

Metalingüajes y Poder Expresivo

Ramiro Caso (UBA – GAF)

jcaso@intramed.net.ar

Tarski | teorías de la verdad
jerarquías de lenguaje | teoría de modelos

El Teorema de Tarski acerca de la Indefinibilidad de la Verdad es un resultado limitativo de la capacidad expresiva de una amplia clase de lenguajes formales interpretados. En su forma más abstracta afirma que, para todo lenguaje L suficientemente expresivo, el predicado de verdad de L no puede ser expresado en L mismo. Como consecuencia de ello, si se desea dar una definición de verdad para uno cualquiera de tales lenguajes, ésta debe construirse en otro lenguaje L' más expresivo, que opere como metalenguaje respecto del primero. Este procedimiento puede iterarse y, bajo ciertas condiciones, el proyecto de dar una definición de verdad para cada etapa del proceso da lugar a una jerarquía linealmente ordenada de lenguajes de poder expresivo creciente, cada uno de los cuales contiene el predicado veritativo del lenguaje que lo precede.¹

En lo que respecta a la noción de *mayor poder expresivo*, Tarski [1944] mantiene que la *riqueza esencial* del metalenguaje respecto del lenguaje objeto es a la vez condición necesaria y suficiente para que una definición materialmente adecuada y formalmente correcta de verdad en L_0 pueda ser construida en L_M , afirmación que llamaremos la Tesis de Tarski:

[ER.] Es posible construir en L_M una definición materialmente adecuada y formalmente correcta de verdad para L_0 si y sólo si L_M es esencialmente más rico que L_0

Dado que la noción de *ser esencialmente más rico que* juega un rol fundamental en la formulación de esta tesis, es deseable una caracterización

¹ O para todos los precedentes, si la jerarquía se construye de modo que los lenguajes satisfagan este requisito más fuerte.

más precisa de la misma.

En una posición ya clásica, DeVidi y Solomon [1999] han argumentado que [ER] resulta o trivial o falsa, debido a que no habría una propiedad suficientemente natural que todo par $L_O - L_M$ posea, tal que haga a la tesis de Tarski a la vez informativa y verdadera, por cuanto la única propiedad que hace verdadera [ER] es “salvajemente” disyuntiva y, como tal, filosóficamente no interesante. En parte como consecuencia de esto, y sobre la base de un supuesto ejemplo, los autores mantienen que es posible construir “jerarquías” tarskianas que no constituyen un ordenamiento lineal estricto, debido a que, bajo cualquier lectura que se haga de ella, la noción de riqueza esencial no satisface la propiedad de asimetría.

La tarea es por tanto doble: en primer lugar, encontrar una propiedad P tal que (i) L_M posee P respecto de L_O , si y sólo si L_M puede dar una definición materialmente adecuada y formalmente correcta de verdad para L_O , y (ii) es lo suficientemente natural como para hacer a la tesis de Tarski informativa a la vez que verdadera; en segundo lugar, debe mostrarse que la propiedad identificada permite construir únicamente jerarquías linealmente ordenadas de lenguajes. En la próxima sección introduciremos la noción de K_M -representabilidad del predicado de verdad de L_O , y en las siguientes sostendremos que la misma es un buen candidato para tal propiedad.

I

Una consideración más detallada del Teorema puede ser de utilidad. En su forma abstracta, sostiene que cualquier lenguaje formal interpretado que satisface ciertas condiciones mínimas no puede expresar su propio concepto de verdad. Primero, caracterizamos la idea de expresabilidad en L , o L -definibilidad:

Definición 1. Un conjunto $C \subset D_L$ es expresable en L (L -definible) si y sólo si existe una fórmula $\varphi(x)$ de L tal que para todo $d \in D_L$

$$U_L \models \varphi(x)[d/x] \text{ sii } d \in C$$

donde U_L es la estructura que da la interpretación de L . Es usual en teoría de

modelos permitir que la interpretación de un lenguaje formal pueda variar. Tarski [1935, 1939, 1944] trata con lenguajes formales plenamente interpretados; reflejamos este uso tarskiano haciendo que la interpretación permanezca constante, con lo cual entendemos un lenguaje L como un par $L = \langle L, U_L \rangle$, donde L es la parte sintáctica de L .

Teorema de Tarski (Forma Abstracta) 1. *Para todo lenguaje L tal que*

- 1. puede expresar su propia sintaxis y*
- 2. sigue las leyes de la lógica clásica*

el conjunto T^ de códigos de las oraciones verdaderas de L no es L -definible*

donde $T^* \subset D_L$ es el subconjunto de elementos de D_L que codifican las oraciones verdaderas de L en L mismo.

Para entender mejor las nociones de codificación y definibilidad en L , es conveniente tener en mente alguna forma particular en que un lenguaje puede representar su propia sintaxis. Para tener definición al respecto, suponemos que L puede hablar de su sintaxis por medio de aritmetización. La idea de que la sintaxis de un lenguaje es aritmetizable conlleva dos afirmaciones: primero, que las expresiones del lenguaje pueden representarse por medio de números naturales, esto es, que es posible asignar a cada expresión un único número natural, de manera de crear una imagen isomórfica del lenguaje en ω , el conjunto de los números naturales (a estos números que representan o codifican expresiones se los suele llamar números de Gödel); segundo, que las propiedades y relaciones sintácticas de las expresiones del lenguaje pueden ser expresadas por medio de propiedades y relaciones aritméticas entre los correspondientes números de Gödel. Como resultado de esto, pueden entenderse las fórmulas de L como hablando indirectamente acerca de la sintaxis de L .

Podemos describir el resultado de Tarski de una manera un tanto diferente: para todo predicado $F(x)$ de L , o bien $U_L \models F(n)$, y n no es el código de una oración verdadera de L , o bien $U_L \models \neg F(n)$, y n es el código de una oración verdadera de L . Por razones de espacio, obviaremos la demostración del teorema.

Ahora bien, una cosa es poder expresar un conjunto en un lenguaje, y otra diferente es poder representar dicho conjunto en una teoría formulada

en ese lenguaje. Dicho de otra manera, pudiera ocurrir que nuestro lenguaje L fuese suficientemente rico como para expresar un conjunto C , y nuestra teoría K formulada en L demasiado pobre como para probar todos los hechos relevantes acerca de C que son verdaderos en L . De hecho, ya sabemos, por [Gödel, 1931], que para lenguajes y teorías suficientemente ricos como para ser interesantes, verdad y probabilidad no coinciden.

Como construir una definición de verdad es tarea de una teoría adecuada formulada en L_M , y no de L_M mismo, y como pudiera ocurrir que L_M fuera lo suficientemente expresivo como para representar el predicado de verdad de L_O , pero K_M no fuera lo suficientemente poderosa como para construir una definición de verdad, determinar las condiciones bajo las cuales un metalenguaje puede dar una definición de verdad para un lenguaje objeto implica considerar las condiciones bajo las cuales el predicado de verdad de L_O es representable en K_M . De modo que introducimos la siguiente definición de representabilidad:

Definición 2. Un conjunto $C \subseteq D_{LK}$ es *representable en K* (K -representable) si y sólo si existe una fórmula $\varphi(x)$ de L_K , tal que para todo $d \in D_{LK}$

1. Si $d \in C$, entonces $K \vdash \varphi(\langle\langle d \rangle\rangle)$
2. Si $d \in C$, entonces $K \vdash \neg\varphi(\langle\langle d \rangle\rangle)$

donde $\langle\langle d \rangle\rangle$ es el nombre de d en L_K . Resulta bastante claro que si un conjunto C es representable en K , entonces C es expresable en L_K . La converso, desde ya, no vale para toda teoría. De modo que el Teorema de Tarski tiene su dual para teorías:

Teorema de Tarski (Forma Abstracta) 2. *Para cualquier teoría K cuyo lenguaje L_K satisface las condiciones 1. y 2. del Teorema de Tarski (Forma Abstracta) 1., el conjunto T^* de códigos de las oraciones verdaderas de L_K no es K -representable*

Es una consecuencia de ambas formas del teorema que, si se desea dar una definición de verdad para un L_O que cumple con las condiciones de los mismos, esto debe hacerse desde una metateoría K_M en cuyo lenguaje es definible el predicado de verdad de L_O , y en la cual es además representable:

Definición 3. Tr es L_M -definible si y sólo si:

1. $U_{L_M} \models Tr(\langle\langle s \rangle\rangle) \equiv s$, para cada oración s de L_O , y
2. $U_{L_M} \models (\forall x)[Tr(\langle\langle x \rangle\rangle) \equiv x]$

Definición 4. Tr es K_M -representable si y sólo si:

1. $K_M \models Tr(\langle\langle s \rangle\rangle) \equiv s$, para cada oración s de L_O , y
2. $K_M \models (\forall x)[Tr(\langle\langle x \rangle\rangle) \equiv x]$

Finalmente, conjeturamos que una teoría K_M es esencialmente más rica que una teoría K_O cuando el predicado de verdad de L_O es K_M -representable.

II

Hemos dicho que DeVidi y Solomon [1999] sostienen que la noción de riqueza esencial es filosóficamente no interesante, por cuanto la única propiedad que todo par $L_O - L_M$ posee, tal que hace a la tesis de Tarski verdadera, es “salvajemente” disyuntiva.

Ahora bien, podría sostenerse que el requisito de K_M -representabilidad no es más que una reformulación trivial de la idea de poder dar una definición adecuada y correcta de verdad, y que por tanto oculta detrás de palabras técnicas una propiedad disyuntiva después de todo. Para mostrar que esto no es así, debemos considerar las diferentes maneras en que el concepto de verdad para L_O puede ser representado en una teoría K_M adecuada.

Tarski da una caracterización de lo que significa “riqueza esencial” en diversos lugares.² Ésta resulta ser también una caracterización de las condiciones bajo las cuales el predicado de verdad de L_O es K_M -representable. De estos pasajes parece seguirse que una teoría K_M es esencialmente más rica que una teoría K_O si K_M satisface alguna de las siguientes tres condiciones:

1. L_M es más rico en formas gramaticales que L_O (L_M posee variables de orden superior a cualquier variable de L_O)
2. K_M tiene más axiomas que K_O (K_M puede probar la existencia de clases

² Cfr. [Tarski, 1939, p. 110; 1944, n. 16].

que cuya existencia no puede probarse en K_O)

3. K_M admite formas de definición no disponibles en K_O (K_M admite definiciones recursivas)

Como todo lo que puede ser introducido por medio de axiomas nuevos puede ser introducido por medio de reglas de inferencia adicionales, una cuarta forma no contemplada explícitamente por Tarski viene dada por:

4. K_M admite reglas de inferencia no disponibles en K_O

Estas cuatro maneras de ser esencialmente más rica una teoría que otra son a su vez maneras diferentes en que un lenguaje o teoría puede ser *extendido*. Recordamos para ello que una teoría en sentido tarskiano es un par ordenado $K = \langle L, \Delta \rangle$, donde L es el lenguaje formal interpretado en el cual K está formulada, y Δ es un sistema deductivo para K , constituido a su vez por un conjunto de axiomas y reglas de inferencia.

La adición de formas gramaticales nuevas – tales como variables de orden superior – es una manera de enriquecer el lenguaje L en el que está formulada una teoría K , y por tanto una manera de volver expresable el predicado de verdad de L_O en L_M . Las otras maneras de enriquecer una teoría no son relativas a la expresividad del lenguaje, sino al poder de prueba de la teoría misma: son maneras de extender aquello que puede ser representado dentro de una teoría. Para tal fin, la nueva teoría K_M debe ser capaz de probar más cosas que las que era capaz de probar K_O , y esto puede lograrse o bien introduciendo nuevos axiomas, relativos a clases adicionales, cuya existencia no puede probarse en la teoría inicial; o bien por medio de reglas de inferencia de un nuevo tipo; o bien, finalmente, admitiendo nuevas formas de definición, en la forma de axiomas o reglas de inferencia apropiados.

De manera que la noción de K_M -representabilidad es disyuntiva, pero no es arbitrariamente disyuntiva: cada una de las maneras en que un lenguaje puede ser esencialmente más rico que otro es una manera en que una teoría formal interpretada puede ser enriquecida. Lo mismo vale, *pari passu*, para la noción de riqueza esencial.³

3 Vease [Ray, 2005] para esencialmente el mismo punto.

III

La segunda de nuestras preocupaciones es relativa a la estructura de la jerarquía. Intuitivamente, una jerarquía tarskiana es un ordenamiento lineal estricto de teorías bajo la relación *ser esencialmente más rico que*. Sin embargo, DeVidi y Solomon [1999] han argumentado que, bajo cualquier lectura que se haga de esta relación, no satisface la propiedad de asimetría. El argumento está montado sobre un supuesto contraejemplo que involucra dos teorías que, se sostiene, pueden darse mutuamente definiciones de verdad correctas y adecuadas. De un resultado ofrecido en [Wang, 1952] se sigue que es posible dar una definición de verdad para Z (la teoría de conjuntos de Zermelo) en P' o P , respectivamente, las extensiones predicativa e impredicativa de G , la teoría de conjuntos finitos. Por otra parte, DeVidi y Solomon argumentan que es posible dar una definición de verdad para P en Z mismo (los detalles aquí no son importantes).⁴ Desde luego, ya sabemos que si Z puede dar una definición de verdad para P , debe poder probar, en particular, todas las oraciones de P que son oraciones-T para Z , con lo cual, en un sentido preciso, Z debería poder dar la propia definición de verdad. Si asumimos que las definiciones de verdad de Z en P y de P en Z son ambas correctas y adecuadas definiciones de verdad, entonces tanto Z como P serían inconsistentes. De modo que el argumento de DeVidi y Solomon debe ser incorrecto. Dejaremos esta cuestión de lado por el momento.

Para mostrar que la jerarquía generada por el requisito de K_M -representabilidad constituye un ordenamiento lineal estricto sobre las teorías, debemos mostrar que la relación T_x es *y-representable* es una relación de orden total estricto sobre teorías, para lo cual debemos mostrar que satisface transitividad, irreflexividad, asimetría y tricotomía. Recordamos que transitividad e irreflexividad implican asimetría. Por su parte, la propiedad de tricotomía asegura que dos elementos cualesquiera del ordenamiento o bien sean comparables, o bien sean el mismo elemento.

Para mostrar que la relación cumple estas condiciones, apelamos a una peculiaridad de la idea de metalenguaje tarskiano. Es una consecuencia del Teorema de Tarski que la definición de verdad para L_0 debe darse desde una K_M adecuada para esta tarea. Uno de los requisitos que debe satisfacer todo

4 [DeVidi y Solomon, 1999] p. 22.

candidato a metalenguaje es que debe contener o bien a L_0 como un fragmento de L_M , o bien una traducción de L_0 en L_M . Ahora bien, dado que el teorema se aplica a lenguajes formales totalmente interpretados, una restricción razonable sobre la traducción es que no solamente preserve valor de verdad y teoremicidad, sino que preserve también interpretación.⁵ Esta preservación de interpretación puede reflejarse requiriendo que la estructura de L_M incluya la estructura de L_0 :⁶

$$(1) \quad U_{L_0} \subset U_{L_M}$$

Que la propiedad de transitividad se satisface parece seguirse bastante directamente de este requisito: sean L , L' y L'' tres lenguajes, y U_L , $U_{L'}$ y $U_{L''}$ sus respectivas estructuras. Supongamos que L' es esencialmente más rico que L , y L'' esencialmente más rico que L' . Dado (1) se sigue:

$$(2) \quad U_L \subset U_{L'} \subset U_{L''}$$

Dado que L'' contiene a L' , posee, en particular, todos los recursos necesarios para expresar el concepto de verdad de L . Para mostrar que además una teoría $K_{L''}$ formulada en L'' puede dar una definición de verdad para L , basta recordar que la traducción de L' a L'' preserva teoremicidad de $K_{L'}$.

Se trata de una relación claramente irreflexiva: esto es lo que se sigue directamente del teorema de Tarski.⁷ Ahora, transitividad e irreflexividad implican asimetría. De modo que parece satisfacer todas las propiedades de un orden lineal estricto, con excepción de la tricotomía. En lo que refiere a esta última, me parece relativamente claro que, por construcción misma de la jerarquía, la relación es tricotómica, en el siguiente sentido: esta propiedad asegura que cualesquiera dos elementos de un ordenamiento o sean comparables, o sean el mismo elemento; y esto parece asegurado por el hecho de que, para cada lenguaje L , construimos exactamente *un* lenguaje L' que pueda contener el predicado de verdad de L . Como consecuencia de la

5 Ray [2005] introduce este requisito para la noción de traducción, precisamente sobre la base de estar tratando con lenguajes formales plenamente interpretados. Fine y McCarthy [1984] adoptan implícitamente esta concepción de la relación entre L_M y L_0 al introducir su primer ejemplo.

6 Cfr. [Fine y McCarthy, 1984], p. 399.

7 En particular, apelando a (1), claramente $U_L \not\subset U_L$.

propiedad de transitividad, no ocurrirá que L'' , el lenguaje que da el predicado de verdad de L' , sea no comparable con L' .

Hemos dicho que DeVidi y Solomon [1999] proporcionan un ejemplo destinado a mostrar que, bajo cualquier lectura que se haga de ella, la relación *ser esencialmente más rico que* no satisface la propiedad de asimetría – con lo cual no puede ser considerada un orden lineal estricto sobre lenguajes. Dado que afirmamos que sí la satisface, debemos considerar en qué punto falla el argumento que ofrecen. En qué consiste el error surge inmediatamente de considerar la restricción que hemos impuesto a la noción de traducción del lenguaje objeto en el metalenguaje: si P tiene una definición materialmente adecuada y formalmente correcta de verdad para Z , entonces, por (1), tenemos:

$$(3) \quad U_Z \subset U_P$$

Por otra parte, si Z tiene una definición materialmente adecuada y formalmente correcta de verdad para P , entonces debemos tener:

$$(4) \quad U_P \subset U_Z$$

Claramente, estas dos condiciones no pueden ser satisfechas conjuntamente. De modo que alguna de ellas debe ser errónea.

Ahora bien, es posible probar en Z la existencia de conjuntos cuya existencia no es probable en P (en particular, es posible probar la existencia de conjuntos infinitos). De manera que, si U_P es la interpretación estándar de P , y U_Z la interpretación estándar de Z :

$$(5) \quad U_Z \not\subset U_P$$

De modo que (3) es errónea: si en algún sentido P tiene una definición de verdad para Z , esta última teoría debe estar interpretada de manera no estándar.⁸ Sea U_{Z^*} la estructura que da a Z su interpretación no estándar. Entonces se cumple:

⁸ La interpretación requerida se logra restringiendo las fórmulas de Z a $V_{\omega+1}$, con lo cual el universo de Z es restringido a elementos dentro de $V_{\omega+1}$.

$$(6) \quad U_{Z^*} \subset U_P$$

El argumento de DeVidi y Solomon supone en consecuencia una variación de la interpretación de Z a lo largo del mismo. Pero entonces se trata de dos lenguajes interpretados diferentes, y por tanto dos lenguajes diferentes en sentido tarskiano. En particular, tenemos:

$$(7) \quad U_{Z^*} \subset U_P \subset U_Z$$

Esto no es un caso de simetría, dado que P es esencialmente más rico que Z^* , mientras que Z es esencialmente más rico que P (y *a fortiori* que Z^*). Dicho de otra manera, Z y Z^* no son un mismo lenguaje con la admirable propiedad de ser esencialmente más rico que sí mismo, sino que son dos lenguajes diferentes (aunque sintácticamente idénticos), uno esencialmente más rico que otro.

Resumimos los resultados obtenidos. En primer lugar, introducimos la noción de K_M -representabilidad como forma de entender la idea de riqueza esencial. Segundo, intentamos mostrar que esta noción captura una propiedad suficientemente natural que poseen los pares $L_O - L_M$. Por último, intentamos mostrar que la relación de K_M -representabilidad del predicado de verdad de L_O es un orden lineal estricto sobre teorías. Concluimos indicando que es necesaria más investigación acerca del concepto de K_M -representabilidad en relación con los resultados de Tarski. Sin embargo, los resultados obtenidos parecen, aunque preliminares, apoyar esta noción como análisis de la idea de riqueza esencial.

Referencias

- [DeVidi y Solomon, 1999] DeVidi, D., Solomon, G.: "Tarski on "Essentially Richer" Metalanguages", *Journal of Philosophical Logic* 28: 1-28, 1999.
- [Fine y McCarthy, 1984] Fine, K., McCarthy, T.: "Truth Without Satisfaction", *Journal of Philosophical Logic* 13: 397-21, 1984
- [Gödel, 1931] Gödel, K.: "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I", *Monatshefte für Mathematik und Physik* 38: 173-98, 1931.
- [Ray, 2005] Ray, G.: "On the Matter of Essential Richness", *Journal of Philosophical Logic* 34: 433-457, 2006.
- [Tarski, 1935] Tarski, A.: "Die Wahrheitsbegriff in den formalisierten sprachen", *Studia Philosophica* 1: 261-405, 1935.
- [Tarski, 1939] Tarski, A.: "On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth", *The Journal of Symbolic Logic* 4: 105-112, 1939.
- [Tarski, 1944] Tarski, A.: "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research* 4: 341-37, 1944.
- [Wang, 1952] Wang, H.: "Truth Definitions and Consistency Proofs", *Transactions of the American Mathematical Society* 73: 243-275, 1952.