Verdad y Demostración*

ALFRED TARSKI

EL TEMA DE ESTE ARTÍCULO NO ES NUEVO. Ha sido discutido frecuentemente en la literatura lógica y filosófica moderna, y no sería fácil contribuir con algo original a la discusión. Me temo que a muchos lectores ninguna de las ideas presentadas en este artículo les resultará esencialmente novedosa; sin embargo, espero que encuentren interesante la manera en que se ha dispuesto y organizado el material.

Como el título lo indica, deseo discutir aquí dos nociones que, aunque son diferentes, están relacionadas: la noción de verdad y la noción de demostración. De hecho, el artículo está dividido en tres secciones. La primera sección se ocupa exclusivamente de la noción de verdad, la segunda trata primariamente de la noción de demostración, y la tercera es una discusión acerca de la relación entre estas dos nociones.

§1. La noción de verdad

La tarea de explicar el significado del término «verdad(ero) —*true*—» será interpretada aquí de una manera restringida. La noción de verdad aparece en muchos contextos diferentes, y hay varias categorías distintas de objetos a las cuales se aplica el término «verdad(ero)». En una discusión psicológica podría hablarse tanto de emociones verdaderas como de creencias verdaderas; en un discurso que pertenezca al dominio de la estética podría analizarse la verdad interna de un objeto de arte. En este artículo, sin embargo, sólo estamos interesados en lo que podría llamarse la noción lógica de verdad. Más específicamente, nos ocuparemos exclusivamente del significado del término «verdad(ero)» cuando este término es usado para referirse a oraciones. Presumiblemente, este era el uso original del término «verdad(ero)» en el lenguaje humano. Las oraciones son tratadas aquí como objetos lingüísticos,

* Publicado originalmente como: Tarski, Alfred. «Truth and Proof». *Scientific American* 220/6 (1969): pp. 63-77.

A. Tarski (✉)
University of California, Berkeley, EUA

Disputatio. Philosophical Research Bulletin
Vol. 4, No. 5, Dic. 2015, pp. 367-396
ISSN: 2254-0601 | www.disputatio.eu

DOCUMENTO

como determinadas secuencias de sonidos o signos escritos (por supuesto, no toda secuencia de este tipo es una oración). Además, cuando hablemos de oraciones, estaremos pensando en las que en gramática se llaman oraciones declarativas, y no en las oraciones interrogativas o imperativas.

Siempre que se explica el significado de cualquier término del lenguaje cotidiano se debe tener en cuenta que el objetivo y el estatus lógico de esa explicación pueden variar de un caso a otro. Por ejemplo, la explicación puede proponerse como una aclaración del uso efectivo del término en cuestión, y se puede preguntar, por lo tanto, si la aclaración es en verdad correcta. En otra ocasión, una explicación puede ser de naturaleza normativa, esto es, puede ofrecerse como una sugerencia de que el término sea usado de alguna manera definida, sin pretender que la sugerencia se adapte a la manera en que el término es efectivamente usado; una explicación de este tipo puede ser evaluada, por ejemplo, desde el punto de vista de su utilidad, mas no de su corrección. Podrían enumerarse también algunas alternativas más.

La explicación que deseamos ofrecer en este caso es, en alguna medida, de un carácter mixto. Lo que se ofrecerá puede, en principio, ser tratado como una sugerencia para una manera definida de usar el término «verdad(ero)», pero esta propuesta estará acompañada por la creencia de que está de acuerdo con el uso predominante de este término en el lenguaje cotidiano.

Nuestra comprensión de la noción de verdad parece estar esencialmente de acuerdo con varias explicaciones que de esta noción se han dado en la literatura filosófica. La que quizás sea la primera explicación se puede encontrar en la *Metafísica* de Aristóteles:

> Decir de lo que es que no es, o de lo que no es que es, es falso, mientras que decir de lo que es que es, o de lo que no es que no es, es verdadero.

Aquí y en la discusión subsiguiente la palabra «falso» significa lo mismo que la expresión «no verdadero», y puede ser reemplazada por ésta última.

El contenido intuitivo de la formulación de Aristóteles parece ser bastante claro. Sin embargo, la formulación deja mucho que desear desde el punto de vista de la precisión y la corrección formal. Por ejemplo, no es suficientemente general; se refiere sólo a oraciones que «dicen» de algo «que es» o«que no es»: en la mayoría de los casos sería difícilmente posible poner a una oración en este molde sin sesgar el sentido de la oración y forzar el espíritu del lenguaje. Esta es quizás una de las razones por las que en la filosofía moderna se han ofrecido diversos substitutos de la formulación aristotélica. Como ejemplos citaremos los siguientes:

> Una oración es verdadera si denota el estado de cosas existente.

> La verdad de una oración consiste en su conformidad (o correspondencia) con la realidad.

Estas formulaciones tienen, sin duda, un tono muy «académico» debido al uso de términos filosóficos técnicos. Sin embargo, me parece que, cuando se analizan más detalladamente, las nuevas formulaciones resultan menos claras e inequívocas que la propuesta por Aristóteles.

La concepción de la verdad que encontró su expresión en Aristóteles (y en formulaciones de origen más reciente relacionadas con ella) se suele llamar la *concepción clásica o semántica de la verdad*. Por semántica entendemos la parte de la lógica que, hablando sin demasiada precisión, se ocupa de las relaciones entre los objetos lingüísticos (como las oraciones) y lo que estos objetos expresan. El carácter semántico del término «verdad(ero)» queda claramente revelado por la explicación ofrecida por Aristóteles, y por algunas formulaciones que se ofrecerán más adelante en este artículo. Se habla a veces de la teoría de la verdad como correspondencia como la teoría que está basada en la concepción clásica.

(En la literatura filosófica moderna se discuten algunas otras concepciones y teorías de la verdad, tales como la concepción pragmática y la teoría de la coherencia. Estas concepciones parecen ser de un carácter exclusivamente normativo y tener poca conexión con el uso efectivo del término «verdad(ero)»; ninguna de ellas ha sido, hasta ahora, formulada con algún grado de claridad y precisión. No se las discutirá en este artículo).

Aquí trataremos de obtener una explicación más precisa de la concepción clásica de la verdad, que pueda reemplazar a la formulación aristotélica, preservando sus intenciones básicas. Para ello recurriremos a algunas técnicas de la lógica contemporánea. También tendremos que especificar el lenguaje de cuyas oraciones nos ocuparemos; esto es necesario aunque más no sea porque una secuencia de sonidos o signos que, verdadera o falsa, sea una oración significativa en un lenguaje, puede no ser significativa en otro lenguaje. Por el momento, supongamos que el lenguaje del que nos ocupamos es el castellano.

Comencemos con un problema simple. Considérese una oración castellana cuyo significado no ofrece ninguna duda, por ejemplo la oración «la nieve es blanca». Por razones de brevedad denotaremos a esta oración con «*S*», de manera que «*S*» se convierte en el nombre de la oración. Nos preguntamos: ¿qué significa decir que *S* es verdadera o que es falsa? La respuesta a esta pregunta es simple: en el espíritu de la explicación aristotélica, cuando decimos que *S* es verdadera queremos decir simplemente que la nieve es blanca, y cuando decimos que *S* es falsa queremos decir que la nieve no es blanca. Al eliminar el símbolo «*S*» llegamos a las siguientes formulaciones:

(1) «la nieve es blanca» es verdadera si y sólo si la nieve es blanca.

(1') «la nieve es blanca» es falsa si y sólo si la nieve no es blanca.

De esta manera, (1) y (1') proporcionan explicaciones satisfactorias del significado de los términos «verdad(ero)» y «falso» cuando estos términos se refieren a la oración «la nieve es blanca». Podemos considerar a (1) y a (1') como definiciones parciales de los términos «verdad(ero)» y «falso»; de hecho, como definiciones de estos términos respecto de una oración en particular. Obsérvese que (1), y (1'), tiene la forma prescrita para las definiciones por las reglas de la lógica, la forma de la equivalencia lógica. Está compuesta de dos partes, los lados izquierdo y derecho de la equivalencia, combinados por la conectiva «si y sólo si». El lado izquierdo es el definiendum, la frase cuyo significado es explicado por la definición; el lado derecho es el definiens, la frase que proporciona la explicación. En este caso el definiendum es la siguiente expresión:

«la nieve es blanca» es verdadera;

el definiens tiene la forma:

la nieve es blanca.

A primera vista podría parecer que (1), considerada como una definición, exhibe un defecto esencial que ha sido ampliamente discutido en la lógica tradicional bajo el nombre de círculo vicioso. La razón es que ciertas palabras, «nieve» por ejemplo, aparecen tanto en el definiens como en el definiendum. Sin embargo, en realidad, estas apariciones tienen un carácter enteramente diferente. La palabra «nieve» es una parte sintáctica, u orgánica, del definiens; de hecho, el definiens es una oración, y la palabra «nieve» es su sujeto. El definiendum también es una oración; expresa el hecho de que el definiens es una oración verdadera. Su sujeto es un nombre del definiens que se forma poniendo el definiens entre comillas. (Cuando decimos algo de un objeto, siempre usamos el nombre de este objeto y no el objeto mismo, aun cuando nos estemos ocupando de objetos lingüísticos). Por varias razones una expresión encerrada entre comillas debe ser tratada gramaticalmente como una única palabra que no tiene partes sintácticas. Por consiguiente, la palabra «nieve» que indudablemente aparece como una parte del definiendum, no aparece allí como una parte sintáctica. Un lógico medieval diría que «nieve» aparece en el definiens *in suppositione formalis* y en el definiendum *in suppositione materialis.* Sin embargo, las palabras que no son partes sintácticas del definiendum no pueden crear un círculo vicioso, y el peligro de un círculo vicioso desaparece.

Las observaciones precedentes tocan algunas cuestiones bastante sutiles y no demasiado simples desde el punto de vista lógico. En lugar de profundizar en ellas, indicaré otra manera en que es posible disipar cualquier temor de círculo vicioso. Al formular (1) hemos aplicado un método común para formar el nombre de una oración, o de cualquier otra expresión, que consiste en poner la expresión entre comillas. Este método tiene muchas virtudes, pero es también la fuente de las dificultades ya discutidas. Para eliminar estas dificultades probemos otro método de formar nombres de expresiones, un método que, de hecho, puede ser caracterizado como una descripción letra por letra de una expresión. Usando este método obtenemos, en lugar de (1), la siguiente extensa formulación:

(2) La secuencia de cuatro palabras, la primera de las cuales es la secuencia de las letras Ele y A, la segunda es la secuencia de las letras Ene, I, E, Uve y E, la tercera es la secuencia de las letras E y Ese, y la cuarta es la secuencia de las letras Be, Ele, A, Ene, Ce y A, es una oración verdadera si y sólo si la nieve es blanca.

La formulación (2) no difiere de la (1) en su significado: (1) puede simplemente ser considerada como una forma abreviada de (2). La nueva formulación es ciertamente mucho menos perspicua que la anterior, pero tiene la ventaja de que no crea la apariencia de un círculo vicioso.

Pueden construirse definiciones parciales de verdad análogas a (1) (o a (2)) para otras oraciones también. Cada una de estas definiciones tiene la forma:

(3) «*p*» es verdadera si y sólo si *p*,

donde «*p*» debe ser reemplazada en ambos lados de (3) por la oración para la cual se construye la definición. Sin embargo, se debe prestar especial atención a aquellas situaciones en las que la oración que se pone en lugar de «*p*» contiene la palabra «verdad(ero)» como parte sintáctica. Entonces la correspondiente equivalencia (3) no puede ser considerada como una definición parcial de verdad dado que, cuando se la trata de esa manera obviamente exhibe un circulo vicioso. Aun en este caso, sin embargo, (3) es una oración significativa y de hecho es una oración verdadera desde el punto de vista de la concepción clásica de la verdad. Como ilustración, imagínese que en una reseña de un libro se encuentra la siguiente oración:

(4) No todas las oraciones de este libro son verdaderas.

Aplicando el criterio aristotélico a (4) vemos que la oración (4) es verdadera si, de hecho, no todas las oraciones del libro en cuestión son verdaderas, y que si no es así (4) es falsa; en otras palabras, podemos afirmar la equivalencia obtenida a partir de (3) tomando a (4) como «*p*». Por supuesto esta equivalencia afirma meramente las condiciones bajo las cuales la oración (4) es verdadera o no es verdadera, pero por sí misma la equivalencia no nos permite decidir cuál es el caso. Para verificar el juicio expresado por (4) se debería leer atentamente el libro reseñado y analizar la verdad de las oraciones contenidas en él.

A la luz de la discusión precedente podemos ahora reformular nuestro problema principal. Estipulamos que el uso del término «verdad(ero)» cuando se refiere a las oraciones del castellano está de acuerdo con la concepción clásica de la verdad sólo cuando nos permite aseverar todas las equivalencias de la forma (3) en las que «*p*» es reemplazado en ambos lados por una oración castellana arbitraria. Si esta condición es satisfecha diremos simplemente que el uso del termino «verdad(ero)» es adecuado. De manera que nuestro problema principal es: ¿podemos establecer un uso adecuado del termino «verdad(ero)» para las oraciones del castellano y, si es así, con qué métodos? Podemos por supuesto plantear una pregunta análoga para las oraciones de cualquier otro lenguaje.

El problema será resuelto completamente si logramos construir una definición general de verdad que sea adecuada, en el sentido de que implique como consecuencias lógicas todas las equivalencias de la forma (3). Si tal definición es aceptada por los castellanoparlantes obviamente establecerá un uso adecuado del término «verdad(ero)».

Bajo ciertos supuestos especiales la construcción de una definición general de verdad es fácil. Supóngase, en efecto que no estamos interesados en la totalidad del lenguaje castellano común sino sólo en un fragmento suyo y que queremos definir el término «verdad(ero)» exclusivamente en referencia a las oraciones de este lenguaje fragmentario; nos referiremos a este lenguaje fragmentario como el lenguaje *L*. Supongamos además que *L* tiene unas reglas sintácticas precisas que nos permiten, en cada caso particular, distinguir una oración de una expresión que no es una oración, y que el número de todas las oraciones del lenguaje *L* es finito (aunque posiblemente sea muy grande). Supóngase finalmente que la palabra verdadero) no aparece en *L* y que el significado de todas las palabras de *L* es suficientemente claro de manera que no tengamos objeciones para usarlas al definir verdad. Bajo estos supuestos procédase como sigue. Primero prepare una lista completa de todas las oraciones de *L*; supóngase por ejemplo que hay exactamente 1.000 oraciones en *L* y acuérdese utilizar los símbolos «s_1», «s_2», ..., «$s_{1.000}$» como abreviatura de las oraciones consecutivas de la lista. Después, para cada una de las oraciones «s_1», «s_2», ..., «$s_{1.000}$» construya una definición parcial de verdad substituyendo «*p*» por esas oraciones en ambos lados del esquema (3). Finalmente forme la conjunción lógica de estas definiciones parciales; en otras palabras combínelas en un enunciado poniendo la conectiva «y» entre dos definiciones parciales cualesquiera. La única cosa que falta por hacer es dar a la conjunción resultante una forma diferente, aunque lógicamente equivalente, de manera de satisfacer los requerimientos formales impuestos a las definiciones por las reglas de la lógica:

(5) Para cada oración x (del lenguaje L) x es verdadera si y sólo si

o

s_1, y x es identica a «s_1»,

o

s_2, y x es identica a «s_2»,

....

o finalmente,

$s_{1.000}$, y x es identica a «$s_{1.000}$».

Hemos llegado así a un enunciado que puede ciertamente ser aceptado como la definición general de verdad que deseábamos: es formalmente correcta y es adecuada en el sentido que implica todas las equivalencias de la forma (3) en las que «p» ha sido reemplazada por cualquier oración del lenguaje L. Hacemos notar al pasar que (5) es una oración castellana pero que obviamente no pertenece al lenguaje L; dado que (5) contiene todas las oraciones de L como partes propias no puede coincidir con ninguna de ellas. Una discusión más profunda aclarará mejor este punto.

Por razones obvias el procedimiento que acabamos de esbozar no puede seguirse si estamos interesados en la totalidad del lenguaje castellano y no meramente en un fragmento suyo. Cuando tratamos de preparar una lista completa de las oraciones castellanas, nos encontramos desde el comienzo con la dificultad de que las reglas de la gramática castellana no determinan precisamente la forma de las expresiones (secuencias de palabras) que deben ser consideradas como oraciones; una expresión particular, una exclamación por ejemplo, puede funcionar como una oración en algún contexto dado, mientras que una expresión de la misma forma no funcionará de ese modo en algún otro contexto. Además, el conjunto de todas las oraciones del castellano es, potencialmente al menos, infinito. A pesar de que es ciertamente verdadero que los seres humanos han formulado sólo un numero finito de oraciones en el habla y la escritura hasta el momento actual, probablemente nadie estaría de acuerdo con que la lista de todas estas oracíones comprende todas las oraciones del castellano. Por el contrario, parece probable que al ver dicha lista cada uno de nosotros pueda fácilmente producir una oración castellana que no está en la lista. Finalmente, el hecho de que la palabra «verdad(ero)» aparezca en el castellano impide por sí mismo la aplicación del procedimiento anteriormente descripto.

No se sigue de estas observaciones que la definición de verdad que buscamos para oraciones arbitrarias del castellano no pueda ser obtenida de alguna otra manera, posiblemente mediante el uso de una idea diferente. Hay, sin embargo, una razón más seria y fundamental que parece impedir esta posibilidad. Lo que es más, la mera suposición de que, mediante un método cualquiera, se ha obtenido un uso adecuado del término «verdad(ero)» (en su referencia a oraciones arbitrarias del castellano) parece llevar a una contradicción. El argumento más simple que proporciona tal contradicción es conocido como la *antinomia del mentiroso* y será desarrollado en las próximas líneas.

Considérese la siguiente oración:

(6) La oración impresa en rojo en «Verdad y Demostración» es falsa.

Acordemos usar «*s*» como abreviatura de esta oración. Mirando «Verdad y Demostración» verificamos fácilmente que «*s*» es la única oración impresa en rojo en «Verdad y Demostración». De esto se sigue, en particular, que

(7) «*s*» es falsa si y sólo si la oración impresa en rojo en «Verdad y Demostración» es falsa.

Por otra parte, «*s*» es indudablemente una oración del castellano. Por lo tanto, suponiendo que nuestro uso del término «verdad(ero)» sea adecuado, podemos afirmar la equivalencia (3) en la cual «*p*» es reemplazada por «*s*». Así podemos aseverar:

(8) «s» es verdadera si y sólo si *s*.

Ahora recordamos que «*s*» ocupa el lugar de la oración entera (6). Por lo tanto, podemos reemplazar «*s*» por 6 en el lado derecho de (8); obtenemos entonces

(9) «*s*» es verdadera si y sólo si la oración impresa en rojo en «Verdad y Demostración» es falsa.

Comparando ahora (8) y (9), concluimos:

(10) «*s*» es falsa si y sólo si «*s*» es verdadera.

Esto lleva a una obvia contradicción: «*s*» resulta ser al mismo tiempo verdadera y falsa. Nos vemos enfrentados así con una antinomia. La formulación precedente de la antinomia del mentiroso se debe al lógico polaco Jan Łukasiewicz.

Se conocen también algunas formulaciones más complicadas de esta antinomia. Imagínese, por ejemplo, un libro de 100 páginas, con sólo una oración impresa en cada página. En la página 1 leemos:

La oración impresa en la página 2 de este libro es verdadera.

En la página 2 leemos:

La oración impresa en la página 3 de este libro es verdadera.

Y así siguiendo hasta la página 99. Sin embargo, en la página 100, la última página del libro, encontramos:

La oración impresa en la página 1 de este libro es falsa.

Supóngase que la oración impresa en la página 1 es, en efecto, falsa. Por medio de un argumento que no es difícil, pero que es muy largo y requiere hojear el libro entero, concluimos que nuestro supuesto era incorrecto. En consecuencia, ahora suponemos que la oración impresa en la página 1 es verdadera —y, mediante un argumento que es tan fácil y tan largo como el original, nos convencemos de que el nuevo supuesto también es falso. De manera que nos vemos nuevamente enfrentados con una antinomia.

Componer muchos otros «libros antinómicos», variantes del que acabamos de componer, resulta ser una cuestión fácil. Cada uno de ellos tiene 100 páginas. Cada página contiene sólo una oración, y, de hecho, una oración de la forma:

La oración impresa en la página *00* de este libro es *XX*.

En cada caso particular. «*XX*» es reemplazado por las palabras «*verdadero*» o «*falso*», mientras que «*00*» es reemplazado por uno de los numerales «1», «2», ..., «100»; el mismo numeral puede aparecer en varias páginas. No toda variante del libro original que se componga de acuerdo a estas reglas produce efectivamente una antinomia. Al lector al que le gusten los enigmas lógicos no le será nada difícil describir todas las variantes que la producen. La siguiente advertencia puede resultar útil a este respecto. Imagínese que en algún lugar del libro, por ejemplo en la página 1, se dice que la oración de la página 3 es verdadera, mientras que en algún otro lugar, por ejemplo en la página 2, se afirma que la misma oración es falsa. De esta información no se sigue en absoluto que nuestro libro sea «antinómico»; sólo podemos concluir que, o bien la oración de la página 1, o bien la oración de la página 2, deben ser falsas. La antinomia surge, sin embargo, cada vez que somos capaces de mostrar que una de las oraciones del libro es tanto verdadera como falsa, independientemente de la verdad o falsedad de las restantes oraciones.

La antinomia del mentiroso tiene un origen muy antiguo. Se la suele atribuir al lógico griego Eubúlides; atormentó a muchos lógicos antiguos y causó la muerte prematura de al menos uno de ellos, Filetas de Cos. Varias otras antinomias y paradojas fueron descubiertas en la Antigüedad, en la Edad Media, y en los tiempos modernos. A pesar de que muchas de ellas ya han sido completamente olvidadas, la antinomia del mentiroso todavía es analizada y discutida en obras contemporáneas. Ha tenido gran impacto en el desarrollo de la lógica moderna, junto con algunas antinomias recientes que fueron descubiertas alrededor de principios de siglo (en particular, la antinomia de Russell).

En la literatura sobre este tema es posible encontrar dos posiciones diametralmente opuestas acerca de las paradojas. Una posición consiste en desentenderse de ellas, en tratarlas como sofisterías, como bromas que no son serias sino maliciosas, y que tienen como finalidad principal mostrar el ingenio del hombre que las formula. La posición opuesta es característica de ciertos pensadores del siglo XIX y tiene representantes, o los tenía hace poco, en ciertas partes del globo. De acuerdo a esta posición, las antinomias constituyen un elemento muy esencial del pensamiento humano; deben aparecer una y otra vez en las actividades intelectuales, y su presencia es la fuente básica del progreso real. Como sucede a menudo, la verdad probablemente esté entre ambas posiciones. Personalmente, como lógico, no podría resignarme a tener a las antinomias como un elemento permanente de nuestro sistema de conocimiento. Con todo, no tengo la menor inclinación de tratar a las antinomias con ligereza. La aparición de una antinomia es para mí un síntoma de enfermedad. Una antinomia nos lleva a un contrasentido, a una contradicción, partiendo de premisas que parecen intuitivamente obvias, usando formas de razonamiento que parecen intuitivamente seguras. Cuando esto sucede, tenemos que someter nuestro modo de pensar a una revisión a fondo, rechazando algunas premisas en las que creíamos o mejorando algunas formas de argumento que usábamos. Hacemos esto con la esperanza no sólo de que la vieja antinomia sea eliminada, sino también de que ninguna nueva aparezca. Para ello, probamos nuestro sistema de pensamiento reformado por todos los medios disponibles, y, ante todo, tratamos de reconstruir la vieja antinomia en el nuevo marco; estas pruebas constituyen una actividad muy importante en el dominio del pensamiento especulativo, semejante a la realización de experimentos cruciales en la ciencia empírica.

Consideremos ahora específicamente la antinomia del mentiroso desde este punto de vista. La antinomia involucra la noción de verdad en referencia a oraciones arbitrarias del castellano común; podría fácilmente reformularse para aplicarse a otros lenguajes naturales. Nos enfrentamos a un problema serio: ¿cómo pueden evitarse las contradicciones inducidas por esta antinomia? Fácilmente se nos podría ocurrir una solución radical al problema consistente simplemente en eliminar la palabra «verdad(ero)» del vocabularrio castellano o, al menos, en abstenerse de usarla en cualquier discusión seria.

Las personas a las cuales tal amputación les parece altamente insatisfactoria e ilegítima podrían preferir una solución de compromiso, que consiste en adoptar lo que podría llamarse (siguiendo al filósofo polaco Tadeusz Kotarbiński) «la concepción nihilista de la teoría de la verdad». De acuerdo con este enfoque, la palabra «verdad(ero)» no tiene ningún significado independiente, pero puede ser usada como componente de las dos expresiones significativas «es verdad que» y «no es verdad que». De esta manera, estas expresiones son tratadas como si fueran una sola palabra sin partes orgánicas. El significado que se les atribuye es tal que pueden ser inmediatamente eliminadas de cualquier oración en la que aparezcan. Por ejemplo, en lugar de decir

es verdad que todos los gatos son negros,

podemos simplemente decir

todos los gatos son negros,

y en lugar de

no es verdad que todos los gatos sean negros

podemos decir

no todos los gatos son negros.

En otros contextos la palabra «verdad(ero)» no tiene sentido. En particular, no puede ser usada como un verdadero predicado que califique a nombres de oraciones. Empleando la terminología de la lógica medieval, podemos decir que la palabra «verdad(ero)» puede ser usada sincategoremáticamente en algunas situaciones especiales, pero no puede nunca ser usada categoremáticamente.

Para darse cuenta de las implicaciones de este enfoque, considérense las oraciones que fueron el punto de partida para la antinomia del mentiroso; esto es, la oración impresa en rojo en este artículo. Desde el punto de vista «nihilista» esta no es una oración significativa, y la antinomia simplemente se desvanece. Desafortunadamente, muchos usos de la palabra «verdad(ero)», que de otro modo parecen bastante legítimos y razonables, se ven similarmente afectados por este enfoque. Imagínese, por ejemplo, que un cierto término que aparece repetidamente en las obras de un matemático antiguo admite varias interpretaciones. Un historiador de la ciencia que estudia sus obras llega a la conclusión de que bajo una de esas interpretaciones todos los teoremas enunciados por el matemático resultan verdaderos; esto le lleva naturalmente a la conjetura de que lo mismo sucederá con cualquiera de las obras de este matemático que no sea conocida actualmente, pero que pueda ser descubierta en el futuro. Sin embargo, el historiador de la ciencia comparte la concepción «nihilista» de la noción de verdad y no tiene la posibilidad de expresar en palabras su conjetura. Uno podría decir que la teoría «nihilista» de la verdad rinde una supuesta pleitesía a algunas formas populares del habla humana, mientras que en realidad elimina la noción de verdad del repertorio conceptual de la mente humana.

Buscaremos, por lo tanto, otra salida a nuestra encrucijada. Trataremos de buscar una solución que mantenga el concepto clásico de verdad esencialmente intacto. La aplicabilidad de la noción de verdad tendrá que sufrir algunasnas; restricciones, pero la noción seguirá estando disponible menos para los fines del discurso académico.

Para ello, tenemos que analizar aquellas características del lenguaje común que son la fuente real de la antinomia del mentiroso. Al llevar a cabo este análisis, nos daremos cuenta inmediatamente de un rasgo notable de este lenguaje: su carácter omnicomprensivo, universal. El lenguaje común es universal y está pensado para serlo. Se espera que proporcione las facilidades adecuadas para expresar cualquier cosa que pueda ser expresada en cualquier lenguaje: está continuamente expandiéndose para satisfacer esta exigencia En particular, es semánticamente universal en el sentido siguiente. En el lenguaje están incluidos, junto con los objetos lingüísticos, tales como las oraciones y los términos, que son componentes de este lenguaje, los nombre, de estos objetos (como sabemos, los nombres de expresiones pueden obtenerse poniendo las expresiones entre comillas); además, el lenguaje contiene términos semánticos tales como «verdad», «nombre», «designación», que dírecta o indirectamente se refieren a la relación entre los objetos lingüísticos y lo expresado por ellos. En consecuencia, podemos formar, para cada oración formulada en el lenguaje, otra oración que dice que la primera oración es verdadera o que es falsa. Usando un «truco» adicional hasta podemos construir en el lenguaje lo que se llama a veces una oración auto-referencial, es decir, una oración *S* que afirma el hecho que *S* misma es verdadera o de que es falsa. En caso de que *S* afirme su propia falsedad, podemos mostrar mediante un sencillo argumento que *S* es tanto verdadera como falsa —y nos enfrentamos nuevamente a la antinomia del mentiroso.

No hay, sin embargo, ninguna necesidad de usar lenguajes universales en todas las situaciones posibles. En particular, tales lenguajes no son necesarios para los propósitos de la ciencia (y por ciencia entiendo aquí todo el dominio de la investigación intelectual). En una rama particular de la ciencia, por ejemplo la química, uno discute ciertos objetos especiales, tales como los elementos, las moléculas, etc., pero no, por ejemplo, objetos lingüísticos tales como oraciones o términos. El lenguaje que está bien adaptado para esta discusión es un lenguaje restringido con un vocabulario limitado; debe contener nombres de objetos químicos, términos tales como «elemento» y «molécula», pero no nombres de objetos lingüísticos; por consiguiente, no tiene que ser semánticamente universal. Lo mismo se aplica a la mayoría de las otras ramas de la ciencia. La situación se torna algo confusa cuando nos ocupamos de la lingüística. Esta es una ciencia en la que estudiamos lenguajes; de modo que el lenguaje de la lingüística debe ciertamente estar provisto de nombres de objetos lingüísticos. Sin embargo, no tenemos que identificar el lenguaje de la lingüística con el lenguaje universal ni con ninguno de los lenguajes que son los objetos de la discusión lingüística y no estamos obligados a suponer que en lingüística usamos uno y el mismo lenguaje para todas las discusiones. El lenguaje de la lingüística debe contener los nombres de los componentes lingüísticos de los lenguajes discutidos, pero no los nombres de sus propios componentes; de manera que, una vez más, no tiene por qué ser semánticamente universal. Lo mismo se aplica al lenguaje de la lógica, o, mejor dicho, a aquella parte de la lógica conocida como metalógica y metamatemática; aquí nos ocupamos nuevamente de ciertos lenguajes, primariamente de los lenguajes de teorías lógicas y matemáticas (aunque discutimos estos lenguajes desde un punto de vista diferente del de la lingüística).

Surge ahora el interrogante de si la noción de verdad puede definirse precisamente, y si, de este modo, se puede establecer, al menos para los lenguajes semánticamente restringidos del discurso científico, un uso consistente y adecuado de esta noción. Bajo ciertas condiciones, la respuesta a esta pregunta resulta ser afirmativa. Las principales condiciones impuestas al lenguaje son que su vocabulario completo debe estar disponible y que sus reglas sintácticas referentes a la formación de oraciones y otras expresiones significativas a partir de las palabras listadas en el vocabulario deben ser formuladas precisamente. Además, las reglas sintácticas deberían ser puramente formales, esto es, deberían referirse exclusivamente a la forma (la figura) de las expresiones; la función y el significado de una expresión debería depender exclusivamente de su forma. En particular, al examinar una expresión, uno debería poder decidir en cada caso si la expresión es o no es una oración. No debería nunca suceder que una expresión funcione como una oración en un lugar, mientras que en algún otro lugar una expresión de la misma forma no funciona de ese modo, o que una oración sea afirmada en un contexto, mientras que una oración de la misma forma pueda ser negada en otro. (En particular, se sigue por consiguiente que pronombres demostrativos y adverbios tales como «éste» y «aquí» no deberían aparecer en el vocabulario del lenguaje). Los lenguajes que satisfacen estas condiciones son llamados lenguajes formalizados. Cuando discutimos un lenguaje formalizado no hay necesidad de distinguir entre expresiones de la misma forma que han sido escritas o proferidas en lugares diferentes; a menudo se habla de ellas como si fueran una y la misma expresión. El lector puede haberse dado cuenta que a veces usamos esta manera de hablar cuando discutimos un lenguaje natural, es decir, uno que no está formalizado; lo hacemos por simplicidad, y sólo en aquellos casos en los que parece no haber riesgo de confusión.

Los lenguajes formalizados son completamente adecuados para la presentación de las teorías lógicas y matemáticas. No veo ninguna razón esencial por la que no puedan ser adaptados para su uso en otras disciplinas científicas y, en particular para el desarrollo de las partes teóricas de las ciencias empíricas. Me gustaría enfatizar que cuando uso el término «lenguajes formalizados» no me refiero exclusivamente a sistemas lingüísticos que se formulan enteramente en símbolos, y no estoy pensando en algo esencialmente opuesto a los lenguajes naturales. Por el contrario, los únicos lenguajes formalizados que parecen tener real interés son aquellos que son fragmentos de lenguajes naturales (fragmentos provistos de vocabularios completos y reglas sintácticas precisas) o aquellos que pueden al menos ser adecuadamente traducidos a los lenguajes naturales.

Hay algunas condiciones más de las cuales depende la realización de nuestro programa. Deberíamos hacer una distinción estricta entre el lenguaje que es el objeto de nuestra discusión y para el cual intentamos, en particular, construir la definición de verdad, y el lenguaje en el cual se formula esta definición y se estudian sus implicaciones. A este ultimo se lo llama *metalenguaje* y al primero *lenguaje-objeto*. El metalenguaje debe ser suficientemente rico; en particular, debe incluir al lenguaje-objeto como parte suya. De hecho, de acuerdo con nuestras estipulaciones, una definición adecuada de verdad implicará como consecuencias todas las definiciones parciales de esta noción, esto es, todas las equivalencias de la forma (3):

«*p*» es verdadera si y sólo si *p*,

donde «*p*» debe ser reemplazada (en ambos lados de la equivalencia) por una oración arbitraria del lenguaje-objeto. Dado que todas estas consecuencias están formuladas en el metalenguaje, concluimos que toda oración del lenguaje-objeto debe también ser una oración del metalenguaje. Además, el metalenguaje debe contener nombres para oraciones (y otras expresiones) del lenguaje-objeto, dado que estos nombres aparecen en el lado izquierdo de las equivalencias anteriores. Debe también contener algunos términos más que son necesarios para la discusión del lenguaje-objeto, de hecho términos que denotan ciertos conjuntos especiales de expresiones, relaciones entre expresiones, y operaciones sobre expresiones; por ejemplo, debemos ser capaces de hablar del conjunto de todas las oraciones o de la operación de yuxtaposición, por medio de la cual, al poner una de dos expresiones dadas inmediatamente después de la otra, formamos una nueva expresión. Finalmente, mediante nuestra definición de verdad, mostramos que los términos semánticos (que expresan relaciones entre las oraciones del lenguaje-objeto y los objetos a los que se refieren estas oraciones) pueden introducirse en el metalenguaje por medio de definiciones. Por consiguiente, concluimos que el metalenguaje que proporciona los medios suficientes para definir verdad debe ser esencialmente más rico que el lenguaje objeto; no puede coincidir con, o ser traducible a este último, ya que de otro modo ambos lenguajes resultarían ser semánticamente universales, y la antinomia del mentiroso podría reconstruirse en ambos. Volveremos a esta cuestión en la ultima sección de este artículo.

Si todas las condiciones anteriores son satisfechas, la construcción de la definición de verdad deseada no presenta ninguna dificultad esencial. Técnicamente, sin embargo, es demasiado complicada para explicarla aquí en detalle. Para cualquier oración dada del lenguaje objeto puede fácilmente formularse la correspondiente definición parcial de la forma (3). Sin embargo, dado que el conjunto de todas las oraciones del lenguaje-objeto es por regla general infinito, mientras que toda oración del metalenguaje es una secuencia finita de signos, no podemos obtener una definición general simplemente formando la conjunción lógica de todas las definiciones parciales. Sin embargo, lo que finalmente obtenemos es en algún sentido intuitivo equivalente a la imaginaria conjunción infinita. De modo aproximado, procedemos de la siguiente manera. Primero, consideramos las oraciones más simples, que no incluyen ninguna otra oración como parte; para esas oraciones logramos definir verdad directamente (usando la misma idea que nos conduce a las definiciones parciales). Luego, haciendo uso de reglas sintácticas que se refieren a la formación de oraciones más complicadas a partir de otras más simples, extendemos la definición a oraciones compuestas arbitrarias; aplicamos aquí el método conocido en matemática como definíción por recursión. (Esta es una mera aproximación al procedimiento real. Debido a algunas razones técnicas, el método de la recursión no se aplica en realidad para definir la noción de verdad, sino la noción semántica relacionada de satisfacción. La verdad es entonces fácilmente definida en términos de satisfacción).

Sobre la base de una definición construida de este modo podemos desarrollar toda la teoría de la verdad. En particular, podemos derivar de ella, además de todas las equivalencias de la forma (3), algunas consecuencias de naturaleza general, tales como las famosas leyes de (no) contradicción y de tercero excluido. De acuerdo a la primera de estas leyes, para ningún par de oraciones, una de las cuales sea la negación de la otra, puede darse el caso de que ambas sean verdaderas; de acuerdo a la segunda, dos oraciones de ese tipo no pueden ser ambas falsas.

§2. La noción de demostración

Sea lo que fuere que pueda lograrse construyendo una definición adecuada de verdad para un lenguaje científico, una cosa parece segura: la definición no conlleva un criterio aplicable para decidir si determinadas oraciones de este lenguaje son verdaderas o falsas (y, de hecho, no está en absoluto concebida con este propósito). Considérese, por ejemplo, una oración del lenguaje de la geometría del colegio secundario, digamos «las tres bisectrices de un triángulo se cruzan en un punto». Si estamos interesados en la cuestión de la verdad de esta oración y recurrimos a la definición de verdad para obtener una respuesta, nos debemos preparar para una decepción. La única información que obtenemos es que la oración es verdadera si las tres bisectrices de un triángulo siempre se cruzan en un punto, y es falsa si no es así; pero sólo una investigación geométrica nos permite decidir cuál es efectivamente el caso. Observaciones análogas se aplican a oraciones del ámbito de cualquier otra ciencia particular; decidir si una de esas oraciones es o no es verdadera es una tarea de la ciencia, y no de la lógica o de la teoría de la verdad.

Algunos filósofos y metodólogos de la ciencia están dispuestos a rechazar cualquier definición que no proporcione un criterio para decidir si un objeto particular dado cae bajo la noción definida o no. En la metodología de las ciencias empíricas esa tendencia está representada por la doctrina del operacionalismo; los filósofos de la matemática que pertenecen a la escuela constructivista parecen exhibir una tendencia similar. En ambos casos, sin embargo, los que sostienen esta opinión parecen ser una pequeña minoría. Un intento serio de llevar a cabo este programa en la práctica (esto es, de desarrollar una ciencia sin usar definiciones indeseables) no ha sido hecho casi nunca. Parece claro que con este programa desaparecería mucho de la matemática contemporánea, y que partes teóricas de la física, química, biología y otras ciencias empíricas serían severamente mutiladas. Las definiciones de nociones tales como las de átomo o gen, y la mayoría de las definiciones de la matemática, no conllevan ningún criterio para decidir si un objeto cae o no bajo el término que se ha definido.

Dado que la definición de verdad no nos proporciona ningún criterio de este tipo y que, al mismo tiempo, la búsgueda de la verdad es correctamente considerada como la esencia de las actividades científicas, aparece como un problema importante el encontrar al menos criterios parciales de verdad y desarrollar procedimientos que nos permitan afirmar o negar la verdad (o, al menos, la probabilidad de verdad) de tantas oraciones como sea posible. Se conocen, en efecto, procedimientos tales; algunos de ellos son usados exclusivamente en las ciencias empíricas y algunos primariamente en las ciencias deductivas. La noción de demostración —la segunda noción que será discutida en este artículo— se refiere justamente a un procedimiento usado para afirmar la verdad de oraciones que es empleado primariamente en la ciencia deductiva. Este procedimiento es un elemento esencial de lo que es conocido como método axiomático, el único método utilizado en la actualidad para desarrollar disciplinas matemáticas.

El método axiomático y, dentro de su marco, la noción de demostración son producto de un largo desarrollo histórico. Algún conocimiento esquemático de este desarrollo es, probablemente, esencial para la comprensión de la noción contemporánea de demostración.

Originalmente, una disciplina matemática era un agregado de oraciones que se referían a una cierta clase de objetos o fenómenos, que se formulaban mediante un cierto repertorio de términos, y que se aceptaban como verdaderas. Este agregado de oraciones no tenía ningún orden estructural. Una oración era aceptada como verdadera, o porque parecía intuitivamente evidente, o sino porque se la demostraba a partir de algunas oraciones intuitivamente evidentes, y de esta manera se mostraba, por medio de un argumento intuitivamente seguro, que eran una consecuencia de estas otras oraciones. El criterio de evidencia intuitiva (y de certeza intuitiva de los argumentos) se aplicaba sin ninguna restricción; toda oración reconocida como verdadera por medio de este criterio era automáticamente incluida en la disciplina. Esta descripción parece ajustarse a, por ejemplo, la ciencia de la geometría tal como era conocida por los antiguos egipcios y por los griegos en su estadio temprano, pre-euclidiano.

Sin embargo, muy pronto se dieron cuenta de que el criterio de evidencia intuitiva está lejos de ser infalible, no tiene carácter objetivo y conduce a menudo a serios errores. Todo el desarrollo subsiguiente del método axiomático puede verse como una expresión de la tendencia a restringir el recurso a la evidencia intuitiva.

Esta tendencia se manifestó en primer lugar en el esfuerzo por demostrar tantas oraciones como fuera posible, y, por consiguiente, por restringir todo lo que fuera posible el número de oraciones aceptadas como verdaderas meramente sobre la base de su evidencia intuitiva. Lo ideal, desde este punto de vista, sería demostrar toda oración que sea aceptada como verdadera. Por razones obvias, este ideal no puede ser realizado. En efecto, demostramos cada oración sobre la base de otras oraciones, demostramos estas otras oraciones sobre la base de algunas otras oraciones, y así siguiendo: si queremos evitar tanto un círculo vicioso como un regreso infinito, el procedimiento debe detenerse en algún lugar. Dos principios emergieron, y fueron subsiguientemente aplicados en la construcción de disciplinas matemáticas, como un compromiso entre ese ideal inalcanzable y las posibilidades realizables. De acuerdo al primero de estos principios, toda disciplina comienza con una lista de un número pequeño de oraciones, llamadas axiomas u oraciones primitivas, que parecen intuitivamente evidentes y que son reconocidas como verdaderas sin ninguna justificación adicional. De acuerdo al segundo princípio, ninguna otra oración se acepta como verdadera en la disciplina a menos que seamos capaces de demostrarla con la ayuda exclusiva de los axiomas y de aquellas oraciones que fueron previamente demostradas. Todas las oraciones que pueden ser reconocidas como verdaderas en virtud de estos dos principios son llamadas teoremas u oraciones demostradas de la disciplina dada. Dos principios análogos se ocupan del uso de los términos en la construcción de la disciplina. De acuerdo al primero de ellos, hacemos al comienzo una lista de unos pocos términos, llamados términos no-definidos o primitivos, que resultan directamente comprensibles y que decidimos usar (al formular y demostrar teoremas) sin explicar su significado; de acuerdo al segundo principio, acordamos no usar ningún término adicional a menos que seamos capaces de explicar su significado definiéndolo con la ayuda de términos no definidos y términos previamente definidos. Estos cuatro principios son los fundamentos del método axiomático; las teorías desarrolladas de acuerdo con estos principios se llaman teorías axiomáticas.

Como es bien sabido, el método axiomático se aplicó al desarrollo de la geometría en los *Elementos* de Euclides alrededor del 300 A.C.. De ahí en más fue usado durante más de 2000 años sin prácticamente ningún cambio de sus principios más importantes (los que, dicho sea de paso, no fueron ni siquiera formulados explícitamente durante un período muy largo de tiempo) ni del enfoque general del tema. Sin embargo, en los siglos XIX y XX el concepto de método axiomático sufrió una profunda evolución. Aquellos aspectos de la evolución que tienen que ver con la noción de demostración son particularmente significativos para nuestra discusión.

Hasta los últimos años del siglo XIX la noción de demostración tuvo primariamente un carácter psicológico. Una demostración era una actividad intelectual que tenía como finalidad convencer a uno mismo y a los demás de la verdad de la oración en cuestión; más específicamente, las demostraciones se usaban en el desarrollo de una teoría matemática para convencer a uno mismo y a los demás de que la oración en cuestión tenía que ser aceptada como verdadera una vez que habían sido aceptadas como tales algunas otras oraciones. No se ponía ninguna restricción a los argumentos usados en las pruebas, excepto que tenían que ser intuitivamente convincentes. En un cierto período, sin embargo, se comenzó a sentir la necesidad de someter la noción a un análisis más profundo, lo que tendría como resultado una restricción del recurso a la intuición también en este contexto. Esto estuvo probablemente relacionado con algunos desarrollos específicos en matemática, en particular con el descubrimiento de las geometrías no euclidianas. El análisis fue hecho por lógicos, comenzando por el lógico alemán Gottlob Frege; condujo a la introducción de una nueva noción, la de *demostración formal*, que resultó ser un sustituto adecuado y una mejora esencial de la vieja noción psicológica.

El primer paso para dotar a una teoría matemática de la noción de demostración formal es la formalización del lenguaje de la teoría, en el sentido discutido previamente en conexión con la definición de verdad. Así, se proporcionan reglas sintácticas formales que, en particular, nos permiten distinguir una oración de una expresión que no es una oración, simplemente observando las formas de las expresiones. El próximo paso consiste en formular unas pocas reglas de una naturaleza diferente, las llamadas reglas de demostración (o de inferencia). Mediante estas reglas, una oración se considera directamente derivable de ciertas oraciones si, hablando en general, su forma está relacionada de una manera prescrita con las formas de esas oraciones. El número de reglas de demostración es pequeño, y su contenido es simple. Igual que las reglas sintácticas, todas tienen un carácter formal, esto es, se refieren exclusivamente a las formas de las oraciones involucradas. Intuitivamente, todas las reglas de derivación aparecen como infalibles, en el sentido de que una oración que es directamente derivable de oraciones verdaderas por medio de cualquiera de estas reglas debe, ella misma, ser verdadera. En realidad, la infalibilidad de las reglas de demostración puede establecerse sobre la base de una definición adecuada de verdad. El ejemplo mejor conocido y más importante de regla de demostración es la regla de separación, también conocida como *modus ponens*. De acuerdo con esta regla (que en algunas teorías es la única regla de demostración), una oración «*q*» es directamente derivable de dos oraciones dadas si una de ellas es la oración condicional «si *p*, entonces *q*», mientras que la otra es «*p*»; aquí «*p*» y «*q*» son, como es usual, abreviaturas de dos oraciones cualesquiera de nuestro lenguaje formalizado. Podemos ahora explicar en qué consiste la prueba formal de una oración dada. Primero, aplicamos las reglas de demostración a los axiomas y obtenemos nuevas oraciones que son directamente derivables de los axiomas; a continuación, aplicamos las mismas reglas a nuevas oraciones, o conjuntamente a nuevas oraciones y axiomas, y obtenemos más oraciones; y continuamos con este proceso. Si luego de un número finito de pasos llegamos a una oración dada, decimos que la oración ha sido demostrada formalmente. Esto puede también expresarse de un modo más preciso de la siguiente manera: una demostración formal de una oración dada consiste en la construcción de una secuencia finita de oraciones tal que (1) la primera oración de la secuencia es un axioma, (2) cada una de las oraciones siguientes es, o bien un axioma, o bien es directamente derivable de alguna de las oraciones que la preceden en la secuencia en virtud de una de las reglas de demostración, y (3) la última oración de la secuencia es la oración a demostrar. Cambiando ligeramente el uso del término «demostración», podemos decir que una demostración formal de una oración es simplemente una secuencia finita de oraciones con las tres propiedades que acabamos de mencionar.

Una teoría axiomática cuyo lenguaje ha sido formalizado y para la cual se ha proporcionado una noción de demostración formal se llama teoría formalizada. Estipulamos que las únicas demostraciones que pueden usarse en una teoría formalizada son demostraciones formales; ninguna oración puede aceptarse como teorema a menos que aparezca en la lista de axiomas o pueda encontrarse para ella una demostración formal. El método de presentación de una teoría formalizada en cada etapa de su desarrollo es, en principio, muy elemental. Primero hacemos una lista de los axiomas, y luego de todos los teoremas conocidos, en un orden tal que cualquier oración de la lista que no sea un axioma pueda ser reconocida directamente como teorema, simplemente comparando su forma con las formas de las oraciones que la preceden en la lista; no hay involucrado ningún proceso complejo de razonamiento y de convencimiento. (No estoy hablando aquí de los procesos psicológicos por medio de los cuales los teoremas fueron, de hecho, descubiertos). El recurso a la evidencia intuitiva ha sido, en verdad, considerablemente restringido; las dudas acerca de la verdad de los teoremas no han sido enteramente eliminadas, pero han sido reducidas a las posibles dudas acerca de la verdad de las pocas proposiciones listadas como axiomas y la infalibilidad de unas pocas y simples reglas de demostración. Puede añadirse que el proceso de introducción de nuevos términos en el lenguaje de la teoría también puede ser formalizado proporcionando reglas formales especiales de definición.

Se sabe ahora que todas las disciplinas matemáticas existentes pueden ser presentadas como teorías formalizadas. Pueden proporcionarse demostraciones formales para los teoremas matemáticos más profundos y complicados, que fueron originalmente establecidos mediante argumentos intuitivos.

§3. La relación entre verdad y demostración

Indudablemente, fue un gran logro de la lógica moderna el haber reemplazado la vieja noción psicológica de demostración, que difícilmente podría haberse hecho clara y precisa, por una noción nueva y simple de un carácter puramente formal. Pero el triunfo del método formal trajo consigo el germen de un futuro revés. Como veremos, la misma simplicidad de la nueva noción resultó ser su talón de Aquiles.

Para evaluar la noción de demostración formal, tenemos que clarificar su relación con la noción de verdad. Después de todo, la demostración formal, como la vieja demostración intuitiva, es un procedimiento destinado a adquirir nuevas oraciones verdaderas. Tal procedimiento será adecuado sólo si todas las oraciones adquiridas con su ayuda resultan ser verdaderas y todas las oraciones verdaderas pueden ser adquiridas con su ayuda. Por consiguiente, surge naturalmente un problema: ¿es la demostración formal realmente un procedimiento adecuado para adquirir la verdad?

En otras palabras: ¿coincide el conjunto de todas las oraciones formalmente demostrables con el conjunto de todas las oraciones verdaderas? Para ser específicos, referimos este problema a una disciplina matemática particular muy elemental, la aritmética de los números naturales (la teoría numérica elemental). Suponemos que esta teoría ha sido presentada como una teoría formalizada. El vocabulario de la teoría es magro. De hecho, está formado por variables tales como «*m*», «*n*», «*p*», ..., que representan números naturales arbitrarios; por los numerales «0», «1», «2», ..., que denotan números particulares; por símbolos que denotan algunas relaciones familiares entre números y operaciones sobre números tales como «=», «<», «+», «–» y, finalmente, por ciertos términos lógicos, a saber, conectivas oracionales («y», «o», «si», «no») y cuantificadores (expresiones de la forma «para todo número *m*» y «para algún número *m*»). Las reglas sintácticas y las reglas de demostración son simples. Cuando hablemos de oraciones en la discusión subsiguiente, siempre estaremos pensando en las oraciones del lenguaje formalizado de la aritmética.

Sabemos por la discusión sobre la verdad de la primera sección que, tomando este lenguaje como lenguaje objeto, podemos construir un metalenguaje apropiado y formular en él una definición adecuada de verdad. Resulta conveniente en este contexto decir que lo que hemos definido de este modo es el conjunto de las oraciones verdaderas; de hecho, la definición de verdad enuncia que ciertas condiciones formuladas en el metalenguaje son satisfechas por todos los elementos de este conjunto (esto es, todas las oraciones verdaderas) y sólo por estos elementos. Más fácilmente todavía podemos definir en el metalenguaje el conjunto de las oraciones demostrables; la definición está completamente de acuerdo con la explicación de la noción de demostración formal que se dio en la segunda sección. Estrictamente hablando, tanto la definición de verdad como la de demostrabilidad pertenecen a una nueva teoría formulada en el metalenguaje y específicamente destinada al estudio de nuestra aritmética formalizada y de su lenguaje. La nueva teoría se llama metateoría o, más específicamente, meta-aritmética. No nos extenderemos aquí sobre la manera en que se construye la metateoría, sobre sus axiomas, términos no-definidos, etc.. Sólo destacaremos que es en el marco de esta metateoría que formularemos y resolveremos el problema de si el conjunto de las oraciones demostrables coincide con el de las oraciones verdaderas.

La solución del problema resulta ser negativa. Daremos aquí una explicación esquemática del método por el cual se ha llegado a esta solución. La idea principal está estrechamente relacionada con la usada por el lógico norteamericano (de origen austríaco) Kurt Gödel en su famoso artículo sobre la incompletitud de la aritmética.

Se ha señalado en la primera sección que el metalenguaje que nos permite definir y discutir la noción de verdad debe ser rico. Contiene todo el lenguaje-objeto como una parte suya y, por lo tanto, podemos hablar en él de números naturales, conjuntos de números, relaciones entre números, y así siguiendo. Pero también contiene los términos necesarios para la discusión del lenguaje-objeto y sus componentes; por consiguiente, podemos hablar en el metalenguaje de expresiones y, en particular, de oraciones, de conjuntos de oraciones, de relaciones entre oraciones, y así siguiendo. Por lo tanto, podemos estudiar en la metateoría las propiedades de estos diversos tipos de objetos y establecer conexiones entre ellos.

En particular, resulta fácil disponer todas las oraciones (desde las más simples hasta las de mayor complejidad) en una secuencia infinita y numerarlas consecutivamente, usando la descripción de oraciones que proporcionan las reglas sintácticas del lenguaje objeto. Correlacionamos así cada oración con un número natural, de tal manera que dos números correlacionados con dos oraciones diferentes son siempre diferentes; en otras palabras, establecemos una correspondencia uno a uno entre oraciones y números. Esto, a su vez, conduce a una correspondencia similar entre conjuntos de oraciones y conjuntos de números, o relaciones entre oraciones y relaciones entre números. En particular, podemos considerar los números de las oraciones demostrables y los números de las oraciones verdaderas; brevemente, los llamaremos números demostrables* y números verdaderos*. Nuestro problema principal queda reducido entonces a la pregunta: ¿son idénticos el conjunto de los números demostrables* y el conjunto de los números verdaderos*?

Para responder negativamente a esta pregunta, es suficiente, por supuesto, indicar una sola propiedad que se aplique a un conjunto pero no al otro. La propiedad que de hecho exhibiremos puede parecer un poco inesperada, un tipo de *deus ex machina.*

La simplicidad intrínseca de las nociones de demostración formal y de demostrabilidad formal desempeñara aquí un papel básico. Hemos visto en la segunda sección que el significado de estas nociones se explica, esencialmente, en términos de ciertas relaciones simples entre oraciones prescritas por unas pocas reglas de demostración; el lector recordará aquí la regla del *modus ponens.* Las correspondientes relaciones entre números de oraciones son igualmente simples; resulta que pueden ser caracterizados en términos de las operaciones y relaciones aritméticas más simples, tales como la adición, la multiplicación y la igualdad —por lo tanto, en términos que aparecen en nuestra teoría aritmética. Se puede describir lo que se ha logrado diciendo que la definición de demostrabilidad ha sido traducida del metalenguaje al lenguaje-objeto.

Por otra parte, la discusión sobre la noción de verdad en los lenguajes comunes sugiere fuertemente la conjetura de que no puede obtenerse una traducción de ese tipo para la definición de verdad; de otra manera, el lenguaje-objeto resultaría ser en un sentido semánticamente universal, y una reaparición de la antinomia del mentiroso seria inminente. Confirmamos esta conjetura mostrando que si el conjunto de los números verdaderos* pudiera ser definido en el lenguaje de la aritmética, la antinomia del mentiroso podría, de hecho, ser reconstruida en este lenguaje. Dado que, sin embargo, estamos tratando con un lenguaje formalizado, la antinomia asumiría una forma más complicada y sofisticada. En particular, no aparecerían en la nueva formulación expresiones con un contenido empírico, tal como «la oración impresa en tal-y-tal lugar», que formaban parte esencial de la formulación original de la antinomia. No entraremos aquí en más detalles.

De manera que el conjunto de los números demostrables* no coincide con el conjunto de los números verdaderos*, dado que el primero es definible en el lenguaje de la aritmética, mientras que el segundo no lo es. En consecuencia, el conjunto de las oraciones demostrables y el conjunto de las oraciones verdaderas tampoco coinciden. Por otra parte, usando la definición de verdad podemos mostrar fácilmente que todos los axiomas de la aritmética son verdaderos y que todas las reglas de demostración son infalibles. Por lo tanto, todas las oraciones demostrables son verdaderas; luego, la conversa no puede darse. Así, nuestra conclusión final es: hay oraciones formuladas en el lenguaje de la aritmética que son verdaderas pero que no pueden ser demostradas sobre la base de los axiomas y las reglas de demostración aceptadas en aritmética.

Se podría pensar que esta conclusión depende esencialmente de ciertos axiomas y reglas de inferencia específicos, elegidos para nuestra teoría aritmética, y que el resultado final de la discusión podría ser diferente si enriqueciéramos de manera apropiada la teoría mediante el añadido de nuevos axiomas o nuevas reglas de inferencia. Un análisis más detallado muestra, sin embargo, que el argumento depende muy poco de las propiedades específicas de la teoría discutida, y que, de hecho, se extiende a la mayoría de las teorías formalizadas. Suponiendo que una teoría incluye a la aritmética de los números naturales como parte (o que, al menos, la aritmética puede ser reconstruida en ella), podemos repetir la porción esencial de nuestro argumento sin prácticamente ningún cambio; de esta manera, concluimos nuevamente que el conjunto de las oraciones demostrables de la teoría es diferente del conjunto de sus oraciones verdaderas. Si, además, podemos mostrar (como sucede frecuentemente) que todos los axiomas de la teoría son verdaderos y que todas las reglas de inferencia son infalibles, concluimos además que hay oraciones verdaderas de la teoría que no son demostrables. Con la excepción de algunas teorías fragmentarias con medios restringidos de expresión, el supuesto sobre la relación de la teoría con la aritmética de los números naturales es generalmente satisfecho, y, por lo tanto, nuestras conclusiones tienen un carácter casi universal. (En lo que respecta a aquellas teorías fragmentarias que no incluyen a la aritmética de los números naturales, sus lenguajes pueden no estar provistos de los medios suficientes para definir la noción de demostrabilidad, y sus oraciones demostrables pueden de hecho coincidir con sus oraciones verdaderas. La geometría elemental y el álgebra elemental de los números reales son los ejemplos mejor conocidos, y quizá los más importantes, de teorías en las que estas nociones coinciden).

El papel dominante que tiene la antinomia del mentiroso en todo el argumento ilumina de manera interesante nuestras observaciones anteriores acerca del papel de las antinomias en la historia del pensamiento humano. La antinomia del mentiroso apareció por primera vez en nuestra discusión como una especie de fuerza maligna con gran poder destructivo. Nos obligó a abandonar todos los intentos de clarificar la noción de verdad para los lenguajes naturales. Tuvimos que restringir nuestros esfuerzos a los lenguajes formalizados del discurso científico. Como salvaguarda contra una posible reaparición de la antinomia, tuvimos que complicar considerablemente la discusión distinguiendo entre un lenguaje y su metalenguaje. Subsiguientemente, sin embargo, en el nuevo marco restringido logramos domesticar su energía destructiva y enjaezarla para propósitos pacíficos y constructivos. La antinomia no ha reaparecido, pero su idea básica ha sido usada para establecer un resultado metalógico significativo de implicaciones de gran alcance.

El hecho de que sus implicaciones filosóficas sean esencialmente negativas no quita nada a la significación de este resultado. En verdad, el resultado muestra que la noción de demostrabilidad no es un sustituto perfecto de la noción de verdad en ningún dominio de la matemática. La creencia de que la demostración formal puede servir como instrumento adecuado para establecer la verdad de todos los enunciados matemáticos no ha resultado fundamentada. El triunfo inicial de los métodos formales ha sido seguido por un serio revés.

Cualquier cosa que se diga para concluir esta discusión resultará seguramente un anticlímax. La noción de verdad para teorías formalizadas puede ahora ser introducida por medio de una definición precisa y adecuada. Puede, por lo tanto, ser usada sin restricción ni reserva alguna en la discusión metalógica. De hecho, se ha convertido en una noción metalógica básica involucrada en problemas y resultados importantes. Por otra parte, la noción de demostración no ha perdido tampoco su significación. La demostración es todavía el único método utilizado para afirmar la verdad de oraciones dentro de cualquier teoría matemática específica. Sin embargo, ahora somos conscientes de que hay oraciones formuladas en el lenguaje de la teoría que son verdaderas pero no demostrables, y no podemos desechar la posibilidad de que alguna de esas oraciones aparezca entre las que nos interesan y estamos tratando de demostrar. Por consiguiente, en algunas situaciones quizá deseemos explorar la posibilidad de ampliar el conjunto de las oraciones demostrables. Con este fin, enriquecemos la teoría dada incluyendo nuevas oraciones en su sistema de axiomas o proporcionándole nuevas reglas de demostración. Al hacerlo, usamos la noción de verdad como guía; ya que no queremos añadir un nuevo axioma o una nueva regla de demostración si tenemos razones para creer que el nuevo axioma no es una oración verdadera, o que la nueva regla de demostración puede producir una oración falsa cuando se la aplica a oraciones verdaderas. El proceso de extender una teoría puede, por supuesto, ser repetido un número arbitrario de veces. De esta manera, la noción de oración verdadera funciona como un límite ideal que nunca puede ser alcanzado, pero al cual tratamos de aproximarnos ampliando gradualmente el conjunto de las oraciones demostrables. (Parece probable, aunque por razones diferentes, que la noción de verdad juegue un papel análogo en el reino del conocimiento empírico). No hay ningún conflicto entre las nociones de verdad y demostración en el desarrollo de la matemática; las dos nociones no están en guerra, sino que viven en pacífica coexistencia.

Versión castellana de CARLOS OLLER

Recibido: *20-Mayo-2015* | Aceptado: *30-Julio2015*

ALFRED TARSKI, fue Profesor de Filosofía en la University of California, Berkeley, EUA. Doctor en Filosofia (PhD) por la Universidad de Varsovia, Polonia. Sus principales áreas de interés fueron la matemática, lógica y filosofía del lenguaje. Entre sus principales publicaciones se cuentan: *Introduction to Logic and to the Methodology of Deductive Sciences* (New York: Dover, 1941), y *Logic, Semantics, Metamathematics* (Oxford: Clarendon Press, 1956).

CARLOS OLLER, es Profesor de Lógica en la Universidad de Buenos Aires y de Teoria de la Argimentación en la Universidad Nacional de La Plata, Argentina. Doctor en Filosofía (PhD). Sus principales áreas de interés son la lógica, la argumentación y epístemología. Ha publicado artículos y traducciones en algunas antologías y revistas académicas de filosofía, tales como: *Análisis Filosófico*, *Journal of Symbolic Logic*, *Logic and Logical Philosophy*, y *Journal of Applied Logic*, entre otras.

DIRECCIÓN POSTAL: Departamento de Filosofía, Universidad de Buenos Aires, Puán 480, CABA, Argentina. e-mail (✉): carlos.a.oller@gmail.com

COMO CITAR ESTE TRABAJO: TARSKI, Alfred «Verdad y Demostración». *Disputatio. Philosophical Research Bulletin* 4:5 (2015): pp. 367-396.