

NOTA CRÍTICA SOBRE *LA PARADOJA DE ORAYEN*

Moretti, A. y G. Hurtado (comps.), *La Paradoja de Orayen*, Eudeba, Buenos Aires, 2003.

AGUSTÍN RAYO

Department of Linguistics and Philosophy
Massachusetts Institute of Technology
arayo@mit.edu

La Paradoja de Orayen es dos cosas en una. Primeramente, es un homenaje al filósofo argentino Raúl Orayen (1942–2003). Pocos filósofos hispanoamericanos han gozado de la solidez intelectual y agudeza filosófica de Orayen, y pocos han sido tan queridos. Se trata, pues, de un homenaje bien merecido y que mucho agradecemos los que tuvimos la fortuna de interactuar con Raúl y aprender de él. En segundo lugar, el libro es una contribución a la filosofía hispanoamericana. Alberto Moretti y Guillermo Hurtado tuvieron el acierto de reconocer el valor de un proyecto que la prematura muerte de Orayen dejó inconcluso, y apreciar su potencial para generar discusión filosófica de alto nivel. El resultado es un volumen que recompensará la atención de sus lectores, y dará al trabajo de Orayen justa prominencia en el mundo hispanoamericano.

El cuerpo principal del libro consiste en cuatro ensayos filosóficos y tres cartas. El primero de los ensayos es un artículo previamente inédito de Orayen, en el que se presenta la paradoja que da título al volumen y se proponen dos soluciones. Los otros tres ensayos —de Carlos Alchourrón, Adolfo García de la Sienra y Mario Gómez Torrente— y las cartas —de Willard Quine, Hilary Putnam y William Hart— son discusiones de la Paradoja. En el ensayo de Alchourrón se argumenta que una lección de la Paradoja es que las teorías de conjuntos sin conjunto universal resultan filosóficamente inadecuadas.¹ En el ensayo de García de la Sienra se argumenta que una lección de la Paradoja es que la teoría de modelos es sólo el primer nivel en una jerarquía de teorías semánticas cada vez más poderosas. En el ensayo de Gómez Torrente se argumenta que la

¹ Una buena discusión de tales teorías puede encontrarse en Forster 1995.

contribución de Orayen consiste en haber hecho patente una dificultad en el análisis de la noción de interpretación (no en haber identificado una paradoja), y se critica una caracterización potencial de esta noción basada en la primera de las soluciones que propone Orayen.

La obra contiene también tres textos suplementarios: un ensayo de Eduardo Barrio en el que se presenta un esbozo de conceptos y resultados metalógicos que preparan al lector para el resto del volumen; una introducción de Moretti, en la que se narra la historia intelectual de la Paradoja y se explican los orígenes de este libro, y un texto de Hurtado en el que se describe el pensamiento de Orayen acerca de otros problemas en la lógica filosófica y en otras áreas de la filosofía. Finalmente, el volumen cuenta con dos bibliografías; la primera, compilada por Hurtado, es una lista de todas las publicaciones filosóficas de Orayen; la segunda, compilada por Barrio, es una lista de lecturas recomendadas. Todo esto da como resultado un libro muy atractivo —y con sólo 108 páginas, un libro que se lee con mucha facilidad—.

En la contribución de Orayen, se razona como sigue: el lenguaje de la teoría de conjuntos trata sobre todos los conjuntos; por tanto, el dominio de un modelo que capture la interpretación deseada del lenguaje de la teoría de conjuntos tendría que consistir en todos los conjuntos. Sin embargo, el dominio de un modelo es un conjunto, y de acuerdo con las teorías de conjuntos estándar, no existe el conjunto de todos los conjuntos. Por tanto, ningún modelo puede capturar la interpretación deseada del lenguaje de la teoría de conjuntos. Este razonamiento es la Paradoja de Orayen.

La pregunta crucial es, pues, la siguiente:

LA PREGUNTA CRUCIAL

¿Podríamos formular la interpretación deseada del lenguaje de la teoría de conjuntos sin apelar a modelos (o, por lo menos, sin apelar al tipo de modelo que da lugar al problema)?

Para poder responder a esta pregunta es necesario saber más acerca de qué es lo que uno espera obtener de una interpretación del lenguaje de la teoría de conjuntos.

Una primera razón por la que podríamos estar interesados en formular tal interpretación es para hacer *semántica*, es decir, caracterizar un predicado de verdad que permita probar, con base en los significados de las expresiones atómicas del lenguaje, cada instancia del siguiente esquema:

(T) e es verdadero si y sólo si p

(donde “ e ” es reemplazado por el nombre de un enunciado del lenguaje objeto y “ p ” es reemplazado por una traducción de ese enunciado al metalenguaje).

Consideremos un ejemplo. Supongamos que \mathcal{L}^E es un lenguaje de primer orden cuya única expresión no lógica es el predicado diádico “ $P(\dots, \dots)$ ”. Supongamos, además, que \mathcal{L}^E es un lenguaje interpretado no en el sentido de tener modelo deseado, sino en el sentido de tener traducción a una versión del castellano con variables y subíndices. El método de traducción puede establecerse de manera composicional. Para fórmulas atómicas de la forma $P^\ulcorner(x_i, x_j)\urcorner$, procedamos como sigue:

$\ulcorner P(x_i, x_j)\urcorner$ se traduce como $\ulcorner x_i$ es amigo de $x_j\urcorner$.

En el caso de los cuantificadores requerimos una traducción que refleje el dominio de discurso de \mathcal{L}^E . Digamos, por ejemplo, que \mathcal{L}^E trata de todos y sólo los elefantes:

$\ulcorner \exists x_i(\dots)\urcorner$ se traduce como \ulcorner existe un elefante, x_i , tal que $\dots\urcorner$

Las expresiones restantes se traducen del modo usual:

$\ulcorner x_i = x_j\urcorner$ se traduce como $\ulcorner x_i$ es idéntico a $x_j\urcorner$

“ $\neg \dots$ ” se traduce como “no es el caso que \dots ”

“ $\dots \vee \dots$ ” se traduce como “o bien es el caso que \dots , o bien es el caso que \dots ”

He aquí una manera de producir una semántica para \mathcal{L}^E . Comenzamos asignándole a cada expresión atómica de \mathcal{L}^E un valor semántico:

- A1. El *dominio* de \mathcal{L}^E es el conjunto de elefantes.
- A2. El valor semántico de “ P ” es el conjunto de pares ordenados $\langle x, y \rangle$, donde x y y están en el dominio de \mathcal{L}^E , y x es amigo de y .
- A3. El valor semántico de “ $=$ ” es el conjunto de pares ordenados $\langle x, x \rangle$, donde x está en el dominio de \mathcal{L}^E .
- A4. El valor semántico de “ \exists ” es la función de subconjuntos del dominio de \mathcal{L}^E a valores de verdad que le asigna “F” al conjunto vacío, y “V” a todo lo demás.

- A5. El valor semántico de “ \neg ” es la función de valores de verdad a valores de verdad que le asigna “F” a “V” y “V” a “F”.
- A6. El valor semántico de “ \forall ” es la función de pares ordenados de valores de verdad a valores de verdad que le asigna “F” a “ $\langle F, F \rangle$ ” y “V” a todo lo demás.

Con esta asignación de valores semánticos a bordo, puede caracterizarse el siguiente predicado de verdad para \mathcal{L}^E :

LA CARACTERIZACIÓN SIMPLE

Una secuencia es una función que asigna a cada variable del lenguaje un objeto en el dominio del lenguaje.

La secuencia s satisface la fórmula ϕ si y sólo si $\langle \phi, s \rangle$ es miembro de todo conjunto α con las siguientes características:

$\langle \ulcorner P(x_i, x_j) \urcorner, s \rangle \in \alpha$ si y sólo si $\langle s(\ulcorner x_i \urcorner), s(\ulcorner x_j \urcorner) \rangle$ es miembro del valor semántico de ‘ P ’.

$\langle \ulcorner x_i = x_j \urcorner, s \rangle \in \alpha$ si y sólo si $\langle s(\ulcorner x_i \urcorner), s(\ulcorner x_j \urcorner) \rangle$ es miembro del valor semántico de “=”.

$\langle \ulcorner \exists x_i(\psi) \urcorner, s \rangle \in \alpha$ si y sólo si hay una secuencia s' tal que: (a) s' difiere de s a lo más en el objeto que le asigna a $\ulcorner x_i \urcorner$, y (b) $\langle \psi, s' \rangle \in \alpha$.

$\langle \ulcorner \neg \psi \urcorner, s \rangle \in \alpha$ si y sólo si $\langle \psi, s \rangle \notin \alpha$.

$\langle \ulcorner \psi \vee \theta \urcorner, s \rangle \in \alpha$ si y sólo si o bien $\langle \psi, s \rangle \in \alpha$ o $\langle \theta, s \rangle \in \alpha$.

ϕ es verdadero si y sólo si toda secuencia satisface ϕ .

No es difícil verificar que la caracterización simple y A1–A6 implican cada instancia del esquema (T) (véase Tarski 1929).

En el caso de la paradoja de Orayen, el lenguaje que nos concierne no es \mathcal{L}^E , sino \mathcal{L}^C . Desde el punto de vista sintáctico, no hay ninguna diferencia entre \mathcal{L}^C y \mathcal{L}^E , pero la interpretación de \mathcal{L}^C es ligeramente diferente de la de \mathcal{L}^E : en lugar de hablar de amistad entre elefantes, hablaremos de membresía entre conjuntos. En particular:

$\ulcorner P(x_i, x_j) \urcorner$ se traduce como $\ulcorner x_i \text{ es miembro de } x_j \urcorner$

$\ulcorner \exists x_i(\dots) \urcorner$ se traduce como $\ulcorner \text{existe un conjunto } x_i \text{ tal que } \dots \urcorner$

Como vimos antes, es posible dar una semántica para \mathcal{L}^E . ¿Será posible dar una semántica análoga para \mathcal{L}^C ? Si uno insiste en utilizar la caracterización simple y una teoría de conjuntos estándar, entonces la respuesta es “no”. Esto puede verificarse con una versión del razonamiento de Orayen: para que fuera posible probar cada instancia del esquema (T) a partir de la caracterización simple, el dominio de \mathcal{L}^C tendría que ser el conjunto de todos los conjuntos, y no hay tal conjunto de acuerdo con las teorías de conjuntos estándar.

Pero no hay ninguna razón para insistir en la caracterización simple, si lo que a uno le interesa es simplemente establecer una asignación composicional de valores de verdad a los enunciados del lenguaje. En el caso de \mathcal{L}^E podríamos utilizar el método siguiente:

LA CARACTERIZACIÓN COMBINADA (caso elefantino)

Una secuencia es una función que asigna a cada variable del lenguaje un elefante.

La secuencia s satisface la fórmula ϕ si y sólo si $\langle \phi, s \rangle$ es miembro de todo conjunto α con las siguientes características:

- $\langle \ulcorner P(x_i, x_j) \urcorner, s \rangle \in \alpha$ si y sólo si $s(\ulcorner x_i \urcorner)$ es amigo de $s(\ulcorner x_j \urcorner)$.
- $\langle \ulcorner x_i = x_j \urcorner, s \rangle \in \alpha$ si y sólo si $s(\ulcorner x_i \urcorner) = s(\ulcorner x_j \urcorner)$.
- $\langle \ulcorner \exists x_i(\psi) \urcorner, s \rangle \in \alpha$ si y sólo si hay una secuencia s' tal que: (a) s' difiere de s a lo más en el objeto que le asigna a $\ulcorner x_i \urcorner$, y (b) $\langle \psi, s' \rangle \in \alpha$.
- $\langle \ulcorner \neg \psi \urcorner, s \rangle \in \alpha$ si y sólo si $\langle \psi, s \rangle \notin \alpha$.
- $\langle \ulcorner \psi \vee \theta \urcorner, s \rangle \in \alpha$ si y sólo si o bien $\langle \psi, s \rangle \in \alpha$ o $\langle \theta, s \rangle \in \alpha$.

ϕ es verdadero si y sólo si toda secuencia satisface ϕ .

No es difícil verificar que la caracterización combinada implica cada instancia del esquema (T).

La caracterización combinada es una semántica para \mathcal{L}^E que no asigna a \mathcal{L}^E un objeto como dominio, y no asigna a “ P ” un objeto como valor semántico. Pero de esto no se sigue que la caracterización combinada no proceda de modo composicional. Tal y como queremos, la caracterización combinada establece valores de verdad para los enunciados de \mathcal{L}^E con base en el significado de sus partes.

Es fácil adaptar la caracterización combinada para el caso de \mathcal{L}^C . Lo único que tenemos que hacer es reemplazar la ocurrencia de “elefante” por “conjunto” y la ocurrencia de “es amigo de” por “ \in ”.

(El resultado es un enunciado que puede formularse en el lenguaje de la teoría de conjuntos sin urelementos.) Dado que en ningún momento se asocia a \mathcal{L}^C un conjunto como dominio, en ningún momento es pertinente la observación de que este dominio tendría que ser el conjunto de todos los conjuntos. Por tanto, el razonamiento de Orayen no puede ser utilizado para determinar si la caracterización combinada es adecuada para \mathcal{L}^C . (De hecho, no es adecuada. Si la teoría de conjuntos que se utiliza es ZFC, una manera rápida de confirmarlo es como sigue: dado que hay al menos tantas secuencias como conjuntos, no hay ningún conjunto que pueda desempeñar el papel de α ; por lo tanto, todo enunciado del lenguaje resulta ser verdadero.)

Esto significa que si la razón por la que queremos formular la interpretación deseada del lenguaje de la teoría de conjuntos es porque nos interesa hacer semántica, entonces no estamos todavía en posición de responder a la pregunta crucial, pues no hemos descartado la posibilidad de que algún método distinto de la asignación de modelos pueda utilizarse para producir una semántica para el lenguaje de la teoría de conjuntos. La caracterización combinada es un ejemplo de un método que el razonamiento de Orayen no descarta, y que no funciona. Pero, ¿no podría haber métodos que el método de Orayen no descartara y que sí funcionaran?

La respuesta depende del tipo de semántica que uno requiera. Si se insiste en una caracterización de verdad para el lenguaje de la teoría de conjuntos que esté formulada en el lenguaje de la teoría de conjuntos, entonces es una consecuencia inmediata del teorema de Tarski que la respuesta es “no”. (Ésta es la observación de fondo en la respuesta que se presenta en la carta de Hart a la estrategia que se ofrece en el artículo de Alchourrón.) Pero si está uno dispuesto a ir más allá del lenguaje de la teoría de conjuntos, entonces la respuesta es “sí”.

He aquí una manera de hacerlo. Comenzamos expandiendo nuestro lenguaje con un nuevo predicado “...satisface...” (que se aplica a una secuencia y una fórmula), y postulamos los siguientes axiomas:

- S1. s satisface $\ulcorner P(x_i, x_j) \urcorner$ si y sólo si $s(\ulcorner x_i \urcorner)$ es miembro de $s(\ulcorner x_j \urcorner)$.
- S2. s satisface $\ulcorner x_i = x_j \urcorner$ si y sólo si $s(\ulcorner x_i \urcorner)$ es idéntico a $s(\ulcorner x_j \urcorner)$.
- S3. s satisface $\ulcorner \exists x_i(\psi) \urcorner$ si y sólo si hay una secuencia s' tal que:
 - (a) s' difiere de s a lo más en el objeto que le asigna a $\ulcorner x_i \urcorner$,
 - (b) el objeto que s le asigna a $\ulcorner x_i \urcorner$ es un conjunto, y
 - (c) s' satisface ψ .

- S4. s satisface $\lceil \neg\psi \rceil$ si y sólo si s no satisface ψ .
- S5. s satisface $\lceil \psi \vee \theta \rceil$ si y sólo si o bien s satisface ψ o bien s satisface θ .

Finalmente, damos la siguiente caracterización de verdad para \mathcal{L}^C :

ϕ es verdadero si y sólo si toda secuencia satisface ϕ .

Con esta caracterización de verdad, S1–S5 implican cada instancia del esquema (T). Por tanto, si lo que nos interesa es simplemente producir una semántica composicional para el lenguaje de la teoría de conjuntos, el razonamiento de Orayen no es un obstáculo para nuestro proyecto.

Estamos, pues, ante la siguiente conclusión. Supongamos que la razón por la que queremos formular la interpretación deseada del lenguaje de la teoría de conjuntos es simplemente que nos interesa hacer semántica. O bien insistimos en formular nuestra semántica en el lenguaje de la teoría de conjuntos, o no. Si lo hacemos, entonces el teorema de Tarski muestra que nuestro proyecto es imposible; si no, entonces la paradoja de Orayen no es un obstáculo para nuestro proyecto. En cualquier caso, el razonamiento de Orayen está de más. Esto sugiere que el proyecto de Orayen no consiste simplemente en dar una semántica para el lenguaje de la teoría de conjuntos. ¿Cuál es el proyecto entonces?

Una posible respuesta es que lo que interesaba a Orayen era dar una caracterización semántica de *consecuencia lógica* para el lenguaje de la teoría de conjuntos. Pero si éste fuera el proyecto, la Paradoja de Orayen no tendría gran alcance. Como señala Gómez Torrente en su contribución al volumen (p. 87), el teorema de completud de Gödel implica que la caracterización usual de consecuencia lógica es adecuada para cualquier lenguaje de primer orden —aun cuando el lenguaje esté interpretado de tal modo que hable de todos los conjuntos, como en el caso de \mathcal{L}^C —. ² (Y basta postular un principio de

² *Prueba:* Para que la caracterización usual de consecuencia lógica sea adecuada basta que se cumplan las siguientes dos condiciones: (a) si hay algún modelo en el que cada enunciado de Γ es verdadero y ϕ es falso, entonces ϕ no es consecuencia lógica de Γ , y (b) si no hay ningún modelo en el que cada enunciado de Γ sea verdadero y ϕ sea falso, entonces ϕ es consecuencia lógica de Γ . Sabemos que (a) se cumple porque si existiera algún modelo en el que cada enunciado de Γ es verdadero y ϕ es falso, sería posible interpretar el vocabulario no lógico de ϕ y Γ de manera que todos los enunciados en Γ resulten verdaderos y ϕ resulte falso, y esto es imposible cuando ϕ es consecuencia lógica de Γ . Sabemos que (b) es verdadero porque si no

reflexión para obtener un resultado similar en el caso de lenguajes de orden superior.)³ Paradoja de Orayen o no, tenemos una buena caracterización semántica de consecuencia lógica para el lenguaje de la teoría de conjuntos.

¿Cuál es entonces el interés de la paradoja? Para responder a esta pregunta es necesario considerar el proyecto de dar una *semántica generalizada* para lenguajes de primer orden, es decir, el proyecto de caracterizar un predicado de verdad-de-acuerdo-con-una-asignación-posible-de-significado. La noción de verdad-en-un-modelo es un caso particular de la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado. A todo modelo le corresponde al menos una asignación posible de significado: una asignación de significado de acuerdo con la cual el lenguaje trata sobre todos y sólo los objetos en el dominio del modelo y de acuerdo con la cual cada predicado no lógico del lenguaje es verdadero de todos y sólo los objetos en el valor semántico que le asigna el modelo. Pero hay asignaciones posibles de significado que no corresponden a ningún modelo. \mathcal{L}^C es un ejemplo porque, como señala Orayen, no hay ningún modelo cuyo dominio consista en todos y sólo los objetos de los que trata \mathcal{L}^C .

Al igual que la noción de verdad-en-un-modelo, la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado puede usarse para dar una caracterización de consecuencia lógica para lenguajes de primer orden. (A grandes rasgos: ϕ es consecuencia de Γ si y sólo si ϕ es verdadero de acuerdo con toda asignación posible de significados tal que: (a) todos los enunciados en Γ son verdaderos, (b) las conectivas lógicas reciben sus significados usuales, (c) el predicado de identidad recibe sus significado usual y (d) $\lceil \exists x_i \dots \rceil$ significa algo de la forma \lceil existe una F , x_i , tal que $\dots \rceil$.)⁴ Aunque las dos caracterizaciones son adecuadas en el sentido de que las dos son satisfechas

existe ningún modelo en el que cada enunciado de Γ sea verdadero y ϕ sea falso, entonces, por el teorema de completud de Gödel, ϕ es consecuencia sintáctica de Γ ; pero un examen de los sistemas deductivos de primer orden revela que cualquier consecuencia sintáctica es una consecuencia lógica.

³ El principio de reflexión de segundo orden dice que cualquier enunciado verdadero en el lenguaje de la teoría de conjuntos de segundo orden es también verdadero en algún modelo. La prueba del resultado no es trivial porque tiene que comenzar con probar que el principio de reflexión implica el principio de Kreisel. Véanse Shapiro 1987, Shapiro 1991, Rayo y Uzquiano 1999, y Rayo 2002.

⁴ Asumo arriba que el lenguaje en cuestión es un lenguaje de primer orden sin funciones ni constantes, y que el cuantificador universal se define en términos del cuantificador existencial de la manera usual. Ciertas modificaciones serían necesarias para otros lenguajes. Para un problema con este tipo de caracterización de consecuencia lógica, véase McGee 1992.

por todos y sólo los argumentos válidos, hay diferencias importantes entre ellas.

Putnam señala una de las diferencias en su carta a Orayen (p. 98). Conforme a la caracterización usual de consecuencia lógica, basada en la noción de verdad-en-un-modelo, el hecho de que un enunciado de \mathcal{L}^C sea lógicamente verdadero no nos da ninguna garantía inmediata de que el enunciado sea verdadero a secas, porque, como muestra el razonamiento de Orayen, no hay ningún modelo que capture el significado de \mathcal{L}^C .⁵ En cambio, cuando caracterizamos consecuencia lógica con base en la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado tenemos una garantía inmediata de que verdad lógica implica verdad a secas. Esto es porque el significado real de un enunciado es una de las asignaciones posibles de significado que la caracterización de consecuencia lógica toma en cuenta. (Dada la asignación real de significado, la noción de verdad-en-una-asignación-posible-de-significado debería poder utilizarse para probar cada instancia del esquema (T).)

Con base en McGee 1992, Gómez Torrente (p. 102) señala otra de las diferencias. Una caracterización de consecuencia lógica basada en la noción de verdad-en-un-modelo es una herramienta que funciona bien en el caso particular de los lenguajes de primer orden, pero puede dejar de funcionar cuando vamos más allá. Si, por ejemplo, agregamos a \mathcal{L}^C una nueva expresión lógica, “ $\exists^C x \dots$ ”, traducida como “las x tales que \dots son tantas como conjuntos hay”, entonces una caracterización de consecuencia lógica basada en la noción de verdad-en-un-modelo llamaría verdades lógicas a enunciados que son, de hecho, falsos (como “ $\neg \exists^C x(x = x)$ ”). Una caracterización de consecuencia lógica basada en la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado, en cambio, puede adaptarse fácilmente a un lenguaje con cuantificadores como “ $\exists^C x \dots$ ”.

En suma: para ciertos propósitos —como el proyecto de dar una caracterización semántica de consecuencia lógica para lenguajes de primer orden— nos basta una aproximación a la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significados; pero no para todos. Una aproximación como la noción de verdad-en-un-modelo es insuficiente cuando lo que queremos es una noción de verdad lógica que nos dé una garantía inmediata de verdad a secas, o cuando queremos una noción de consecuencia lógica que funcione cuando vamos

⁵ Una consecuencia de las observaciones anteriores es que teorema de completud de Gödel nos da una garantía no inmediata.

más allá de los lenguajes de primer orden. Nos enfrentamos, pues, a una versión generalizada de la pregunta crucial:

LA PREGUNTA CRUCIAL (versión generalizada)
 ¿Podríamos hacer semántica generalizada para el lenguaje de la teoría de conjuntos sin apelar a modelos (o, por lo menos, sin apelar al tipo de modelo que da lugar al problema)?

En su contribución al volumen, Orayen ofrece dos soluciones a la paradoja. Cada una puede leerse como una manera de responder a esta pregunta. La primera solución consiste en utilizar *predicados castellanos* para representar el dominio de un lenguaje y los valores semánticos de sus expresiones no lógicas. (Podríamos decir, por ejemplo, que el predicado castellano "... es un conjunto" representa el dominio de \mathcal{L}^C , y que el predicado castellano "... es miembro de ..." representa el valor semántico de "P" en \mathcal{L}^C .) Una semántica generalizada para \mathcal{L}^C puede entonces construirse como sigue. El primer paso consiste en caracterizar una nueva noción de modelo, y una noción generalizada de secuencia:

Un *modelo_{pred}* para \mathcal{L}^C es un par ordenado $\langle D, E \rangle$, donde D es un predicado monádico del castellano (que representa el dominio de \mathcal{L}^C) y E es un predicado diádico del castellano (que representa el valor semántico de 'P').

Una *secuencia* es una función que le asigna a cada par $\langle \langle D, E \rangle, v \rangle$ (donde $\langle D, E \rangle$ es un modelo_{pred} para \mathcal{L}^C y v es una variable de \mathcal{L}^C) un objeto que cae bajo D .

A continuación, introducimos un nuevo predicado *triádico* "...satisface ... con respecto a ..." (que se aplica a una secuencia, una fórmula y un modelo_{pred}), y postulamos los siguientes axiomas:

- S1^G s satisface $\ulcorner P(x_i, x_j) \urcorner$ con respecto a $\langle D, E \rangle$ si y sólo si E es verdadero de $s(\ulcorner x_i \urcorner)$ y $s(\ulcorner x_j \urcorner)$.
- S2^G s satisface $\ulcorner x_i = x_j \urcorner$ con respecto a $\langle D, E \rangle$ si y sólo si $s(\ulcorner x_i \urcorner)$ es idéntico a $s(\ulcorner x_j \urcorner)$.
- S3^G s satisface $\ulcorner \exists x_i(\psi) \urcorner$ con respecto a $\langle D, E \rangle$ si y sólo si hay una secuencia s' tal que: (a) s' difiere de s a lo más en el objeto que le asigna a $\ulcorner x_i \urcorner$, (b) D es verdadero del objeto que s' le asigna a $\ulcorner x_i \urcorner$, y (c) s' satisface ψ con respecto a $\langle D, E \rangle$.

S4^G s satisface $\lceil \neg\psi \rceil$ con respecto a $\langle D, E \rangle$ si y sólo si s no satisface ψ con respecto a $\langle D, E \rangle$.

S5^G s satisface $\lceil \psi \vee \theta \rceil$ con respecto a $\langle D, E \rangle$ si y sólo si o bien s satisface ψ con respecto a $\langle D, E \rangle$, o bien s satisface θ con respecto a $\langle D, E \rangle$.

Finalmente, caracterizamos la noción de verdad-en-un-modelo_{pred}:

ϕ es verdadero en $\langle D, E \rangle$ si y sólo si toda secuencia satisface ϕ con respecto a $\langle D, E \rangle$.

La noción de verdad-en-un-modelo_{pred} tiene una ventaja importante sobre la noción de verdad-en-un-modelo: existe un modelo_{pred} que corresponde al significado real de \mathcal{L}^C ; concretamente, el modelo_{pred} $\langle \dots \text{es un conjunto}, \dots \text{es miembro de } \dots \rangle$. Pero tiene también una desventaja crucial. Dado que no hay más que un número contable de predicados castellanos, hay asignaciones posibles de significado a las que no les corresponde ningún modelo_{pred}. Peor aún, hay modelos a los que no les corresponde ningún modelo_{pred}; es decir, existe un modelo m tal que no existen predicados castellanos D y E tales que D es verdadero de todos y sólo los objetos en el dominio de m y E es verdadero de todas y sólo las parejas de objetos en el valor semántico que m le asigna a “ P ”. Orayen muestra que esta limitación no importa para ciertos propósitos (p. 48); en particular, no importa cuando lo que nos interesa es caracterizar una noción de consecuencia lógica para lenguajes de primer orden. (Se sigue de un teorema en Quine (1986, pp. 53–54) que hay suficientes modelos_{pred} para falsificar cada argumento inválido en un lenguaje de primer orden.) Pero eso no significa que la limitación no sea real. (Éste es un punto que enfatiza Gómez Torrente en su contribución, pp. 91–92.)

De hecho, la limitación no es una particularidad de la noción de verdad-en-un-modelo_{pred}. Imaginemos que alguien caracteriza un nuevo tipo de modelo: digamos, los modelos*. Mientras cada modelo* sea un individuo, es posible probar que hay asignaciones posibles de significado que no corresponden a ningún modelo*. La prueba es como sigue.⁶ Consista en lo que consista ser G , hay una asignación posible de significados para \mathcal{L}^C de acuerdo con la cual el predicado “ P ” es verdadero de la pareja $\langle x, x \rangle$ si y sólo si x es G . Por tanto, si toda asignación posible de significados para \mathcal{L}^C corresponde a algún modelo*, tiene que ser cierto lo siguiente:

⁶ Véanse Williamson 2003, y Rayo y Williamson 2003.

- (1) Hay un modelo* M^G de acuerdo con el cual, para todo individuo x , “ P ” es verdadero de la pareja $\langle x, x \rangle$ si y sólo si x es G .

Pero, dado que los modelos* son individuos, podemos definir un nuevo predicado, “ R ”, como sigue:

- (2) Para todo individuo x , x es R si y sólo si x no es un modelo* de acuerdo con el cual “ P ” es verdadero de la pareja $\langle x, x \rangle$.

Reemplazando “ G ” por “ R ” en (1) y utilizando (2) obtenemos:

- (3) Hay un modelo M^R de acuerdo con el cual, para todo individuo x , “ P ” es verdadero de la pareja $\langle x, x \rangle$ si y sólo si x no es un modelo* de acuerdo con el cual “ P ” es verdadero de la pareja $\langle x, x \rangle$.

Una instancia de (3) es la siguiente:

- (4) Hay un modelo M^R de acuerdo con el cual “ P ” es verdadero de la pareja $\langle M^R, M^R \rangle$ si y sólo si M^R no es un modelo* de acuerdo con el cual “ P ” es verdadero de la pareja $\langle M^R, M^R \rangle$.

Y (4) implica una contradicción.

Cuando Orayen comenzaba a reflexionar sobre estos temas, durante la primera mitad de la década de los ochenta, habría sido razonable pensar que una consecuencia inmediata del resultado anterior es que la versión generalizada de la pregunta crucial debe ser contestada negativamente: el proyecto de dar una semántica generalizada para \mathcal{L}^C es, estrictamente hablando, imposible. Lo máximo a lo que podemos aspirar es a una buena aproximación a la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado. Y la noción de verdad-en-un-modelo_{pred} no es mala aproximación cuando lo que nos interesa es caracterizar una noción adecuada de consecuencia lógica para lenguajes de primer orden con la propiedad adicional de que la verdad a secas de un enunciado sea consecuencia inmediata de su verdad lógica.

Pero mucho ha ocurrido desde la década de los ochenta. En particular, nuestra comprensión de los lenguajes de orden superior se ha transformado radicalmente.⁷ Y teniendo un lenguaje de orden superior a nuestra disposición es posible dar una respuesta bien distinta

⁷ El texto fundamental es Boolos 1984. Véanse también Boolos 1985a, Boolos 1985b, Lewis 1991, McGee 1997, Hossack 2000, McGee 2000, Rayo y Yablo 2001, Oliver y Smiley 2001, Rayo 2002 y Williamson 2003, entre otros. Para propuestas rivales, véanse Quine 1986, cap. 5, Resnik 1988, Parsons 1990 y Linnebo 2003, entre otros.

a la pregunta crucial. La observación básica es la siguiente. Normalmente es un *conjunto* de individuos el que desempeña el papel de dominio en nuestros modelos, pero un lenguaje de orden superior nos permite fijar la atención en ciertos individuos —formen o no un conjunto— y dejar que sean *ellos* quienes juntos cumplan ese papel. De modo similar, un lenguaje de orden superior nos permite rechazar la idea de que sea un *conjunto* de parejas de individuos el que desempeñe el papel del valor semántico de “*P*” en nuestros modelos, y dejar que sean *las parejas mismas* las que juntas cumplan ese papel. Una consecuencia de la prueba que acabamos de considerar es que esto no es posible si se piensa en un modelo como un individuo. Pero un lenguaje de orden superior nos permite fijar la atención en ciertos individuos —formen o no un conjunto— y dejar que sean *ellos* quienes juntos desempeñen el papel de modelo.

Podemos decir, por ejemplo, que las Ms forman un *supermodelo* para \mathcal{L}^C si y sólo si: (a) cada una de las Ms es o bien una pareja de la forma $\langle \text{“dom”}, x \rangle$ o bien una pareja de la forma $\langle \text{“P”}, \langle x, y \rangle \rangle$, y (b) al menos una de las Ms es de la forma $\langle \text{“dom”}, x \rangle$. (Las x tales que $\langle \text{“dom”}, x \rangle$ es una de las Ms cumplen el papel de *dominio* en el supermodelo, y las parejas $\langle x, y \rangle$ tales que $\langle \text{“P”}, \langle x, y \rangle \rangle$ es una de las Ms desempeñan el papel de *valor semántico* de “*P*” en el supermodelo.) Para caracterizar la noción de verdad-en-un-supermodelo comenzamos generalizando la noción de secuencia:

Las Ss forman una *supersecuencia* si y sólo si: (a) cada una de las Ss es una pareja de la forma $\langle v, x \rangle$ donde v es una variable de \mathcal{L}^C , y (b) para cada variable de \mathcal{L}^C hay precisamente una pareja de la forma $\langle v, x \rangle$ entre las Ss.

A continuación, expandimos nuestro lenguaje con un nuevo predicado “...satisfacen ... con respecto a ...”, que se aplica a las Ss, una fórmula y las Ms (donde las Ss forman una supersecuencia y las Ms forman un supermodelo), y postulamos los siguientes axiomas:⁸

S1^S las Ss satisfacen $\ulcorner P(x_i, x_j) \urcorner$ con respecto a las Ms si y sólo si $\langle \text{“P”}, \langle y, z \rangle \rangle$ es una de las Ms, donde $\langle \ulcorner x_i \urcorner, y \rangle$ y $\langle \ulcorner x_j \urcorner, z \rangle$ están entre las Ss.

⁸El nuevo predicado es un predicado de orden superior. Véanse Rayo 2002 y Rayo (en prensa).

- S2^S las Ss satisfacen $\ulcorner x_i = x_j \urcorner$ con respecto a las Ms si y sólo si la y tal que $\langle \ulcorner x_i \urcorner, y \rangle$ es una de las Ss es idéntica a la z tal que $\langle \ulcorner x_j \urcorner, z \rangle$ es una de las Ss.
- S3^S las Ss satisfacen $\ulcorner \exists x_i(\psi) \urcorner$ con respecto a las Ms si y sólo si las S's forman una supersecuencia tal que: (a) las S's difieren de las Ss a lo más en la pareja cuyo primer miembro es $\ulcorner x_i \urcorner$, (b) si $\langle \ulcorner x_i \urcorner, x \rangle$ es una de las S's, entonces $\langle \text{"dom"}, x \rangle$ es una de las Ms, y (c) las S's satisfacen $\ulcorner \psi \urcorner$ con respecto a las Ms.
- S4^S las Ss satisfacen $\ulcorner \neg \psi \urcorner$ con respecto a las Ms si y sólo si las Ss no satisfacen $\ulcorner \psi \urcorner$ con respecto a las Ms.
- S5^S las Ss satisfacen $\ulcorner \psi \vee \theta \urcorner$ con respecto a las Ms si y sólo si o bien las Ss satisfacen $\ulcorner \psi \urcorner$ con respecto a las Ms, o bien las Ss satisfacen $\ulcorner \theta \urcorner$ con respecto a las Ms.

Finalmente, caracterizamos la noción de verdad-en-un-supermodelo:

ϕ es verdadero de acuerdo con las Ms si y sólo si no es el caso que existan individuos —las Ss— tales que las Ss forman una supersecuencia pero las Ss no satisfacen ϕ con respecto a las Ms.

Es fácil verificar que la noción de verdad-en-un-supermodelo captura la noción de verdad-de-acuerdo-con-una-asignación-posible-de-significado, en el sentido siguiente: dada cualquier asignación posible de significado para \mathcal{L}^C hay unas Ms tales que: (a) las Ms forman un supermodelo, (b) x está entre los individuos de los que trata \mathcal{L}^C si y sólo si $\langle \text{"dom"}, x \rangle$ es una de las Ms, y (c) " P " es verdadero de x y y si y sólo si $\langle \text{"P"}, \langle x, y \rangle \rangle$ es una de las Ms. (Y, por supuesto, tenemos también lo inverso: si las Ms forman un supermodelo, entonces hay una asignación posible de significado para \mathcal{L}^C tal que (a) x está entre los individuos de los que trata \mathcal{L}^C si y sólo si $\langle \text{"dom"}, x \rangle$ es una de las Ms, y (b) " P " es verdadero de x y y si y sólo si $\langle \text{"P"}, \langle x, y \rangle \rangle$ es una de las Ms.) Una ventaja de ello es que, cuando la consecuencia lógica se caracteriza como verdad de acuerdo con todo supermodelo, el resultado es una noción que puede adaptarse fácilmente para funcionar más allá de los lenguajes de primer orden, y que, de la verdad a secas de un enunciado, hace la consecuencia inmediata de su verdad lógica.⁹

⁹ Véanse Rayo y Uzquiano 1999 y Rayo y Williamson 2003.

La posibilidad de hacer semántica generalizada para lenguajes de primer orden utilizando lenguajes de orden superior no es un resultado trivial, aunque tampoco es una victoria absoluta. El problema es que para hacer semántica generalizada para un lenguaje de orden n es necesario utilizar recursos de al menos orden $n + 1$.¹⁰ Esto genera una jerarquía infinita de nociones de verdad-de-acuerdo-con-una-asignación-posible-de-significados. El enésimo miembro de la jerarquía nos permite describir lenguajes de orden n , pero está formulado en un lenguaje de al menos orden $n + 1$. Nunca alcanzamos, pues, una perspectiva absoluta: una noción de verdad-de-acuerdo-con-una-asignación-posible-de-significados apropiada para lenguajes de cualquier orden.

La segunda de las soluciones que propone Orayen para la paradoja también está basada en una jerarquía, como deja en claro García de la Sienna (pp. 76–79); pero en lugar de tratarse de una jerarquía ideológica como la que acabamos de discutir, se trata de una jerarquía ontológica. En ningún momento se apela a lenguajes de orden superior. La idea es que el universo de conjuntos (o, si se prefiere, el universo de objetos tipo-conjunto) está estratificado en niveles, y sólo debemos estudiar asignaciones de significado conforme a las cuales \mathcal{L}^C trata de conjuntos que están por debajo de algún nivel en la estratificación. Podemos, sin embargo, proceder en etapas. En una primera etapa nos limitamos a asignaciones de significado de acuerdo con las cuales \mathcal{L}^C trata de conjuntos de “nivel” 0. En la segunda etapa nos limitamos a asignaciones de acuerdo con las cuales \mathcal{L}^C trata de conjuntos de “nivel” 0 o 1, y así sucesivamente. Pero, una vez más, el resultado es que nunca alcanzamos una perspectiva absoluta: una noción de verdad-de-acuerdo-con-una-asignación-posible-de-significados que abarque todas las asignaciones posibles de significado.¹¹ (La confluencia de resultados no es casual: la jerarquía ontológica puede verse como una reificación de la jerarquía ideológica.)

El teorema de Tarski nos enseña que tenemos que aprender a vivir con un cierto tipo de desilusión: es imposible hacer semántica para un lenguaje (suficientemente expresivo) sin utilizar un lenguaje que sea de mayor riqueza expresiva. La paradoja de Orayen sugiere que tenemos que aprender a vivir con un nuevo tipo de desilusión: es imposible hacer semántica generalizada para un lenguaje sin utilizar

¹⁰ Para una formulación más precisa de este resultado y su prueba, véase Rayo (en prensa). Es posible probar una versión del resultado para órdenes transfinitos.

¹¹ Para una propuesta similar, véanse Parsons 1974 y Parsons (en prensa).

un lenguaje que sea no sólo sea de mayor riqueza expresiva, sino también de mayor riqueza lógica u ontológica.¹²

REFERENCES

- Alchourrón, C., 2003, “Sobre la adecuación filosófica de las teorías de conjuntos”, en Moretti y Hurtado 2003, pp. 61–69.
- Beall, J., (comp.), 2003, *Liars and Heaps*, Oxford University Press, Oxford.
- Boolos, G., 1998, *Logic, Logic and Logic*, Harvard, Cambridge, Mass.
- , 1985a, “Nominalist Platonism”, *Philosophical Review*, vol. 94, pp. 327–344. (Reimpreso en Boolos 1998).
- , 1985b, “Reading the *Begriffsschrift*”, *Mind*, vol. 94, pp. 331–334. (Reimpreso en Boolos 1998).
- , 1984, “To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables)”, *The Journal of Philosophy*, vol. 81, pp. 430–449. (Reimpreso en Boolos 1998.)
- Forster, T.E., 1995, *Set Theory with a Universal Set: Exploring an Untyped Universe*, *Oxford Logic Guides*, vol. 31, 2a. ed., Oxford Science Publications, Oxford.
- García de la Sienra, A., 2003, “La Paradoja de Orayen”, en Moretti y Hurtado 2003, pp. 71–81.
- Gómez Torrente, M., 2003, “Notas sobre la Paradoja de Orayen”, en Moretti y Hurtado 2003, pp. 83–94.
- Hart, W., 2003, “Carta de William Hart”, en Moretti y Hurtado 2003, pp. 99–100.
- Hossack, K., 2000, “Plurals and Complexes”, *The British Journal for the Philosophy of Science*, vol. 51, no. 3, pp. 411–443.
- Lewis, D., 1991, *Parts of Classes*, Blackwell, Oxford.
- Linnebo, Ø., 2003, “Plural Quantification Exposed”, *Noûs*, vol. 37, pp. 71–92.
- McGee, V., 2000, “‘Everything’”, en Sher y Tieszen 2000.
- , 1997, “How We Learn Mathematical Language”, *Philosophical Review*, vol. 106, pp. 35–68.
- , 1992, “Two Problems with Tarski’s Theory of Consequence”, *Proceedings of the Aristotelian Society*, vol. 92, pp. 273–292.
- Moretti, A. y G. Hurtado (comps.), 2003, *La Paradoja de Orayen*, Eudeba, Buenos Aires.
- Oliver, A., y T. Smiley, 2001, “Strategies for a Logic of Plurals”, *Philosophical Quarterly*, vol. 51, pp. 289–306.
- Orayen, R., 2003, “Una paradoja en la semántica de la teoría de conjuntos”, en Moretti y Hurtado 2003, pp. 35–59.

¹² Quisiera agradecer a Atocha Aliseda, Maite Ezcurdia y Mario Gómez Torrente por sus valiosos comentarios, y al equipo editorial de *Crítica* por su paciencia con mis idiosincrasias computacionales.

- Parsons, C., 1990, "The Structuralist View of Mathematical Objects", *Synthese*, vol. 84, pp. 303–346.
- , 1983, *Mathematics in Philosophy*, Cornell University Press, Ithaca.
- , 1974, "Sets and Classes", *Noûs*, vol. 8, pp. 1–12. (Reimpreso en Parsons 1983.)
- (en prensa), "The Problem of Absolute Universality", en Rayo y Uzquiano (en prensa).
- Putnam, H., 2003, "Carta de Putnam", en Moretti y Hurtado 2003, pp. 97–98.
- Quine, W.V.O., 1986, *Philosophy of Logic*, 2a. ed., Harvard University Press, Cambridge, Mass.
- Rayo, A., 2002, "World and Objects", *Noûs*, vol. 36, pp. 436–464.
- (en prensa), "Beyond Plurals", en Rayo y Uzquiano (en prensa).
- Rayo, A. y G. Uzquiano, 1999, "Toward a Theory of Second-Order Consequence", *The Notre Dame Journal of Formal Logic*, vol. 40, pp. 315–325.
- (comps.) (en prensa), *Absolute Generality*, Oxford University Press, Oxford.
- Rayo, A. y T. Williamson, 2003, "A Completeness Theorem for Unrestricted First-Order Languages", en Beall 2003.
- Rayo, A. y S. Yablo, 2001, "Nominalism Through De-Nominalization", *Noûs*, vol. 35, no. 1, pp. 74–92.
- Resnik, M., 1988, "Second-Order Logic Still Wild", *Journal of Philosophy*, vol. 85, no. 2, pp. 75–87.
- Shapiro, S., 1991, *Foundations without Foundationalism: A Case for Second-Order Logic*, Clarendon Press, Oxford.
- , 1987, "Principles of Reflection and Second-Order Logic", *Journal of Philosophical Logic*, vol. 16, pp. 309–333.
- Sher, G. y R. Tieszen (comps.), 2000, *Between Logic and Intuition*, Cambridge University Press, Nueva York/Cambridge.
- Tarski, A., 1929, "The Concept of Truth in Formalized Languages", *Logic, Semantics and Meta-Mathematics*, 2a. ed., Hackett, Indiana (1983), pp. 152–278.
- Williamson, T., 2003, "Everything", en J. Hawthorne y D. Zimmerman (comps.), *Philosophical Perspectives 17: Language and Philosophical Linguistics*, Blackwell, Oxford, pp. 415–465.

Recibido el 19 de octubre de 2004; aceptado el 1 de diciembre de 2004.