así sucesivamente. Podemos asumir que en cada ocasión de una proferencia, cuando un hablante usa la palabra «verdadero», le agrega un subíndice implícito que va creciendo a medida que, al reflexionar más y más, accede a niveles cada vez más altos en su

propia jerarquía de Tarski[11].[117]

Desafortunadamente esta forma de ver las cosas parece infiel a los hechos. Si alguien hace una proferencia como (1), no agrega un subíndice, ni explícito ni implícito, a su proferencia de «falso» que determine el «nivel de lenguaje» en el que habla. Un subíndice implícito no causaría ningún problema si estuviésemos seguros del «nivel» de las proferencias de *Nixon*; podríamos entonces abarcarlos a todos, en la proferencia de (1) o incluso en la del más fuerte@

(4) Todas las proferencias de Nixon sobre Watergate son falsas,

[11] El artículo de Charles Parsons «The Liar Paradox», *Journal of Philosophical Logic,* III, 4, octubre de 1974, pp. 380–412, puede tomarse tal vez como si proporcionara un argumento similar al que se esboza en este párrafo. Sin embargo puede considerarse que una gran parte de su artículo queda confirmada, y no refutada, por el presente enfoque. Véase en particular su nota 19 en la que expresa su esperanza de que haya una teoría que evite los subíndices explícitos. El punto fijo mínimo (véase la Sección III más adelante) evita los subíndices explícitos, pero tiene, no obstante, una noción de nivel; en este respecto, puede compararse con la teoría estándar de los conjuntos como opuesta a la teoría de los tipos. El hecho de que los niveles no sean intrínsecos a las oraciones, es peculiar a la presente teoría y es algo adicional a la ausencia de la subindicación explícita.

La asignación de niveles intrínsecos ortodoxa garantiza liberarse del «carácter arriesgado» en el sentido explicado anteriormente en la Sección I. Con respecto a (4) y (5) más adelante, la mera asignación de niveles intrínsecos, que eliminaría su carácter riesgoso, también les impediría «buscar sus propios niveles» (véanse pp. 14–15). Si queremos permitir que las oraciones busquen sus propios niveles, parece obvio que también tenemos que permitir oraciones riesgosas. En ese caso, tenemos que considerar que las oraciones tratan de expresar proposiciones y tenemos que permitir vacíos de valores de verdad. Véase la Sección III más adelante.

escogiendo simplemente un subíndice más alto que el de cualquier nivel contenido en los proferencias de Nixon sobre Watergate. Generalmente, sin embargo, un hablante no tiene ninguna manera de conocer los «niveles» de las proferencias relevantes de Nixon. Así, pues, Nixon pudo haber dicho «Dean es un mentiroso» o «Haldman dijo la verdad cuando dijo que Dean mintió», etcétera, y los «niveles» de éstos pueden aun depender de los niveles de las proferencias de Dean y así sucesivamente. Si se obliga al hablante a asignarle de antemano un «nivel» a (4) [o a la palabra «falso» en (4)], puede estar inseguro acerca de qué tan alto haya de ser el nivel; si, por ignorar el «nivel» de las proferencias de Nixon, escoge un nivel demasiado bajo, su proferencia de (4) falla en su propósito. La idea de que un enunciado como (4) debiera tener un «nivel», en sus usos normales, es convincente intuitivamente. Es, sin embargo, igualmente obvio intuitivamente que el «nivel» de (4) no debe de depender solamente de la forma de (4) (como sería el caso si se les asignaran subíndices explícitos a «falso», o tal vez a «proferencias»); el hablante tampoco debe asignarlo por adelantado, sino que más bien su nivel debe depender de los hechos empíricos relativos a lo que Nixon ha proferido. Mientras más altos sean los «niveles» de Nixon, más alto será el «nivel» de (4). Esto significa que, en algún sentido, se debe per-[118]mitir que un enunciado encuentre su propio nivel, lo suficientemente alto como para que diga lo que se propone decir. No debe tener un nivel intrínseco fijado de antemano, como en la jerarquía de Tarski.

Hay otra situación que resulta aún más difícil de acomodar dentro de los confines del enfoque ortodoxo. Supongamos que Dean afirma (4) en tanto que Nixon por su parte afirma:

(5) Todo lo que dice Dean sobre Watergate es falso.

Al afirmar Dean la oración omniabarcante (4) desea incluir en su alcance la afirmación (5) (como una de las afirmaciones de Nixon sobre Watergate de las que dice que son falsas); Nixon, por su parte, al afirmar (5) quiere hacer lo mismo con la afirmación (4) de Dean. Ahora bien, en cualquier teoría que pretenda asignar «niveles» intrínsecos a tales enunciados, de manera que un enunciado de determinado nivel sólo pueda hablar de la verdad o falsedad de los enunciados de niveles inferiores, es claramente imposible que ambas afirmaciones tengan éxito: si los dos enunciados están en el mismo nivel, ninguno de los dos puede hablar sobre la verdad o la falsedad del otro, mientras que si no están en el mismo nivel, el que está en un nivel más alto puede hablar del de nivel inferior, pero no a la inversa. Sin embargo, intuitivamente, podemos con frecuencia asignar valores de verdad no ambiguos a (4) y a (5). Supongamos que Dean hizo al menos un enunciado verdadero sobre Watergate [distinto de (4)]. Entonces, independientemente de cualquier evaluación de (4), podemos decidir que el (5) de Nixon es falso. Si todas las otras afirmaciones de Nixon sobre Watergate también son falsas, la afirmación (4) de Dean es verdadera; si alguna de ellas es verdadera, (4) es falsa. Nótese que en el último caso, podríamos haber juzgado que (4) es falsa sin evaluar (5), en tanto que en el primer caso la evaluación de (4) como verdadera dependía de la evaluación previa de (5) como falsa. Bajo otro conjunto diferente de supuestos empíricos sobre la veracidad de Nixon y Dean, (5) hubiera sido verdadera [y su evaluación como verdadera dependería de una evaluación

previa de (4) como falsa]. Me parece difícil acomodar estas intuiciones dentro de los confines del enfoque ortodoxo.

Algunos otros defectos del enfoque ortodoxo resultan más difíciles de explicar en un esbozo breve, aunque han constituido una parte sustancial de mi investigación. Un problema es el de los niveles transfinitos. Es fácil afirmar dentro de los confines del enfoque ortodoxo: [119]

(6) La nieve es blanca

y afirmar que (6) es verdadera, que «(6) es verdadera» es verdadera, que «"(6) es verdadera" es verdadera» es verdadera, y así sucesivamente; a las distintas figuraciones con la secuencia de «es verdadera» se les asignan subíndices cada vez mayores. Es algo mucho más difícil afirmar que todos los enunciados en la secuencia que acabamos de describir son verdaderos. Para hacer esto, necesitamos un metalenguaje de nivel transfinito, por encima de todos los lenguajes de nivel finito. Para mi sorpresa, he descubierto que el problema de definir los lenguajes de nivel transfinito presenta dificultades técnicas sustanciales que nunca han sido seriamente investigadas[12].

(Hilary Putnam y sus discípulos esencialmente investigaron el problema —descrito de diferente manera y con una motivación matemática en apariencia completamente diferente— para el caso especial en el que empezamos en el nivel más bajo con el lenguaje de la

[12] El problema de los niveles transfinitos tal vez no es tan difícil de resolver de manera canónica en el nivel ω, pero se vuelve cada vez más agudo en los niveles ordinales superiores.

teoría elemental del número.) He obtenido algunos resultados positivos sobre el problema, así como algunos resultados negativos; no puedo detallarlos aquí. Pero dado el estado que presenta actualmente la literatura sobre el tema debería decirse que si la «teoría de los niveles de lenguaje» ha de incluir una explicación de los niveles transfinitos, entonces uno de los principales defectos de la teoría es simplemente su inexistencia. Podemos decir que la literatura existente define «la jerarquía de lenguajes de Tarski» sólo para los niveles finitos, lo cual difícilmente puede considerarse adecuado. Mi propio trabajo incluye una ampliación de la teoría ortodoxa a los niveles transfinitos, pero aún está incompleto. La falta de espacio no sólo me impide describir el trabajo, sino también me impide mencionar las dificultades matemáticas que convierten al problema en algo sumamente no trivial.

Podemos sólo mencionar algunos otros problemas. Fue para mí una sorpresa que el enfoque ortodoxo no garantice en absoluto de manera obvia la fundamentación [*groundednes*] en el sentido intuitivo antes mencionado. El concepto de verdad para los enunciados matemáticos Σ_1 es él mismo Σ_1 y este hecho puede ser usado para [120] construir enunciados de la forma de (3). Aun cuando estén en cuestión las definiciones irrestrictas de verdad, los teoremas estándar nos permiten fácilmente construir una cadena descendente de lenguajes de primer orden L_0, L_1, L_2, ..., tal que L_i contiene un predicado de verdad para L_{i+1}. No sé si dicha cadena pueda engendrar oraciones infundadas, ni siquiera sé bien cómo formular aquí el problema; algunas cuestiones técnicas sustanciales en esta área tienen todavía que resolverse.

Casi toda la literatura reciente que busca alternativas al enfoque ortodoxo —mencionaré especialmente los escritos de Bas van Fraassen y Robert L. Martin—[13] está de acuerdo en una sola idea básica: habrá de haber solamente un predicado de verdad, aplicable a oraciones que contienen el predicado mismo; no obstante, la paradoja ha de evitarse al permitir vacíos de valores de verdad y al declarar que las oraciones paradójicas en particular padecen de semejante vacío. Me parece que estos escritos sufren a veces de un defecto menor y casi siempre de un defecto mayor. El defecto menor es que algunos de ellos critican una versión caricaturizada del enfoque ortodoxo, no el enfoque genuino[14].

[13] Véase Martin (ed.), *The Paradox of the Liar*, New Haven, Yale, 1970, así como las referencias ahí mencionadas.

[14] Véase la nota 9 anterior. Martin, por ejemplo, en su trabajo «Toward a Solution to the Liar Paradox», *Philosophical Review*, LXXXVI, 3, julio de 1967, pp. 279–311 y «On Grelling's Paradox», ibid. LXXVII, 3, julio de 1968, pp. 325–331, atribuye a «la teoría de los niveles de lenguaje» todo tipo de restricciones sobre la autorreferencia las cuales deben considerarse simplemente como refutadas, incluso para los lenguajes clásicos, por el trabajo de Gödel. Quizá hay o haya habido algunos teóricos que creyeran que todo lo que se dice de un lenguaje debe tener lugar en un metalenguaje distinto. Esto importa poco; el asunto principal es: ¿qué construcciones pueden llevarse a cabo dentro de un lenguaje clásico y qué construcciones requieren vacíos de valores de verdad? Casi todos los casos de autorreferencia mencionados por Martin pueden llevarse a cabo por los métodos ortodoxos gödelianos, sin necesidad de invocar predicados parcialmente definidos ni vacíos de valores de verdad. En la nota 5 de su segundo articulo, Martin se percata de la demostración de Gödel de que los lenguajes suficientemente ricos contienen su propia sintaxis, pero parece no darse cuenta de que ese trabajo convierte en irrelevante la mayor parte de su polémica contra los «niveles de lenguaje».

En el otro extremo, algunos autores aún parecen pensar que es útil para el tratamiento de las paradojas semánticas algún tipo de prohibición general sobre la autorreferencia. En el caso de las oraciones autorreferenciales me parece que ésta es una posición sin esperanzas.

El defecto mayor es que casi invariablemente estos escritos son meras sugerencias y no teo-[121]rías genuinas. Casi nunca hay una formulación semántica precisa de un lenguaje que sea por lo menos lo suficientemente rico como para hablar de su propia sintaxis elemental (ya sea directamente o mediante la aritmetización) y contener su propio predicado de verdad. Sólo en el caso en que dicho lenguaje fuese formulado con precisión formal podría decirse que se ha presentado una teoría de las paradojas semánticas. Idealmente, una teoría debería mostrar que la técnica puede aplicarse a lenguajes arbitrariamente ricos sin importar cuáles sean sus otros predicados «ordinarios» distintos a la verdad. Hay un sentido más en el que el enfoque ortodoxo suministra una teoría, en tanto que la literatura reciente sobre el tema no lo hace. Tarski muestra cómo puede proporcionar una definición matemática de verdad —para un lenguaje clásico de primer orden cuyos cuantificadores tienen como recorrido un conjunto— usando los predicados del lenguaje objeto además de la teoría de los conjuntos (lógica de orden superior). La literatura alternativa abandona el objetivo de dar una definición matemática de verdad y se contenta con tomar la verdad como un primitivo intuitivo. Un solo artículo que he leído dentro del género «vacíos de verdad» —un trabajo reciente de Martin y Peter Woodruff—[15] podría considerarse como un inicio de intento de satisfacer cualquiera de estos *desiderata* para una teoría. Sin

[15] En la terminología del presente artículo, el artículo de Martin y Woodruff prueba la existencia de puntos fijos máximos (no el punto fijo mínimo) dentro del contexto del enfoque trivalente débil. No desarrolla la teoría mucho más allá. Creo que el artículo no ha sido todavía publicado, pero será incluido en un volumen de próxima aparición dedicado a Yehoshua Bar–Hillel. Aunque anticipa parcialmente el enfoque aquí presentado, no era de mi conocimiento cuando realicé este trabajo.

embargo, la influencia de esta literatura sobre mi propia propuesta resultará obvia[16].[122]

III. LA PRESENTE PROPUESTA

No considero que ninguna propuesta, incluyendo la que he de presentar aquí, sea definitiva en el sentido de suministrar la interpretación del uso ordinario de «verdadero», o de dar la solución a las paradojas semánticas. Por el contrario, por ahora no he pensado a fondo en una justificación filosófica detallada de la propuesta, ni estoy seguro de cuáles son las áreas exactas y las limitaciones de su aplicabilidad. Espero que el modelo aquí suministrado tenga dos virtudes: primera, que proporcione un área rica en propiedades matemáticas y relativas a la estructura formal; segunda, que estas propiedades recojan en buena medida algunas intuiciones importantes. Así, pues, el modelo ha de ser puesto a prueba por su fertilidad técnica. No tiene que recoger todas las intuiciones, pero se espera que recoja muchas de ellas.

---

[16] De hecho tenía yo conocimiento de relativamente poca literatura sobre este tema cuando inicié el trabajo sobre el enfoque aquí presentado. Incluso ahora desconozco buena parte de esa literatura, de manera que es difícil trazar las conexiones. El trabajo de Martin parece ser el más cercano al presente enfoque en lo que respecta a sus consecuencias formales, no así en lo que respecta a sus bases filosóficas.

Hay también una literatura considerable sobre enfoques trivalentes o similares de las paradojas de la teoría de los conjuntos; aunque la desconozco en detalle parece estar estrechamente relacionada con el presente enfoque. Debería mencionar a Gilmore, Fitch y Feferman.

Siguiendo la literatura mencionada anteriormente, propongo investigar los lenguajes que permiten vacíos de verdad. A la manera de Strawson[17], podemos considerar una oración como un intento de hacer un enunciado, expresar una proposición, o cosas similares. La significatividad de una oración o el carácter de estar bien formada, radica en el hecho de que hay circunstancias especificables bajo las que tiene condiciones de verdad determinadas (bajo las que expresa una proposición), no en el hecho de que siempre exprese una proposición. Una oración como (1) es siempre significativa, pero bajo distintas circunstancias puede no «hacer un enunciado» o no «expresar una proposición». (No trato aquí de ser totalmente preciso filosóficamente.)

Para desarrollar cabalmente estas ideas, necesitamos un esquema semántico que nos permita manejar predicados que puedan estar sólo parcialmente definidos. Dado un dominio no vacío D, un predicado monádico P(x) se interpreta mediante un par (S_1, S_2) de conjuntos disyuntas de D. S_1 es la extensión de P(x) y S_2 es su antiextensión. P(x) ha de ser verdadero de los objetos en S_1, falso de aquéllos en S_2, [123] de otra manera será indefinido. La generalización de esto para predicados n–ádicos es obvia.

[17] Interpreto a Strawson como si sostuviera que «el actual rey de Francia es calvo» no logra constituir un enunciado pero que, sin embargo, es significativa, pues da las direcciones (condiciones) para hacer un enunciado. Aplico esta idea a las oraciones paradójicas sin comprometerme con respecto a su alegato original de las descripciones. Debería aclarar que la doctrina de Strawson es un tanto ambigua y que he elegido una de las interpretaciones preferidas, la cual, creo yo, también es la preferida por Strawson hoy en día.

Un esquema apropiado para manejar las conectivas es la lógica trivalente fuerte de Kleene. Supongamos que ¬P es verdadera (falsa) si P es falsa (verdadera) y que es indefinida si P es indefinida. Una disyunción es verdadera si al menos uno de los disyuntos es verdadero, sin importar si el otro de los disyuntos es verdadero, falso o indefinido[18]; es falsa si ambos disyuntos son falsos, de otra manera es indefinida. Las otras funciones de verdad pueden definirse en términos de la disyunción y de la negación de la manera usual. (En particular, entonces, una conjunción será verdadera cuando los dos conjuntos son verdaderos, falsa si al menos un conjunto es falso; de otra manera será indefinida.) (∃x)A(x) es verdadera si A(x) es verdadera para alguna asignación de un elemento de D a x; falsa si A(x) es falsa para todas las asignaciones a x, de otra manera será indefinida. (x)A(x) puede definirse como ¬(∃x) ¬A(x). Es, por lo tanto, verdadera si A(x) es verdadera para todas las asignaciones a x, falsa si A(x) es falsa para por lo menos una de dichas asignaciones, de otra manera es indefinida. Podríamos convertir lo anterior en una definición formal más precisa de la satisfacción, pero no nos tomaremos esa molestia[19]. [124]

---

[18] Así, la disyunción de «la nieve es blanca» con una oración del tipo del Mentiroso será verdadera. Si hubiésemos considerado que una oración del tipo del Mentiroso carece de significado, presumiblemente hubiéramos tenido que considerar que cualquier oración compuesta que la contuviera carecería también de significado.

[19] Las reglas de evaluación son las de S. C. Kleene en su *Introduction lo Metamathematics*. Nueva York, Van Nostrand, 1952, Sección 64, pp. 332–340. La noción de Kleene de tablas regulares es equivalente (para la clase de evaluaciones que él considera) a nuestra exigencia de la monotonicidad de N más adelante.

Me ha sorprendido mucho oír que el uso que hago de la evaluación de Kleene se compara ocasionalmente con la propuesta de quienes están en favor de abandonar la

Queremos apresar una intuición que de alguna manera es del siguiente tipo: Supóngase que estamos explicando la palabra «verdadero» a una persona que todavía no la entiende. Podemos decir que tenemos derecho a afirmar (o negar) con respecto a una oración que es verdadera precisamente cuando las circunstancias son tales que podemos afirmar (o negar) la oración misma. Nuestro interlocutor puede entonces entender lo que significa, por ejemplo, atribuir la verdad a (6) («la nieve es blanca»), pero puede aun sentirse desconcertado con respecto a las atribuciones de verdad a aquellas oraciones que contienen la palabra misma «verdadero». Dado que inicialmente no entendió estas oraciones, carecería igualmente de valor explicativo,

---

lógica estándar «para la mecánica clásica» o de postular valores de verdad extra, es decir, además de la verdad y la falsedad, etcétera. Esta reacción me sorprende a mi tanto como presumiblemente sorprendería a Kleene quien intenté escribir (como lo hago yo aquí) un trabajo de resultados matemáticos estándar susceptible de ser probado en la matemática convencional. «Indefinido» no es un valor de verdad extra, de la misma manera que —en el libro de Kleene— no es un número extra en la sección 63. Tampoco debería decirse que «la lógica clásica» no vale en general, ni que (en Kleene) el uso de funciones parcialmente definidas invalida la ley de la conmutatividad para la adición. Si algunas oraciones expresan proposiciones, cualquier función de verdad tautológica de ellas expresa una proposición verdadera. Obviamente las fórmulas que tienen componentes que no expresan proposiciones, incluso aquellas con forma de tautologías, pueden tener funciones de verdad que tampoco expresan proposiciones. (Esto sucede bajo la evaluación de Kleene pero no en la de van Fraasen.) Las meras convenciones para manejar los términos que no designan números no deberían de ser llamadas cambios en la aritmética; las convenciones para manejar las oraciones que no expresan proposiciones no son, en ningún sentido filosóficamente importante, «cambios en la lógica». La expresión «lógica trivalente», ocasionalmente usada aquí no debiera dar lugar a confusiones. Todas nuestras consideraciones pueden formalizarse en un metalenguaje clásico.

inicialmente, explicarle que llamar a esas oraciones «verdaderas» («falsas») equivale a afirmar (negar) la oración misma.

Sin embargo, la noción de verdad, como una noción que se aplica incluso a varias oraciones que contienen en sí mismas la palabra «verdadero», puede irse aclarando gradualmente a medida que reflexionamos más. Supongamos que consideramos la oración

(7) Alguna oración impresa en el *New York Daily News* del 7 de octubre de 1971, es verdadera.

(7) es un ejemplo típico de una oración que comprende el concepto mismo de verdad, de manera que, si (7) no es clara, tampoco lo será

(8) (7) es verdadera.

Sin embargo, si el sujeto en cuestión está dispuesto a afirmar «la nieve es blanca», estará dispuesto a afirmar de conformidad con las reglas «(6) es verdadera». Pero supongamos que entre las afirmaciones impresas en el *New York Daily News* del 7 de octubre de 1971 se encuentra (6) misma. Dado que nuestro sujeto está dispuesto a afirmar «(6) es verdadera» y a afirmar también «(6) está impresa en el *New York Daily News* del 7 de octubre de 1971», deducirá (7) me-[125]diante una generalización existencial. Una vez que esté dispuesto a afirmar (7), también estará dispuesto a afirmar (8). De este modo, el sujeto será capaz eventualmente de atribuir la verdad a más y más enunciados que contienen la noción misma de verdad. No hay ninguna razón para

suponer que todos los enunciados que contienen «verdadero» habrán de decidirse de esta manera, pero la mayor parte se decidirán. De hecho, nuestra sugerencia es que las oraciones «fundadas» pueden caracterizarse como aquellas que eventualmente llegan a tener un valor de verdad en este proceso.

Por supuesto, una oración típicamente infundada como (3) no recibirá ningún valor de verdad en el proceso que acabamos de esbozar. En particular, nunca será llamada «verdadera». Pero el sujeto no puede expresar este hecho diciendo «(3) no es verdadera». Dicha afirmación entraría directamente en conflicto con la estipulación según la cual se debe negar que una oración es verdadera precisamente en las circunstancias en las que uno negaría la oración misma. Al imponer esta estipulación hemos hecho una elección deliberada (véase más adelante).

Veamos cómo podemos dar a estas ideas una expresión formal. Sea L un lenguaje de primer orden del tipo clásico, interpretado, con una lista finita (o incluso denumerable) de predicados primitivos. Se asume que las variables recorren un dominio no vacío D y que los predicados primitivos n–arios se interpretan mediante relaciones n-arias (totalmente definidas) sobre D. La interpretación de los predicados de L se mantiene fija a lo largo de la discusión siguiente. Asumamos también que el lenguaje L es lo suficientemente rico como para poder expresar en L la sintaxis de L (digamos, mediante la aritmetización) y que algún esquema de codificación [*coding scheme*] codifica secuencias finitas de elementos de D en [*into*] elementos de D. No tratamos de presentar rigurosamente estas ideas; la noción de estruc-

tura «aceptable» de Y. N. Moschovakis lo haría[20]. Debo enfatizar que una buena parte de lo que haremos a continuación puede obtenerse cuando consideramos hipótesis mucho más débiles sobre L[21]. [126]

Supongamos que ampliamos L a un lenguaje Ł añadiéndole un predicado monádico T(x) cuya interpretación sólo necesita definirse parcialmente. Una interpretación de T(x) se da mediante un «conjunto parcial» (S_1, S_2) en donde S_1, como dijimos antes, es la extensión de T(x), S_2 es la antiextensión de T(x) y T(x) es indefinido para entidades fuera de S_1 U S_2. Sea Ł(S_1, S_2) la interpretación de Ł que resulta de interpretar T(x) mediante el par (S_1, S_2), quedando como antes los otros predicados de L[22]. Sea S'_1 el conjunto de (códigos de)[23] las oraciones verdaderas de Ł(S_1, S_2) y sea S'_1 el conjunto de todos los elementos de D que o no son (códigos de) oraciones de Ł(S_1, S_2) o son (códigos de)

---

[20] *Elementary Introduction on Abstract Structures,* Amsterdam, North Holland, 1974. La noción de estructura aceptable se desarrolla en el capitulo 5.

[21] Es innecesario suponer, como lo hicimos por mor de simplicidad, que todos los predicados en L están totalmente definidos. La hipótesis de que L contiene un artificio para codificar secuencias finitas sólo es necesaria si añadimos a L la satisfacción más que la verdad. Otras hipótesis pueden hacerse mucho más débiles para la mayor parte del trabajo.

[22] Ł es, así, un lenguaje con todos los predicados interpretados menos T(x). T(x) no está interpretado. El lenguaje Ł (S_1, S_2) y los lenguajes Łα definidos más adelante, son lenguajes obtenidos a partir de Ł al especificar una interpretación para T(x).

[23] Escribo entre paréntesis «códigos de» o «números de Gödel de» en varios lugares para recordar al lector que la sintaxis puede representarse en L mediante la asignación de números de Gödel o algún otro artificio codificador. Por descuido algunas veces omito la cualificación entre paréntesis, identificando las expresiones con sus códigos.

oraciones falsas de Ł(S_1, S_2). La elección de (S_1, S_2) determina de manera única a S'_1 y S'_2. Si T(x) ha de interpretarse como la verdad para el lenguaje mismo L que contiene al propio T(x), obviamente debemos tener $S_1=S'_1$ y $S_2=S'_2$. [Esto significa que si A es una oración cualquiera, A satisface (o falsifica) T(x) si y sólo si A es verdadera (falsa) conforme a las reglas de evaluación.]

Un par (S_1, S_2) que satisface esta condición se llama un punto fijo. Para que una determinada elección de (S_1, S_2) interprete T(x), establézcase que $\varphi((S_1, S_2)) = (S'_1, S'_2)$. φ es entonces una función unitaria definida sobre todos los pares (S_1, S_2) de subconjuntos disyuntos de D y los «puntos fijos» (S_1, S_2) son literalmente los puntos fijos de φ; es decir, son aquellos pares (S_1, S_2) tales que $\varphi(S_1, S_2) = (S'_1, S'_2)$. Si ($S_1$, S_2) es un punto fijo, algunas veces llamamos también a Ł(S_1, S_2) un punto fijo. Nuestra tarea básica es probar la existencia de puntos fijos e investigar sus propiedades.

Construyamos primeramente un punto fijo. Lo haremos considerando una «jerarquía de lenguajes» determinada. Comenzamos por definir el lenguaje interpretado $Ł_0$ como Ł(Λ, Λ) en donde Λ es el conjunto vacío; es decir, $Ł_0$ es el lenguaje en el que T(x) es totalmente indefinido. (Nunca es un punto fijo.) Para cualquier entero α, supongamos que hemos definido $Ł_\alpha = (S_1, S_2)$. Entonces establezca-[127]mos que $Ł_{\alpha+1}$= Ł(S'_1, S'_2), donde, como antes, S'_1 es el conjunto de (códigos de) oraciones verdaderas de $Ł_\alpha$ y S'_2 es el conjunto de todos los elementos de D que o no son (códigos de) oraciones de $Ł_\alpha$ o son (códigos de) oraciones falsas de $Ł_\alpha$.

La jerarquía de lenguajes que acabamos de dar es análoga a la jerarquía de Tarski para el enfoque ortodoxo. T(x) se interpreta en $Ł_{\alpha+1}$, como el predicado de verdad para $Ł_{\alpha}$. Pero surge un fenómeno interesante en el presente enfoque que se expondrá con detalle en los siguientes párrafos.

Digamos que (S^+_1, S^+_2) amplía a (S_1, S_2) [simbólicamente, $(S^+_1, S^+_2) \geq (S_1, S_2)$ o $(S_1, S_2) \leq (S^+_1, S^+_2)$] si y sólo si $S_1 \subseteq S^+_1$, $S_2 \subseteq S^+_2$,. Intuitivamente esto significa que si T(x) se interpreta por (S^+_1, S^+_2) la interpretación concuerda con la interpretación dada por (S_1, S_2) en todos los casos en los que esta última es definida; la única diferencia es que una interpretación por (S^+_1, S^+_2) puede dar lugar a que T(x) sea definida para algunos casos en los que era indefinida cuando se interpretaba por (S_1, S_2). Ahora, una propiedad básica de nuestras reglas de evaluación es la siguiente: φ es una operación monótona (que preserva el orden) sobre ≤; esto es, si $(S_1, S_2) \leq (S^+_1, S^+_2)$, $\varphi((S_1, S_2)) \leq \varphi((S^+_1, S^+_2))$. En otras palabras, si $(S_1, S_2) \leq (S^+_1, S^+_2)$ entonces cualquier oración que sea verdadera (o falsa) en $Ł(S_1, S_2)$ retiene su valor de verdad en $Ł(S^+_1, S^+_2)$. Lo que esto significa es que si la interpretación de T(x) se amplía dándole un valor de verdad definido a algunos casos previamente indefinidos, ningún valor de verdad previamente establecido cambiará ni se hará indefinido; cuando mucho, algunos valores de verdad previamente indefinidos se vuelven definidos. Esta propiedad —hablando técnicamente la monotonicidad de φ— es crucial para todas nuestras construcciones.

Dada la monotonicidad de φ, podemos deducir que para cada α, la interpretación de T(x) en $Ł_{\alpha+1}$ amplía la interpretación de T(x) en $Ł_{\alpha}$. El hecho es obvio para $\alpha = 0$, dado que, en $Ł_0$, T(x) es indefinido para toda x, cualquier interpretación de T(x) lo amplía automáticamente. Si la afirmación vale para $Ł_{\beta}$ —esto es, si la interpretación de T(x) en $Ł_{\beta+1}$, amplía la de T(x) en $Ł_{\beta}$— entonces cualquier oración verdadera o falsa en $Ł_{\beta}$, permanece verdadera o falsa en $Ł_{\beta+1}$. Si vemos las definiciones, esto dice que la interpretación de T(x) en $Ł_{\beta+2}$ amplía la interpretación de T(x) en $Ł_{\beta+1}$. Hemos, pues, probado por inducción que la interpretación de T(x) en $Ł_{\alpha+1}$, siempre amplía la interpretación de T(x) en $Ł_{\alpha}$ para toda a finita. Se sigue que el predicado T(x) crece, tanto en su extensión como en su antiextensión, a [128] medida que a crece. A medida que α crece un mayor número de oraciones llegan a ser declaradas verdaderas o falsas, pero una vez que una oración es declarada verdadera o falsa, conservará su valor de verdad en todos los niveles superiores.

Hasta aquí, hemos definido solamente los niveles finitos de nuestra jerarquía. Para α finita, sea $(S_{1,\alpha}, S_{2,\alpha})$ la interpretación de T(x) en $Ł_{\alpha}$. Tanto $S_{1,\alpha}$ como $S_{2,\alpha}$ crecen (como conjuntos) a medida que a crece. Hay entonces una manera obvia de definir el primer nivel «transfinito», llamémosle «$Ł_{\omega}$». Defínase simplemente $Ł_{\omega} = Ł(S_{1,\omega}, S_{2,\omega})$ en donde $S_{1,\omega}$ es la unión de todos los $S_{1,\alpha}$ para α finita y $S_{2,\omega}$ similarmente, es la unión de $S_{2,\alpha}$ para α finita. Dado $Ł_{\omega}$, podemos entonces definir $Ł_{\omega+1}$, $Ł_{\omega+2}$, $Ł_{\omega+3}$, etcétera, de la misma manera como lo hicimos para los niveles finitos. Cuando volvemos a llegan a un nivel «límite», tomamos una unión como lo hicimos antes.

Formalmente, definimos los lenguajes $Ł_\alpha$ para cada ordinal α. Si α es un ordinal sucesor ($\alpha = \beta + 1$), sea $Ł_\alpha = Ł(S_{1,\alpha}, S_{2,\alpha})$ en donde $S_{1,\alpha}$ es el conjunto de (códigos de) oraciones verdaderas de $Ł_\beta$ y $S_{2,\alpha}$ el conjunto consistente en todos los elementos de D que o son (códigos de) oraciones falsas de $Ł_\beta$ o no son (códigos de) oraciones de Ł. Si λ es un ordinal límite, $Ł_\lambda = (S_{1,\lambda}, S_{2,\lambda})$ en donde $S_{1,\lambda} = U_{\beta<\lambda} S_{1,\beta}$, $S_{2,\lambda} = U_{\beta<\lambda} S_{2,\beta}$. Así, en los niveles «sucesores» tomamos el predicado de verdad sobre el nivel previo y en los niveles límite (transfinitos) tomamos la unión de todas las oraciones declaradas verdaderas o falsas en niveles anteriores. Aun cuando incluyamos los niveles transfinitos, sigue siendo verdadero que la extensión y la antiextensión de T(x) crecen al crecer α.

Hay que notar que «crece» no significa «crece estrictamente»; hemos afirmado que $S_{1,\alpha} \subseteq S_{1,\alpha+1}$ (i = 1, 2), lo cual permite que sean iguales. ¿Continúa el proceso indefinidamente con cada vez más oraciones que se declaran verdaderas o falsas, o llega el momento en el que el proceso se para? Es decir, ¿hay un nivel ordinal σ para el cual $S_{1,\sigma} = S_{1,\sigma+1}$ y $S_{2,\sigma} = S_{2,\sigma+1}$, de manera que ningún «nuevo» enunciado se declare verdadero o falso en el siguiente nivel? La respuesta debe ser afirmativa. Las oraciones de Ł forman un conjunto. Si a cada nivel se decidieran nuevas oraciones de Ł, eventualmente agotaríamos Ł en algún nivel y ya no seríamos capaces de decidir ninguna más. Esto puede fácilmente convertirse en una prueba formal (la técnica es elemental y bien conocida por los lógicos) de que hay un nivel ordinal o tal que $(S_{1,\sigma}, S^{2,}_{\sigma}) = (S_{1,\sigma+1}, S_{2,\sigma+1})$. Pero dado que $(S_{1,\sigma+1}, S_{2,\sigma+1}) = \varphi((S_{1,\sigma}, S_{2,\sigma}))$, esto significa que $(S_{1,\sigma}, S_{2,\sigma})$ es un punto fijo. También puede probarse que es un punto fijo «mínimo» o «me-[129]nor»:

cualquier punto fijo amplía ($S_{1,\sigma}$, $S_{2,\sigma}$). Esto es, si una oración se evalúa como verdadera o falsa en $Ł_\sigma$, tiene el mismo valor de verdad en *cualquier* punto fijo.

Relacionemos con nuestras ideas intuitivas la construcción de un punto fijo que acabamos de dar. En la etapa inicial ($Ł_0$), T(x) es completamente indefinido. Esto corresponde a la etapa inicial en la que el sujeto no tiene ninguna comprensión de la noción de verdad. Dada una caracterización de la verdad mediante las reglas de evaluación de Kleene, el sujeto puede fácilmente ascender al nivel. $Ł_1$. Esto es, puede evaluar varios enunciados como verdaderos o falsos sin saber nada sobre T(x) —en particular, puede evaluar todas aquellas oraciones que no contienen T(x)—. Una vez que ha hecho la evaluación, amplia T(x), como en $Ł_1$. Entonces puede usar la nueva interpretación de T(x) para evaluar más oraciones como verdaderas o falsas y ascender a $Ł_2$, etcétera. Eventualmente, cuando el proceso se vuelve «saturado», el sujeto alcanza el punto fijo $Ł_\sigma$. (Al ser un punto fijo, $Ł_\sigma$ es un lenguaje que contiene su propio predicado de verdad.) Así, la definición formal que acabamos de dar constituye un buen paralelo de la construcción intuitiva previamente formulada[24].

[24] Una comparación con la jerarquía de Tarski: La jerarquía de Tarski usa un nuevo predicado de verdad en cada nivel, siempre cambia. Los niveles límite de la jerarquía de Tarski, que no han sido definidos en la literatura, pero que en alguna medida han sido definidos en mi propio trabajo, son enredosos de caracterizar.

La presente jerarquía usa un solo predicado de verdad, el cual crece cada vez más al aumentar los niveles hasta alcanzar el nivel del punto fijo mínimo. Los niveles límite se definen fácilmente. Los lenguajes en la jerarquía no son el objeto de interés Primordial, pero si son aproximaciones cada vez mejores al lenguaje mínimo con su propio predicado de verdad.

Hemos estado hablando de un lenguaje que contiene su propio predicado de verdad. Sin embargo, sería realmente más interesante ampliar un lenguaje arbitrario a otro lenguaje que contenga su propio predicado de *satisfacción*. Si L contiene un nombre para cada uno de los objetos de D y se define una relación de denotación (si D es no denumerable, esto significa que L contiene un número no denumerable de constantes), la noción de satisfacción se puede reemplazar de manera efectiva (para la mayoría de los propósitos) por la de verdad: por ejemplo, en lugar de decir que A(x) es satisfecho por un objeto a, podemos decir que A(x) se vuelve verdadero cuando la variable se reemplaza por un nombre de a. Basta entonces la construcción anterior. De manera alternativa, podemos ampliar L a Ł añadiendo un [130] predicado binario de satisfacción *Sat(s, x)* en el que *s* recorre secuencias finitas de elementos de D y x recorre fórmulas. Definimos una jerarquía de lenguajes, paralela a la que construimos antes para el caso de la verdad, que eventualmente alcanza un punto fijo —un lenguaje que contiene su propio predicado de satisfacción—. Si L es denumerable pero D no lo es, la construcción con la sola verdad se cierra en un ordinal contable, pero la construcción con la satisfacción puede cerrarse en un ordinal no contable. Más adelante continuaremos concentrándonos, con el fin de lograr simplicidad en la exposición, en la construcción con la verdad, pero la construcción con la satisfacción es más básica[25].

[25] Considérese el caso en el que L tiene un nombre canónico para cada elemento de D. Podemos entonces considerar pares (A, T), (A, F), en donde A es verdadero, o falso, respectivamente. Las reglas de Kleene corresponden a condiciones de clausura

La construcción puede generalizarse de manera que permita una notación en L mayor que la de la lógica de primer orden. Por ejemplo, podríamos tener un cuantificador que significara «para un número no contable de x», o un cuantificador del tipo de «la mayoría de», un lenguaje con infinitas conjunciones, etcétera. Hay una manera bastante canónica de ampliar, en el estilo de Kleene, la semántica de dichos cuantificadores y conectivas de tal manera que permitan vacíos de valores de verdad, pero no daremos aquí los detalles.

Constatemos que nuestro modelo satisface algunos de los desiderata mencionados en las secciones anteriores. Sin duda alguna es una teoría en el sentido exigido: cualquier lenguaje, incluyendo los que contienen teoría del número o sintaxis, puede ampliarse a un lenguaje con su propio predicado de verdad y el concepto de verdad [131] asociado se define matemáticamente mediante técnicas de la teoría de

---

sobre un conjunto de dichos pares: por ejemplo, si (A(α), F) $\in$ S para todo nombre del elemento α de D, póngase (($\exists$x)A(x), F) en S; si ((A(α), T) $\in$ S, póngase (($\exists$x)A(x), T) en S, etcétera. Considérese el más pequeño conjunto S de pares clausurados bajo los análogos de las reglas de Kleene, que contiene (A, T) (o(A, F)) para cada A atómica verdadera (o falsa) de L y clausurada conforme a las dos condiciones siguientes: (i) si (A, T) $\in$ S, (T(k), T) $\in$ S; (ii) si (A, F) $\in$ S, (T(k), F) $\in$ S, en donde «k» es una abreviatura de un nombre de A. Fácilmente se muestra que el conjunto corresponde (en el sentido obvio) al punto fijo mínimo [por tanto, está clausurado bajo las condiciones conversas de (i) y (ii)]. Usé esta definición para mostrar que el conjunto de verdades en el punto fijo mínimo (sobre una estructura aceptable) es inductivo en el sentido de Moschovakis. Probablemente es más simple que la definición dada en el texto. La definición dada en el texto tiene, entre otras ventajas, la de una definición de «nivel», facilitando una comparación con la jerarquía de Tarski y permitiendo la generalización cómoda a otros esquemas de evaluación distintos al de Kleene.

los conjuntos. No hay ningún problema con respecto a los lenguajes de nivel transfinito en la jerarquía.

Dada una oración A de Ł, definamos que A será fundada si tiene un valor de verdad en el punto fijo más pequeño $Ł_0$; de otra manera será infundada. Lo que hasta ahora ha sido, hasta donde yo sé, un concepto intuitivo sin ninguna definición formal, se vuelve un concepto definido con precisión en la presente teoría. Si A es fundada, defínase el nivel de A como el ordinal más pequeño a tal que A tiene un valor de verdad en $Ł_\alpha$.

Si Ł contiene teoría del número o sintaxis, no hay ningún problema de construir oraciones gödelianas que «dicen de sí mismas» que son falsas (oraciones del Mentiroso) o verdaderas [como en (3)]; puede mostrarse fácilmente que todas ellas son infundadas en el sentido de la definición formal. Si, por ejemplo, se usa la forma gödeliana de la paradoja del Mentiroso, la oración del Mentiroso puede tomar la forma siguiente:

(9) $(x)\,(P(x) \supset \sim T(x))$

en la que P(x) es un predicado sintáctico (o aritmético) que satisface únicamente (el número gödeliano de) la propia oración (9). De manera similar (3) toma la forma siguiente:

(10) $(x)(Q(x) \supset T(x))$

en la que Q(x) es satisfecho únicamente por (el número gödeliano de)

la oración (10). Bajo estas hipótesis, es fácil probar mediante una inducción sobre α que ni (9) ni (10) tendrán un valor de verdad en ningún $Ł_{\alpha}$; esto es, que son infundadas. Otros casos intuitivos de falta de fundamentación resultan de la misma manera.

En el modelo presente se aprecia con claridad el rasgo de los enunciados ordinarios que he enfatizado, a saber, que no hay ninguna garantía intrínseca de su seguridad (de que sean fundados) y que su «nivel» depende de hechos empíricos. Considérese, por ejemplo, (9) una vez más, sólo que ahora P(x) es un predicado empírico cuya extensión depende de hechos empíricos desconocidos. Si resulta que P(x) es verdadero solamente de la oración (9) misma, (9) será infundada como antes. Si la extensión de P(x) consiste enteramente de oraciones fundadas de los niveles, digamos, 2, 4 y 13, (9) será fundada y tendrá el nivel 14. Si la extensión de P(x) consiste de [132] oraciones fundadas de un nivel finito arbitrario, (9) será fundada y tendrá el nivel ϖ; y así sucesivamente.

Consideremos ahora los casos (4) y (5). Podemos formalizar (4) mediante (9), interpretando P(x) como «x es una oración que Nixon afirma acerca de Watergate» [Olvídese, por mor de simplicidad, que «acerca de Watergate» introduce un componente semántico en la interpretación de P(x).] Formalicemos (5) como

(11) $(x)\,(Q(x) \supset \sim T(x))$

interpretando Q(x) de la manera obvia. Para completar el paralelo con (4) y (5), supongamos que (9) está en la extensión de Q(x) y (11) está en

la extensión de P(x). Nada garantiza ahora que (9) y (11) hayan de ser fundadas. Supóngase, sin embargo, paralelamente a la discusión intuitiva anterior, que alguna oración verdadera satisface Q(x). Si el nivel más bajo de dicha oración es α, entonces (11) será falsa y fundada en el nivel $\alpha+1$. Si además todas las oraciones, diferentes de (11), que satisfacen P(x) son falsas, (9) será entonces fundada y verdadera. El nivel de (9) será por lo menos $\alpha+2$, debido al nivel de (11). Por otro lado, si alguna oración que satisface P(x) es fundada y verdadera, entonces (9) será fundada y falsa con nivel $\beta+1$,en donde β es el nivel más bajo de aquella oración. Para que el presente modelo pueda asignar niveles a (4) y (5) [(9) y (11)] es crucial que los niveles dependan de hechos empíricos y no que sean asignados de antemano.

Dijimos que los enunciados como (3), a pesar de ser infundados, no son intuitivamente paradójicos. Exploremos esto en términos del modelo propuesto. El punto fijo más pequeño de $Ł_\sigma$ no es el único punto fijo. Formalicemos (3) mediante (10), en donde Q(x) es un predicado sintáctico (de L) verdadero solamente de la propia oración (10). Supongamos que, en lugar de empezar nuestra jerarquía de lenguajes con T(x) completamente indefinido, hubiésemos empezado estableciendo que T(x) es verdadero de (10), de otra manera sería indefinido. Podemos entonces continuar la jerarquía de lenguajes exactamente como antes. Es fácil ver que si (10) es verdadera en el lenguaje de un nivel determinado, permanecerá verdadera en el siguiente nivel [usando el hecho de que Q(x) es verdadero solamente de (10), falso de todo lo demás]. A partir de esto podemos mostrar como antes que la interpretación de T(x) en cada nivel amplía todos los

niveles anteriores y que en algún nivel la construcción se cierra dando lugar a un punto fijo. La diferencia [133] es que (10), que carecía de valor de verdad en el punto fijo menor, es ahora verdadera.

Esto sugiere la siguiente definición: una oración es paradójica si no tiene valor de verdad en ningún punto fijo. Esto es, una oración paradójica A es tal que si $\varphi((S_1, S_2)) = (S_1, S_2)$, entonces A no es un elemento de S_1 ni un elemento de S_2.

(3) [o su versión formal (10)] es infundada, pero no paradójica. Esto significa que podríamos usar consistentemente el predicado «verdadero» de manera que se le diese un valor de verdad a (3) [o a (10)], aunque el proceso mínimo para asignar valores de verdad no se lo daría. Supongamos, por otro lado, con respecto a (9), que P(x) es verdadero de (9) misma y falso de todo lo demás, de manera que (9) es una oración del Mentiroso. Entonces el argumento de la paradoja del Mentiroso produce fácilmente una prueba de que (9) no puede tener un valor de verdad en ningún punto fijo. De manera que (9) es paradójica en nuestro sentido técnico. Nótese que, si el hecho de que P(x) es verdadero de (9) y falso de todo lo demás es meramente un hecho empírico, el hecho de que (9) sea paradójica será él mismo empírico. (Podríamos definir las nociones de «intrínsecamente paradójico», «intrínsecamente fundado» y otras, pero no lo haremos aquí.)

La situación parece ser intuitivamente la siguiente: Aunque el punto fijo más pequeño es probablemente el modelo más natural para el concepto intuitivo de verdad, y es el modelo generado por las instrucciones que nosotros dimos al sujeto imaginario, los otros puntos fijos nunca entran en conflicto con estas instrucciones. Po-

dríamos usar consistentemente la palabra «verdadero» de manera que otorgara un valor de verdad a una oración como (3) sin violar la idea de que se debe afirmar que una oración es verdadera precisamente en el caso en que hubiéramos afirmado la oración misma. No puede sostenerse lo mismo con respecto a las oraciones paradójicas.

Podemos probar, usando el lema de Zorn, que todo punto fijo puede ampliarse a un punto fijo máximo, en donde un punto fijo máximo es un punto fijo que no tiene ninguna extensión propia que sea también un punto fijo. Los puntos fijos máximos asignan «tantos valores de verdad como es posible»; no podrían asignarse más de manera consistente con el concepto intuitivo de verdad. Las oraciones como (3), aunque sean infundadas, tienen un valor de verdad en todo punto fijo máximo. Existen, sin embargo, oraciones infundadas que tienen valores de verdad en algunos puntos fijos máximos, pero no en todos. [134]

Resulta igualmente fácil construir puntos fijos que hacen falsa a (3), que construir puntos fijos que la hacen verdadera. De manera que la asignación de un valor de verdad a (3) es arbitraria. Ciertamente cualquier punto fijo que no asigne ningún valor de verdad a (3) puede ampliarse a puntos fijos que la hacen verdadera y a puntos fijos que la hacen falsa. Las oraciones fundadas tienen el mismo valor de verdad en todos los puntos fijos. Hay, sin embargo, oraciones infundadas no paradójicas que tienen el mismo valor de verdad en todos los puntos fijos en los que tienen un valor de verdad. Un ejemplo es el siguiente:

(12) o (12) o su negación es verdadera.

Es fácil mostrar que hay puntos fijos que hacen verdadera a (12) y ninguno que la haga falsa. No obstante, (12) es infundada (no tiene ningún valor de verdad en el punto fijo mínimo).

Llámese «intrínseco» a un punto fijo si y sólo si no asigna a ninguna oración un valor de verdad que entre en conflicto con su valor de verdad en cualquier otro punto fijo. Esto es, un punto fijo (S_1, S_2) es intrínseco si y sólo si no hay ningún otro punto fijo (S^+_1, S^+_2) y ninguna oración A de L' tal que $A \in (S_1 \cap S^+_2) \cup (S_2 \cap S^+_1)$. Decimos que una oración tiene un valor de verdad intrínseco si y sólo si algún punto fijo intrínseco le otorga un valor de verdad; es decir, A tiene un valor de verdad intrínseco si y sólo si hay un punto fijo intrínseco (S_1, S_2) tal que $A \in S_1 \cap S_2$. (12) es un buen ejemplo.

Hay oraciones no paradójicas que tienen el mismo valor de verdad en todos los puntos fijos en los que tienen valor de verdad, pero que, sin embargo, carecen de valor de verdad intrínseco. Considérese $P \vee \neg P$, en donde P es cualquier oración no paradójica infundada. Entonces, $P \vee \neg P$ es verdadera en algunos puntos fijos (a saber, en aquellos en los que P tiene un valor de verdad) y en ningún punto fijo es falsa. Sin embargo, supóngase que hay puntos fijos que hacen verdadera a P y puntos fijos que hacen falsa a P. [Por ejemplo, digamos, si P es (3).] Entonces, $P \vee \neg P$ no puede tener un valor de verdad en ningún punto fijo intrínseco, pues de acuerdo a nuestras reglas de evaluación, no puede tener un valor de verdad a menos de que uno de sus disyuntos lo tenga[26]. [135]

[26] Si usamos la técnica de superevaluación en lugar de las reglas de Kleene, $P \vee \neg P$ siempre será fundada y verdadera y tenemos que cambiar el ejemplo.

No hay ningún punto fijo que sea «el más grande» y que amplié cualquier otro punto fijo; efectivamente, cualesquiera dos puntos fijos que otorguen diferentes valores de verdad a la misma fórmula no tienen ninguna extensión en común. Sin embargo, no es difícil mostrar que hay un punto fijo intrínseco que es el más grande (y, ciertamente, que los puntos fijos intrínsecos forman una red [lattice] completa bajo $\underline{\bullet}$. El punto fijo intrínseco más grande es la única interpretación «más grande» de T(x) que es consistente con nuestra idea intuitiva de la verdad y que no hace una elección arbitraria en las asignaciones de verdad. Es, pues, en tanto que modelo, un objeto de interés teórico especial.

Es interesante comparar la «jerarquía de lenguajes de Tarski» con el presente modelo. Desgraciadamente esto es muy difícil de hacerse con toda generalidad sin introducir los niveles transfinitos, tarea que se omite en el presente esbozo. Pero podemos decir algo sobre los niveles finitos. Intuitivamente parecería que los predicados «verdadero$_n$» de Tarski son todos ellos casos especiales de un solo predicado de verdad. Por ejemplo, dijimos antes que «verdadero$_1$» significa «es una oración verdadera que no contiene verdad». Desarrollemos formalmente esta idea. Sea $A_1(x)$ un predicado sintáctico (aritmético) verdadero justamente de las fórmulas de Ł que no contienen T(x), es decir, de todas las fórmulas de L. $A_1(x)$, al ser sintáctico, es en sí mismo una fórmula de L, como lo son todas las otras fórmulas sintácticas que se mencionan más adelante. Defínase «$T_1(x)$» como «$T(x) \wedge A_1(x)$». Sea $A_2(x)$ un predicado sintáctico que se aplica a todas aquellas fórmulas cuyos predicados atómicos son los de L más «$T_1(x)$». [De manera más

precisa, la clase de dichas fórmulas puede definirse como la clase más pequeña que incluye todas las fórmulas de L y $T(x_i) \wedge A_1(x_i)$, para cualquier variable x_i clausuradas bajo la cuantificación y las funciones de verdad.] Defínase entonces $T_2(x)$ como $T(x) \wedge A_2(x)$. En general, podemos definir $A_{n+1}(x)$ como un predicado sintáctico que se aplica precisamente a las fórmulas construidas a partir de los predicados de L y $T_n(x)$, y $T_{n+1}(x)$ como $T(x) \wedge A_{n+1}(x)$. Asumamos que T(x) es interpretada por el punto fijo más pequeño (o cualquier otro). Entonces es fácil probar por inducción que cada predicado $T_n(x)$ es totalmente definido, que la extensión de $T_0(x)$ consiste precisamente en las fórmulas verdaderas del lenguaje L, en tanto que la extensión de $T_{n+1}(x)$ consiste en las fórmulas verdaderas del lenguaje obtenido al añadir $T_n(x)$ a L. Esto significa que todos los predicados de verdad de la jerarquía finita de Tarski son defi-[136]nibles dentro de $Ł_\sigma$, y que todos los lenguajes de esa jerarquía son sublenguajes de $Ł_\sigma$[27]. Este tipo de resultado podría ampliarse al transfinito si hubiéramos definido la jerarquía transfinita de Tarski.

Hay otros resultados más difíciles de formular en el presente esbozo. Las oraciones en la jerarquía de Tarski se caracterizan por ser seguras (intrínsecamente fundadas) y por ser intrínseco su nivel, dado independientemente de los hechos empíricos. Resulta natural conjetu-

[27] Suponemos que la jerarquía de Tarski define $L_0 = L$, $L_{n+1} = L + T_{n+1}(x)$ (verdad, o satisfacción, para L_n). De manera alternativa, podríamos preferir la construcción inductiva $L_0 = L$, $L_{n+1} = L_n + T_{n+1}(x)$, en la que el lenguaje de cada nuevo nivel contiene todos los predicados de verdad previos. Es fácil modificar la construcción presentada en el texto de manera que concuerde con la segunda definición. Las dos jerarquías alternativas son equivalentes en lo que respecta al poder expresivo en cada nivel.

rar que toda oración fundada con nivel intrínseco n es, en algún sentido, «equivalente» a una oración de nivel n en la jerarquía de Tarski. Dadas las definiciones adecuadas de «nivel intrínseco», «equivalente» y otras similares, pueden formularse y probarse teoremas de esta clase, e incluso pueden ampliarse al transfinito.

Hasta aquí hemos asumido que los vacíos de verdad han de manejarse de acuerdo a los métodos de Kleene. No es de ninguna manera necesario hacer esto. Casi cualquier esquema para manejar vacíos de verdad puede ser usado, con tal de que se conserve la propiedad básica de la monotonicidad de φ; esto es, a condición de que al ampliar la interpretación de T(x) nunca cambie el valor de verdad de ninguna oración de Ł, sino que, a lo más, se otorguen valores de verdad a los casos que se hallaban previamente indefinidos. Dado cualquier esquema de este tipo, podemos usar los argumentos anteriores para construir el punto fijo mínimo y otros puntos fijos, definir los niveles de las oraciones y las nociones de «fundado», «paradójico», etcétera.

Un esquema que puede usarse de esta manera es la noción de superevaluación introducida por van Fraassen[28]. La definición es fácil para el lenguaje Ł. Dada una interpretación (S_1, S_2) de T(x) en Ł, llámese verdadera (falsa) a una fórmula A si y sólo si resulta verdadera (falsa) conforme a la evaluación ordinaria clásica bajo toda interpretación (S^+_1, S^+_2) que amplía (S_1, S_2) y es totalmente definida, es decir, que es tal que $S^+_1 \cup S^+_2 = D$. Podemos entonces definir como antes la jerarquía $\{Ł_\alpha\}$ y el punto fijo mínimo $Ł_\sigma$. Bajo la interpreta-[137]ción-

[28] Véase su articulo «Singular Terms, Truth-value Gaps and Free Logic» publicado en *The Journal of Philosophy.* LXII1, 17, septiembre 15 de 1966, pp. 481–495.

superevaluación, todas las fórmulas que pueden probarse en la teoría clásica de la cuantificación se vuelven verdaderas en $Ł_\sigma$; bajo la evaluación de Kleene solamente se podía decir que eran verdaderas en el caso de ser definidas. Gracias al hecho de que $Ł_\sigma$ contiene su propio predicado de verdad, no tenemos que expresar este hecho mediante un esquema, o mediante un enunciado de un metalenguaje. Si PQT(x) es un predicado sintáctico verdadero justamente de las oraciones de Ł que pueden probarse en la teoría de la cuantificación, podemos afirmar:

(13) $(x)\,(PQT(x) \supset T(x))$

y (13) será verdadera en el punto fijo mínimo.

Hemos usado aquí superevaluaciones en las que se toman en cuenta todas las ampliaciones totales de la interpretación de T(x). Es natural considerar que hay restricciones sobre la familia de las extensiones totales; dichas restricciones son motivadas por las propiedades intuitivas de la verdad. Por ejemplo, podríamos considerar solamente las interpretaciones consistentes (S^+_1, S^+_2), en donde (S^+_1, S^+_2) es consistente si y sólo si S_1 no contiene ninguna oración junto con su negación. Podríamos entonces definir que A es verdadera (falsa) con T(x) interpretada por (S_1, S_2) si y sólo si A es verdadera (falsa) clásicamente cuando A se interpreta por cualquier extensión consistente totalmente definida de (S_1, S_2).

(14) $(x)\,\neg(T(x) \wedge T(neg(x)))$

será verdadera en el punto fijo mínimo. Si hemos restringido las extensiones totales admisibles a aquellas que definen conjuntos consistentes, máximos de oraciones, en el sentido usual, resultará verdadera en el punto fijo mínimo[29], no sólo (14), sino incluso

(x) $(\text{Oraci}(x) \supset T(x) \vee T(\text{neg}(x)))$

Sin embargo, esta última fórmula debe interpretarse cuidadosamente, pues aún no es el caso, ni siquiera bajo la interpretación-superevaluación en cuestión, que haya algún punto fijo que haga ver-[138]dadera a cualquier fórmula o su negación. (Las fórmulas paradójicas siguen careciendo de valor de verdad en todos los puntos fijos.) El fenómeno se halla asociado con el hecho de que, bajo la interpretación-superevaluación, puede ser verdadera una disyunción sin que de esto se siga que algún disyunto sea verdadero.

No es el propósito del presente trabajo hacer ninguna recomendación particular entre el enfoque trivalente fuerte de Kleene, los enfoques de superevaluación de van Fraassen, o cualquier otro esquema (como la lógica trivalente débil de Frege, preferida por Martin y Woodruff, aunque me inclino tentativamente a considerar que este último es excesivamente aparatoso). Ni siquiera es mi propósito presente hacer alguna recomendación firme entre el punto fijo mínimo de un esquema particular de evaluación y los otros muchos puntos

[29] Una paradoja del Mentiroso debida a H. Friedman muestra que hay límites a lo que puede hacerse en esta dirección.

fijos[30]. Ciertamente no hubiéramos podido definir la diferencia intuitiva entre «fundado» y «paradójico» si no hubiéramos echado mano de los puntos fijos no mínimos. Mi propósito, más bien, es suministrar toda una familia de instrumentos flexibles que pueden explorarse simultáneamente y cuya fertilidad y consonancia con la intuición pueden constatarse.

Tengo alguna incertidumbre con respecto a que haya una cuestión fáctica definida sobre si el lenguaje natural maneja los vacíos de verdad —por lo menos aquellos que surgen en conexión con las paradojas semánticas— mediante los esquemas de Frege, Kleene, van Fraassen, o quizá algún otro. Ni siquiera estoy completamente seguro de que haya una cuestión de hecho definida con respecto a si el lenguaje natural debiera evaluarse mediante el punto fijo mínimo o mediante otro, dada la variedad de esquemas que se pueden elegir para manejar los vacíos[31]. Por el momento no estamos buscando el esquema correcto. [139]

---

[30] Aunque el punto fijo mínimo se distingue ciertamente por ser natural en muchos respectos.

[31] No es mi intención afirmar que no hay ninguna cuestión de hecho definida en estas áreas, o incluso que yo mismo no pueda estar en favor de algunos esquemas de evaluación frente a otros. Pero mis ideas personales son menos importantes que la variedad de herramientas a nuestra disposición, de manera que, para los propósitos de este esbozo, asumo una posición agnóstica. (Hago notar que si se asume el punto de vista de que la lógica se aplica en primer lugar a las proposiciones, y que estamos solamente formulando convenciones sobre cómo manejar las oraciones que no expresan proposiciones, el atractivo del enfoque que introduce la superevaluación disminuye frente al enfoque de Kleene. Véase la nota 18.)

El presente enfoque puede aplicarse a los lenguajes que contienen operadores modales. En este caso, no solamente consideramos la verdad, sino que nos es dado un sistema de mundos posibles, a la manera usual en la teoría modal de los modelos, y evaluamos la verdad y T(x) en cada mundo posible. La definición inductiva de los lenguajes $Ł_\alpha$ que se aproximan al punto fijo mínimo tiene que modificarse conformemente. No podemos dar aquí los detalles[32].

La aplicación del enfoque presente a los lenguajes con operadores modales, irónicamente, puede ser de algún interés para aquellos a quienes les desagradan los operadores intensionales y los mundos posibles y prefieren considerar las modalidades y las actitudes preposicionales como predicados de oraciones verdaderas (o de ejemplares particulares de oraciones). Montague y Kaplan, haciendo uso de las aplicaciones elementales de las técnicas gödelianas, han señalado que dichos enfoques pueden conducir probablemente a paradojas semánticas similares a la del Mentiroso[33]. A pesar de que se co-[140]noce la

[32] Otra aplicación de las técnicas presentes es a la cuantificación sustitucional «impredicativa», en la que los términos de la clase de sustitución contienen cuantificadores sustitucionales del tipo dado. (Por ejemplo, un lenguaje que contiene cuantificadores sustitucionales que tienen como sustituyentes oraciones arbitrarias del lenguaje mismo.) En general, es imposible introducir dichos cuantificadores en los lenguajes clásicos sin vacíos de verdad.

[33] Richard Montague, «Syntactical Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability», *Acta Philosophica Fennica. Proceedings of a Colloquium on Modal and Many Valued Logics,* 1963, pp. 153–167; David Kaplan y Richard Montague, «A Paradox Regained», *Notre Dame Journal of Formal Logic,* 1, 3, julio de 1960, pp. 79–90.

En la actualidad se sabe que los problemas surgen solamente si las modalidades y

dificultad desde hace tiempo, la extensa literatura en favor de dichos tratamientos, en general, ha ignorado simplemente el problema en lugar de indicar cómo debería solucionarse (por ejemplo, ¿mediante una jerarquía de lenguajes?). Ahora bien, si admitimos un operador de necesidad y un predicado de verdad, podríamos definir un predicado de necesidad Nec(x) aplicado a las oraciones, o bien mediante □T(x) o mediante T(nec(x)) dependiendo de nuestro gusto[34], y tratarlo de acuerdo al esquema de mundos posibles esbozado en el párrafo anterior. (No creo que ningún predicado de necesidad de oraciones deba considerarse intuitivamente como derivado, definido en términos de un operador y un predicado de verdad. Pienso también que esto es

---

las actitudes son predicados aplicados a oraciones o a sus ejemplares particulares. Los argumentos de Kaplan–Montague no se aplican a las formalizaciones estándar que toman las modalidades o las actitudes proposicionales como operadores intensionales. Incluso si quisiéramos cuantificar sobre los objetos de las creencias, los argumentos no se aplican si se considera que los objetos de las creencias son proposiciones y si estas últimas se identifican con conjuntos de mundos posibles.

Sin embargo, si cuantificamos sobre proposiciones, pueden surgir paradojas en conexión con las actitudes proposicionales dadas determinadas premisas empíricas apropiadas. [Véase, por ejemplo, A. N. Prior, «On a Family of Paradoxes», *Notre Dame Journal of Formal Logic,* 11, 1, enero de 1961, pp. 16–32]. También es posible que queramos individuar las proposiciones (en conexión con las actitudes proposicionales, pero no con las modalidades) de una manera más fina y no mediante conjuntos de mundos posibles. Es posible que dicha «estructura fina» pueda permitir la aplicación de los argumentos gödelianos, del tipo de los usados por Montague y Kaplan, directamente a las proposiciones.

[34] La segunda versión es mejor en términos generales, en tanto que formalización del concepto propuesto por quienes hablan de las modalidades y de las actitudes como predicados de oraciones. Esto es verdad especialmente para el caso de las actitudes proposicionales.

cierto con respecto a las actitudes proposicionales.) Podemos incluso «dar una patada a la escalera» y tomar como primitivo Nec(x), tratándolo en un esquema de mundos posibles como si estuviese definido por un operador más un predicado de verdad. Observaciones similares valen para las actitudes proposicionales si, haciendo uso de los mundos posibles, estamos dispuestos a tratarlas como operadores modales. (Personalmente pienso que dicho tratamiento supone considerables dificultades filosóficas.) Es posible que el presente enfoque pueda ser aplicado a los supuestos predicados de oraciones en cuestión sin usar ni operadores intensionales ni mundos posibles, pero por el momento, no tengo ninguna idea de cómo hacer esto.

Parece probable que muchos de quienes han trabajado sobre el enfoque de las paradojas semánticas que introduce los vacíos de verdad, hayan tenido esperanzas de encontrar un lenguaje universal en el que todo lo que de alguna manera se puede enunciar, se pueda expresar. (La prueba dada por Gödel y Tarski de que un lenguaje no puede contener su propia semántica, se aplicaba sólo a los lenguajes que no tienen vacíos de verdad). Ahora bien, los lenguajes considerados en el presente enfoque contienen sus propios predicados de verdad e incluso sus propios predicados de satisfacción y así, en esta medida, aquellas esperanzas se han realizado. Sin embargo, el presente enfoque ciertamente no pretende suministrar un lenguaje universal y dudo que pueda alcanzarse semejante meta. Primero, la in-[141]ducción que define el punto fijo mínimo se lleva a cabo en un metalenguaje de la teoría de los conjuntos, no en el lenguaje objeto mismo. Segundo, hay afirmaciones que podemos hacer sobre el lenguaje objeto que no podemos hacer en el lenguaje objeto. Por ejemplo, las oraciones del

Mentiroso no son verdaderas en el lenguaje objeto, en el sentido de que el proceso inductivo nunca las hace verdaderas; pero estamos imposibilitados para decir esto en el lenguaje objeto debido a nuestra interpretación de la negación y del predicado de verdad. Si pensamos que el punto fijo mínimo —digamos, bajo la evaluación de Kleene— nos proporciona un modelo para el lenguaje natural, entonces, el sentido en el que podemos decir, en el lenguaje natural, que una oración del Mentiroso no es verdadera, tiene que concebirse como asociado a alguna etapa posterior en el desarrollo del lenguaje natural, una etapa en la que los hablantes reflexionan sobre el proceso de generación que conduce al punto fijo mínimo. Ésta no es en sí misma parte de dicho proceso. La necesidad de ascender a un metalenguaje puede ser una de las debilidades de la presente teoría. El fantasma de la jerarquía de Tarski está aún con nosotros[35].

El enfoque que hemos adoptado aquí presupone la siguiente versión de la «Convención T» de Tarski, adaptada al enfoque trivalente: Si

[35] Nótese que el metalenguaje en el que escribimos este articulo puede considerarse como si no contuviera ningún vacío de verdad. Una oración, o tiene o no tiene un valor de verdad en un punto fijo determinado.

Las nociones semánticas tales como «fundado», «paradójico», etcétera, pertenecen al metalenguaje. Me parece que esta situación es intuitivamente inaceptable en contraste con la noción de verdad, ninguna de estas otras nociones ha de encontrarse en el lenguaje natural con toda su claridad prístina antes de que los filósofos reflexionen sobre su semántica (en particular, sobre las paradojas semánticas). Si abandonamos la meta de un lenguaje universal, los modelos del tipo presentado en este trabajo resultan plausibles en tanto que modelos del lenguaje natural en una etapa anterior a que reflexionemos sobre el proceso de generación asociado con el concepto de verdad, la etapa que continúa con la vida cotidiana de los hablantes que no son filósofos.

«k» es una abreviatura de un nombre de una oración A, T(k) será verdadera, o falsa respectivamente, si y sólo si A es verdadera, o falsa. Esto recoge la intuición de que T(k) tendrá un vacío de verdad si A lo tiene. Una intuición alternativa[36] afirmaría que, si A es falsa o indefinida, entonces A no es verdadera y T(k) deberá ser falsa y su[142] negación verdadera. De acuerdo a esta posición, T(x) será un predicado, totalmente definido y no habrá ningún vacío de verdad. La Convención T de Tarski debe presumiblemente restringirse de alguna manera.

No es difícil modificar el presente enfoque de tal manera que podamos acomodar dicha intuición alternativa. Tómese cualquier punto fijo L'(S_1, S_2). Modifíquese la interpretación de T(x) a manera de hacerlo falso de cualquier oración fuera de S. [Llamamos a esto «cerrar» T(x).] Una versión modificada de la Convención T de Tarski vale en el sentido del condicional $T(k) \vee T(neg(k)) . \supset . A = T(k)$. En particular, si A es una oración paradójica, podemos ahora afirmar $\neg T(k)$. De manera equivalente, si A tenía un valor de verdad antes de que se cerrara T(x), entonces A = T(k) es verdadera.

Dado que el lenguaje objeto obtenido al cerrar T(x) es un lenguaje clásico con todos los predicados totalmente definidos, es posible definir a la manera tarskiana usual un predicado de verdad para ese lenguaje. Este predicado no coincidiría en extensión con el predicado T(x) del lenguaje objeto y ciertamente es razonable suponer que

[36] Creo que puede defenderse la primacía de la primera intuición, y es por esta razón que he enfatizado el enfoque basado en dicha intuición. La otra intuición surge solamente después de haber reflexionado sobre el proceso que encama la primera intuición. Véase lo anteriormente dicho.

realmente es el predicado del metalenguaje el que expresa el concepto «genuino» de verdad del lenguaje objeto cerrado; el T(x) del lenguaje cerrado define la verdad para el punto fijo antes de que el lenguaje se cerrara. De manera que aún no podemos evitar la necesidad de un metalenguaje.

El hecho de parecer evasiva la meta de un lenguaje universal ha llevado a algunos a concluir que son estériles aquellos enfoques que aceptan los vacíos de verdad, o cualquier enfoque que intente acercarse más al lenguaje natural de lo que lo hace el enfoque ortodoxo. Espero que la fertilidad del presente enfoque y su concordancia con las intuiciones sobre el lenguaje natural en una gran cantidad de casos, arrojen dudas sobre tales actitudes negativas.

Hay aplicaciones matemáticas y problemas puramente técnicos que no he mencionado en este esbozo; rebasarían el campo de un artículo destinado a una revista filosófica. Así, hay el problema —que puede contestarse con bastante generalidad— de caracterizar el ordinal o en el que se cierra la construcción del punto fijo mínimo. Si L es un lenguaje de la aritmética de primer orden, resulta que σ es ϖ_1, el primer ordinal no recursivo. Un conjunto es la extensión de una fórmula con una variable libre en $Ł_\sigma$ si y sólo si es π^1_1 y es la extensión de una fórmula totalmente definida si y sólo si es hiperaritmético. Los lenguajes $Ł_\alpha$ que se aproximan al punto fijo mínimo dan una versión «libre de notación» [*notationfree*] de la jerarquía hipera-[143]ritmética que resulta interesante. De manera más general, si L es el lenguaje de una estructura aceptable, en el sentido de Moschovakis, v si se usa la evaluación de Kleene, un conjunto es la extensión de una fórmula

monádica en el punto fijo mínimo si y solo si es inductivo en el sentido de Moschovakis[37].

[37] Leo Harrington me informa que ha probado la conjetura de que un conjunto es la extensión de una fórmula monádica totalmente definida si y sólo si es hiperelemental. Si L es una teoría del número, el caso especial de II^1_1, y los conjuntos hiperaritméticos es independiente de si se usa la formulación de Kleene o la de van Fraassen. Esto no es así para el caso general en el que la formulación de van Fraassen conduce mas bien a los conjuntos II^1_1 que a los conjuntos inductivos.