

MUNDOS NOCIONALES: ENCAJANDO LAS PIEZAS DEL ROMPECABEZAS DE FREGE*

Juan José Acero

Abstract

In this paper, an analysis of Frege's identity puzzle is put forward that does not make use of Fregean senses. The analysis proposes that identity statements (among others) can be informative in so far as their acceptance may potentially have significant effects on people's belief system. To be exact, informativeness is analyzed in terms of non-conservative extensions of sets of sentences. In the way of giving the knack of such an idea, an extensive use of Dennett's notion of notional world is made, aiming at the conclusion that different cognitive values for a subject *S* mean different individuals (properties, relations) in *S*'s notional world.

En estas páginas voy a ocuparme de un tema clásico de la filosofía del lenguaje: del rompecabezas fregeano de la identidad. Se me disculpará, quizás, que vuelva sobre algo tan debatido si adelanto que el análisis del rompecabezas que presentaré combina una posición antifregeana en semántica con una concepción computacional en teoría de la mente. El mostrar que una cuestión tan importante para la filosofía contemporánea pueda desembocar en semejante alianza teórica podría, a lo mejor, inducir al lector a pasar por alto mi atrevimiento.

El rompecabezas de Frege y la «nueva teoría de la referencia»

En un reciente artículo, H. Wettstein se pregunta si la semántica, tal y como ha venido concibiéndose en las últimas décadas no se habrá erigido sobre una condición de adecuación errónea. La condición que Wettstein pone en su punto de mira es la de la necesidad de proporcionar una explicación satisfactoria del *rompecabezas de Frege*

* Este trabajo forma parte del proyecto núm. PB87-0834-C03-02, financiado por la Dirección General de Investigación Científica y Técnica. Manuel García-Carpintero, Francisco Padilla y M^a José Frápolli han contribuido con sus comentarios críticos a precisar y reelaborar muchas de las ideas que aquí se exponen. No son, sin embargo, ni responsables de ellas ni necesariamente las comparten.

concerniente al valor cognitivo de los términos singulares. Este rompecabezas pone sobre el tapete de un modo particularmente agudo el problema de explicar cómo es que una oración como

(1) Terry Lennox es Cisco Maioranos

puede resultar informativa para alguien (que crea en su verdad), mientras que no vale decir lo mismo de

(2) Terry Lennox es Terry Lennox¹

Es, por decirlo de un modo equivalente al empleado, el problema de dar una explicación de por qué alguien puede no creer en lo que dice la primera oración aun creyendo en lo que dice la segunda. Efectivamente, hay aquí un problema, pues como todos sabemos Terry Lennox es Cisco Maioranos. El propio Frege propuso una solución cuya fuerza e influencia han sido notables. Según esa solución, una y otra oración, aun poseyendo la misma estructura sintáctica y difiriendo tan solo en lo relativo a uno de sus constituyentes, poseen diferente valor cognitivo. No expresan el mismo sentido, el mismo pensamiento. Eso se debe, dado lo que es común y lo que es específico de 1) y (2), a que los nombres propios «Terry Lennox» y «Cisco Maioranos» tienen sentidos o valores cognitivos distintos. Creer que Terry Lennox es Terry Lennox no es creer que Terry Lennox sea Cisco Maioranos.

En opinión de Wettstein, esta solución ha llevado a la semántica contemporánea a un callejón sin salida. Semejante situación se debe al hecho de que, por un lado, a lo largo de la última década se han ido acumulando uno tras otro diversos argumentos en favor de la idea de que ciertas expresiones de nuestro lenguaje -nombres propios, algunas expresiones deícticas- significan en la medida en que se limitan a referir a alguna entidad y, por otro lado, al hecho de que no parece haber forma de hacer encajar las piezas del rompecabezas de Frege con las instrucciones de la «nueva teoría de la referencia» (o teoría de la referencia directa). Puesto que intentos anteriores de resolver el puzzle, debidos a D. Kaplan y a J. Perry, le parecen a Wettstein insuficientes, propone que «llevemos a cabo una ruptura más radical (que la de Perry y Kaplan) con la perspectiva de Frege. El nuevo teórico debería rechazar del todo la condición de adecuación de Frege» (H. Wettstein, 1986, 200)².

Esta propuesta se halla, además, en consonancia con la diferente concepción de la semántica que sostuvo Frege y que ahora propugna el nuevo teórico de la referencia. A Frege le preocupaba, antes que otra cosa, la potencialidad del lenguaje para servir de vehículo expresivo al ámbito del pensamiento, un dominio autónomo e independiente del lenguaje. En esa tarea, la noción de sentido aglutina aspectos epistemológicos y lingüísticos, con un obvio predominio de los primeros sobre los segundos. De este planteamiento «se llega a una noción de sentido con una función teórica diferente de [la que poseen] las modernas nociones de significado», como lo ha diagnosticado magníficamente T. Burge³. El nuevo teórico de la referencia persigue sacar a relucir

¹ El ejemplo se basa en la clásica obra de Raymond Chandler *El largo adiós*. Hay versión española en Barcelona: Bruguera, 1981.

² «Has Semantics Rested on a Mistake?», *Journal of Philosophy*, vol. LXXXIII (1986) 185-209. Una respuesta a Wettstein es la que le da J. Perry en «Cognitive Significance and New Theories of Reference», *Noûs*, 22 (1988) 1-18.

³ T. Burge, «Sinning Against Frege», *Philosophical Review*, vol. LXXXVIII (1979), p. 399.

las reglas por las que se rige la práctica lingüística y su investigación es, entonces, una forma de antropología: «la *antropología* de esos acuerdos que constituyen el lenguaje natural» (H. Wettstein, 1986, p. 200).

Si se acepta la propuesta de Wettstein, los problemas del análisis semántico y los problemas del valor cognitivo pertenecen a esferas independientes. Ya no será un problema para la semántica la de determinar en qué consiste el valor cognitivo de una oración y en explicar cuándo ese valor no es trivial. Las oraciones describen situaciones y tienen condiciones de verdad; pero ésta es una cosa y la perspectiva cognitiva del hablante otra de un orden muy diferente. Del lado del mundo tendremos estados de cosas; del lado de la mente tendremos perspectivas cognitivas. Hablar de individuos, propiedades o estados de cosas es pertinente para el análisis semántico. Hablar de pensamientos o, en general, de valores cognitivos es sólo pertinente para el análisis de la mente, acaba diciendo Wettstein.

Mundos nocionales como perspectivas cognitivas

Si el rompecabezas de Frege es un dato para una teoría de la significación cognitiva, ¿a favor de qué o en contra de qué habla ese dato? Mi respuesta a esta pregunta es que una diferencia de valor cognitivo es siempre una diferencia de mundo nocional.

El término «mundo nocional» es un término recientemente acuñado, pero el concepto al que responde ha tenido una larga vida. Se ilustra suponiendo que hay alguien, un tal Philip Marlowe, para quien es cierto lo siguiente:

(3) Philip Marlowe cree que Cisco Maioranos es mexicano y suspendiendo el juicio a propósito de si la creencia de Marlowe es verdadera y del modo en que ha llegado a tenerla (por ejemplo, de si el procedimiento es epistemológicamente fiable o no). En vez de obrar así, *lo que hacemos es inferir o postular cómo sería el mundo en el caso de que la creencia de Marlowe fuese verdadera*. En consonancia con ello, habríamos de decir que en el mundo nocional de Marlowe habría alguien llamado «Cisco Maioranos» y que ese sujeto sería mexicano. Un mundo en que las creencias de Marlowe fuesen verdaderas podría no ser el mundo real. Es lo que D. Dennett ha denominado un *mundo nocional*, el tipo de entidad abstracta que obtenemos cuando «ponemos entre paréntesis» la vida doxástica de alguien (como Marlowe), despreocupándonos de su conexión con el mundo y tratamos de reconstruir las razones por las que se conduce como en realidad se conduce, aunque estas razones *puedan* ir por un camino y el mundo siga otro bien distinto. El mundo nocional de quien sea se postula sin suponer que existe otra entidad que el dueño de ese mundo. «Un mundo nocional debería verse como un tipo de mundo *de ficción* diseñado por un teórico, un observador al margen, para caracterizar los estados psicológicos restringidos de un sujeto»⁴.

⁴ D. Dennett, «Beyond Belief», en A. Woodfield, ed.: *Thought and Object*, Oxford University Press, 1982, p. 38. De estas palabras de Dennett la que más importa es el adjetivo «restringido» («*narrow*»). Hay que tomarlo con precaución. Putnam ilustraba la idea que motiva el uso de este adjetivo con el caso de alguien que tuviera celos de una persona... que podría no existir. Y la misma moraleja se encuentra en la

Mi convicción es que la noción del mundo nocional es de primera importancia en una teoría del valor cognitivo. Todos seguramente creemos en la verdad de (2), pero sólo algunos pocos creen en la verdad de (1). «Terry Lennox» y «Cisco Maioranos» tienen diferente valor cognitivo. Muy bien; pero ¿qué significa esto? Significa que una y otra creencia pueden llevarnos, como observadores marginales, a diseñar mundos nociónales diferentes entre sí. Si en el mundo nocional de Philip Marlowe hay alguien llamado «Terry Lennox» y le atribuimos a Marlowe la creencia expresada por (1), entonces es casi automático que también le atribuyamos que

(4) Marlowe cree que Terry Lennox es mexicano.

Pero si un hecho del mundo nocional de Marlowe es que Terry Lennox no es mexicano, entonces, como teóricos que deseamos preservar la posibilidad misma de ese mundo nocional, se nos planteará el problema de qué creencia preservar, si la que describe (3) o la que describe (4).

Llamemos *sistema de creencias de un sujeto S* al conjunto de enunciados (o de proposiciones) que describe el mundo nocional de S. La adición de (1) a un sistema lingüístico de creencias *puede alterar sustancialmente* la composición de dicho sistema. Esa alteración reflejará un cambio en el correspondiente mundo nocional. Sin embargo, la adición de (2) no tiene ningún efecto sobre el sistema (ni sobre el mundo). (Naturalmente, esta conclusión depende de que se asuma que la lógica clásica proporciona el metro que nos permite medir la existencia de esos efectos). Un potencial cambio en el mundo nocional, tal y como se concebía el mundo con anterioridad a la adición al sistema del enunciado de identidad no-trivial, frente a ningún cambio en absoluto.

La noción de cambio sustancial en un sistema de creencias puede hacerse más precisa. Sea Γ un sistema de creencias y sea $\Gamma * \alpha$ el resultado de añadir a Γ la creencia de que α . Finalmente, sea $\text{Con}(\Sigma)$ el conjunto de las consecuencias lógicas del sistema Σ . Entonces

Definición 1. La adición de la creencia de que α al sistema de creencias Γ *altera sustancialmente* Γ , si $\text{Con}(\Gamma)$ difiere de $\text{Con}(\{\Gamma * \alpha\}) \sim \{\alpha\}$.

Definición 2. Un enunciado E *tiene valor cognitivo no-nulo*, si existe algún sistema de creencias Γ tal que E expresa el pensamiento de que α y la adición de la creencia de que α al sistema Γ altera sustancialmente a éste.

Importa apereibirse de que tres párrafos más arriba he usado dos veces la forma verbal «puede» y una vez el adjetivo «potencial». Esto no es casual. Un sistema de creencias podría no contener ni el nombre «Terry Lennox» ni el nombre «Cisco Maioranos», en cuyo caso la adición de (1) no daría lugar a ningún cambio sustantivo en ese sistema de creencias. Lo que importa, sin embargo, es que pueda haber sistemas que sí sufran alteraciones sustantivas. Esos, sostengo, son los casos que importan. Cuando nos encontramos en uno de esos casos, es preciso hacer alguna revisión en

historia del aventurero que en medio del desierto ve un oasis... donde no lo hay. Con «restringido» se desea transmitir la idea de un estado mental que no es fiel índice de un estado de cosas extra-mental. Pero al hablar de mundos nociónales, casos como éstos habrán de ser seguramente casos extremos, pues las creencias que atribuimos a una persona dependerán en gran medida del tipo de nexos con su medio externo que apreciamos. La cuestión difícil es, así pues, en qué medida podemos despreocuparnos de las conexiones del sujeto con el mundo.

nuestro sistema conceptual. Y si esa revisión tiene como resultado el que nuestro mundo nocional se aproxime algo más al mundo real, queda más que justificado que se considere que la creencia que desencadenó el cambio proporcionaba información acerca del mundo. Y *a fortiori* que posee un valor informativo probado. Eso ocurre con los enunciados de la forma « $t_i = t_j$ » (donde « t_i » y « t_j » son términos singulares cualesquiera, pero no ocurre con los enunciados de la forma « $t_i = t_i$ ».

La objeción positivista

Consideraré a continuación dos objeciones al análisis precedente. La primera se basa en una premisa que está muy extendida y dice esto: la diferencia de valor cognitivo entre (1) y (2) resulta de la diferencia de valor cognitivo entre dos términos singulares, «Terry Lennox» y «Cisco Maioranos». Pero el análisis precedente resuelve una diferencia de valor cognitivo en una diferencia puramente formal o sintáctica. Que los términos en cuestión posean diferente valor cognitivo significa que hay alguien, Marlowe, por ejemplo, para quien uno y otro término llevan asociados diferentes items de información o diferentes formas de pensar en un individuo (que posiblemente no es el mismo en cada caso). Esto parece tener que ver con hechos relativos a cómo un sujeto *entiende* ambos términos. Y si esto no es semántica en sentido estricto -como Wettstein nos pide que aceptemos-, no es sintaxis tampoco, en ningún caso.

Esta objeción es característica de quien todavía no ha logrado desembarazarse de algunos tics positivistas sobre la naturaleza de la lógica. Mi diagnóstico sobre las diferencias entre (1) y (2) se basa, si así se desea expresarlo, en ciertas diferencias de índole formal que existente entre (1) y (2). La unión de un sistema de creencias Γ -y dejo aquí abierta la cuestión de si un sistema así ha de estar cerrado bajo la relación de consecuencia lógica- *más* (2) es siempre una *extensión conservadora* de Γ . Pero no siempre sucedería lo mismo en el caso de (1), pues ahora sí que importa la particular composición de Γ . Así que aquí tenemos un análisis preciso de la diferencia entre ambas oraciones y, por ello, de la diferencia de valor cognitivo entre los términos «Terry Lennox» y «Cisco Maioranos», pues hay un Γ , el de Philip Marlowe, en donde esto no sucede. Si ahora se objeta que esa diferencia se ha plasmado en términos puramente formales o bien -como diría un positivista- haciendo uso de un conjunto de reglas de transformación, por lo que se ha puesto de relieve únicamente una diferencia sintáctica, entonces el objetor apela a la premisa de que invocar diferencias en los resultados del uso de un conjunto de reglas de transformación es invocar diferencias exclusivamente sintácticas. Y eso porque las reglas de transformación son reglas sintácticas. Si eso mismo se dijera de las reglas que los manuales de hace algunas décadas denominaban reglas de formación -es decir, las reglas que definen recursivamente nociones como término o fórmula-, la objeción sería correcta, a mi modo de ver. Pero las llamadas reglas de transformación son otra historia. Estas reglas explotan diferencias formales, obvio es decirlo, pero esto no es en sí mismo un argumento contra nada. No lo es *porque diferencias formales pueden corresponder* en el sistema de

creencias a *diferencias en el mundo nocional de un sujeto*, es decir, a diferencias en el modo de concebir individuos o en el modo de entender expresiones⁵.

La hipótesis del «lenguaje del pensamiento» y el punto de vista heterofenomenológico

El mundo nocional de un sujeto es, según lo visto, un modelo de las creencias de dicho sujeto. Lo que yo denominado sistema de creencias constituye una representación de dicho mundo, pero no estoy sugiriendo que se identifiquen las creencias de nadie con un conjunto de entidades de naturaleza lingüística. En particular, no estoy adhiriéndome aquí a la tesis de que las actitudes proposicionales (y en particular las creencias) sean relaciones con fórmulas de un sistema de representación mental⁶. (No deseo tampoco excluir esa posibilidad, sino adoptar una posición neutral al respecto). Traigo a colación el concepto de mundo nocional, a fin no sólo de decir qué es lo que alguien cree, sino también para disponer de un modo útil de explicar qué es lo que sucede cuando las creencias de esa persona cambian, cuando nuevas creencias se añaden a otras que ya se tenían. En unas ocasiones, el mundo nocional perderá alguna de su borrosidad o indeterminación al ganar en información, es decir, al incorporarse a él nuevos miembros; en otras, el cambio consistirá en la eliminación de algunos de sus ingredientes anteriores y en su sustitución por otros nuevos. El problema surge cuando los sistemas de creencias se analizan como sistemas de oraciones (o de algo análogo a oraciones o a conjuntos de oraciones, como redes semánticas o marcos). En tal caso, es decir, dentro del (mal) llamado enfoque *sintáctico*, la descripción de esos procesos debe tomar nota de las transformaciones sufridas por un conjunto de *oraciones* (o de fórmulas de algún sistema canónico o expresiones de un sistema de representación).

La objeción que suele hacerse a semejante punto de vista es que afina demasiado. Puesto que el enunciado α difiere del enunciado $\alpha \& \alpha$, o del enunciado $(\alpha \vee \beta \& (\alpha \& \sim \beta))$, por poner dos ejemplos, un sistema que contuviera al primero pero no a los otros dos sería distinto del que constara de todos esos (o de sólo dos de los) enuncia-

⁵ El desliz que comete la objeción citada es el que cometen los que arremeten contra la IA fuerte diciendo que los computadores tienen sintaxis, pero ¿no? semántica. Cf. J. Searle, «Minds, Brains and Programs», *Behavioral and Brain Sciences*, 3 (1980), pp. 417-424; *Intentionality* (Cambridge University Press, 1983), cap. 10; *Mentes, cerebro y ciencia* (Madrid, Cátedra, 1985), cap. 2. La respuesta a esto consiste en decir que si los ordenadores tienen el tipo de sintaxis que tienen, entonces también tienen *alguna* semántica. El problema no es si sólo tienen sintaxis, sino si poseen la semántica necesaria. A mi modo de ver, Fodor entendió esto desde el principio del debate sobre el solipsismo metodológico. Véase su respuesta a Searle: J. Fodor, «Searle on What Only Brains Can Do», *Behavioral and Brain Sciences*, 3 (1980), pp. 431 y ss. En psicología cognitiva, sin embargo, la idea de que diferencias formales desempeñan una función representacional viene de antiguo. En el reciente *mental model approach*, por ejemplo, ésta es una tesis que se asume explícitamente como un ingrediente del concepto de modelo mental. Cf. P. Johnson-Laird, *Mental Models, Towards a Cognitive Science of Language, Inference, and Consciousness* (Cambridge University Press, 1983), p. 156.

⁶ La tesis se expone en los ensayos de la parte III de la obra de J. Fodor, *RePresentations*, Brighton, Sussex: Harvester Press, 1981.

dos. Sin embargo, se dice, no es obvio cuál pueda ser la motivación semántica en que fundar una elección entre uno y otro. Si lo importante son los mundos nocionales mismos, la búsqueda de principios que pongan limitaciones a conjuntos de oraciones puede parecer completamente *ad hoc*. La objeción suele tenerse presente en contextos en que deseamos distinguir entre creencias *explícitas* y creencias *implícitas*, es decir, entre creencias que efectiva y *activamente* tenemos y creencias que sería razonable que se nos atribuyera, dadas las primeras, por seguirse lógicamente de éstas. La objeción viene a señalar, entonces, que no parece haber un modo justificado de efectuar semejante distinción⁷.

Mi actitud hacia esa objeción es ambivalente, y eso depende de si se acepta o no *identificar* sistemas de creencias con conjuntos de enunciados (u oraciones). De hacerlo, mi respuesta a la objeción sería la siguiente. El objetor asume que las reglas por las que se rige el cambio en un sistema de creencias son los principios de inferencia lógica y que esos principios se aplican bajo el principio de salvar la consistencia lógica del sistema. Asume también que los sistemas de creencias están cerradas bajo la relación de deducibilidad⁸. Ambas cosas, sin embargo, no sólo nos llevan a una idealización excesiva; son también seguramente un error. Los principios de inferencia lógica, ha argüido recientemente G. Harman⁹, no son los principios que regulan el cambio en nuestros sistemas de creencias y deseos. Si nuestro objetivo es proporcionar una descripción del mundo nocional de una persona, entonces es un problema empírico determinar por qué principios se rigen los cambios que ese mundo nocional sufre -es decir, qué principios son principios de razonamiento-. En este sentido, por lo tanto, hay razones para pensar que un sistema de creencias puede entenderse como un sistema lingüístico, y que cuando hablamos de la adición de un nuevo elemento al sistema -como yo hacía antes- estamos tomando una opción analítica que no es tan fácilmente refutable. Se trata, al menos, de una opción que tiene entre sus méritos el de proponer criterios de identidad perfectamente concretos para sistemas de creencias y el proponerse proporcionar descripciones igualmente detalladas de los cambios de dichos sistemas.

De otro lado, entiendo que el enfoque *sintáctico* no es en absoluto incompatible con una concepción *semántica* del análisis de los sistemas de creencias, una vez que se lo entiende tras las precauciones tomadas. Según ese punto de vista, un mundo nocional se identifica con un conjunto de situaciones. Una *situación* se concebiría, entonces, como un mundo posible parcial; es decir, como una entidad en virtud de la cual sólo una parte de las oraciones del lenguaje recibe un valor veritativo¹⁰. Y no existe

⁷ Véase, por ejemplo, H. Lebesque: «A Logic of Implicit and Explicit knowledge», *Proceedings of the National Conference on Artificial Intelligence*, Austin: University of Texas, 1984.

⁸ Esos supuestos se adoptan en ocasiones, en los estudios de la dinámica de las creencias, por razones de sencillez teórica. Eso ocurre, por ejemplo, en P. Gärdenfors, *Knowledge in Flux*, Cambridge, Mass: The M.I.T. Press, 1988.

⁹ Véase especialmente los dos primeros capítulos de *Change in View*, Cambridge, Mass, The MIT Press, 1986

¹⁰ Esta perspectiva es la característica de la hoy ya asentada *Semántica de situaciones*, inicialmente propuesta por J. Barwise y J. Perry. La aplicación de esta idea a los estudios de representación del conocimiento, como alternativa a los enfoques sintácticos y semánticos clásicos, se propone por ejemplo en H. Lebesque, «A Logic of Implicit and Explicit Belief», *loc. cit.*

incompatibilidad entre ambos puntos de vista porque cualquiera de las dos siguientes opciones es perfectamente asumible: (i) Aceptamos como sistema de creencias un conjunto de oraciones, no cerrado por la relación de deducibilidad, y postular que el mundo nocional está formado por las situaciones que harían verdaderas a las oraciones del sistema, y sólo a ellas. En este caso, la cuestión central consiste en determinar qué tipos de situación se considerarán interpretaciones aceptables de las oraciones del conjunto. (ii) Aceptar que un mundo nocional está formado por un determinado conjunto de situaciones y postular que su descripción estaría formada por un conjunto de oraciones que serían verdaderas en virtud de tan solo las situaciones del mundo nocional. A mi modo de ver, es indiferente el enfoque que adoptemos.

Supongamos ahora, en segundo lugar, que recusamos la anterior identificación. Pese a todo, entiendo que es intuitivamente útil pensar en términos de los efectos que tiene sobre un conjunto de oraciones, cuya consistencia lógica manifiesta deseamos preservar, la adición de otra u otras. Obtenemos así una reconstrucción precisa de la noción de significado cognitivo, y éste es un objetivo que tiene un valor en sí mismo. Si así lo hacemos, adoptamos una actitud filosófica minimalista en sus compromisos: la del heterofenomenólogo de Dennett¹¹. Nuestro punto de vista es el de una *tercera persona* cuya tarea consiste elaborar un texto que interpretar, a partir de la conducta lingüística de un sujeto -posiblemente un zombie-, tal y como ésta resulta de someter a dicho individuo a entrevistas o a situaciones experimentales en que ha de resolver rompecabezas o emitir juicios sobre una variedad de situaciones o problemas diferentes. El texto no se entiende como una sucesión de signos sin interpretar, sino como el vehículo con el que ese sujeto lleva a cabo actos de habla, con el que expresa sus estados mentales (deseos y creencias) y anuncia sus razones para ello. Desde ese punto de vista, (1) y (2) pueden ser parte de un texto que se sujeta a una interpretación *intencional*, y nuestra interpretación consiste en especificar desde fuera el mundo del sujeto heterofenomenológico. Esto no nos sitúa en una posición incómoda: «Del mismo modo que el crítico literario o el hermeneuta de la ficción aceptan que el texto *constituya* un mundo (de ficción), un mundo determinado por fiat por el texto..., así nuestro experimentador, a quien denominaremos el heterofenomenólogo, acepta que el texto del sujeto constituya lo que llamaremos el mundo del sujeto heterofenomenológico»¹². Aceptamos por fiat que hay algo que las oraciones de Marlowe describen, un mundo nocional, y simplemente tratamos de describirlo. Pero la descripción es una cosa -un texto-; y lo descrito, otra -una ficción.

¿Qué hacer, entonces, con la identificación a que he venido refiriéndome? No tomaré partido al respecto. Sugiero que en adelante obremos dialécticamente: que examinemos cada paso del hilo argumental de una forma y, una vez hecho esto, que adoptemos el otro punto de vista. Por mi parte, intercalaré uno y otro sin más aviso.

¹¹ D. Dennett, «How to Study Human Consciousness Empirically or Nothing Comes to Mind», *Synthese*, 53 (1982), pp. 159-180.

¹² *Loc. cit.*, p. 166.

Valor cognitivo (o la utilidad de la identidad)

El análisis precedente permite resolver el rompecabezas de Frege sin necesidad de invocar ningún tipo de valor semántico intensional. Y, por lo tanto, sin reivindicar la noción semántica de sentido. Se explica la diferencia entre (1) y (2), pero no porque se sienta la necesidad de asociar un sentido con cada nombre propio, sino por el diferente efecto potencial que sobre algún conjunto de creencias pueda ejercer la adición de una creencia de una u otra forma. En particular, una creencia de la forma que posee (2) es tautológica porque, en un sentido, no tiene para ningún sistema consecuencias que puedan considerarse propias. Si nos preguntamos qué supondría para un sistema de creencias la adición de (2), la respuesta habrá de ser que el sistema poseerá después un nuevo elemento, si antes no hubiésemos considerados incluidos en el sistema todas las oraciones deducibles de los miembros del conjunto. Y si se considera al sistema de esta otra manera, entonces el cambio sufrido es inocuo; el sistema no dispondrá de nuevas creencias que no sean verdades lógicas.

No hay duda, por otra parte, que existe entre (1) y (2) una diferencia notable en el respecto citado. Una parte del mundo nocional de Philip Marlowe en *El largo adiós*, en un momento dado de la narración, puede describirse por medio del siguiente conjunto de oraciones:

MUNDO NOCIONAL I:

Universo: Terry Lennox, Randy Starr, Cisco Maioranos, etc.

Hechos:

- (a) La policía americana persiguió a Terry Lennox
- (b) Terry Lennox huyó a México
- (c) Randy Starr asegura que Terry Lennox ha sido asesinado
- (d) Terry Lennox era norteamericano
- (e) Cisco Maioranos trae noticias sobre la presunta muerte de Terry Lennox
- (f) Cisco Maioranos es mexicano
- (g) No hay pruebas convincentes de la muerte de Terry Lennox
- (h) Cisco Maioranos está vivo

Sin embargo, en el desenlace de la historia se le hace evidente a Marlowe que la oración (1) es verdadera, y *esto*, es decir, la incorporación de (1) al sistema de creencias de Marlowe tiene efectos de peso sobre su mundo nocional. No es sólo que algunas de las creencias acerca de Terry Lennox se puedan aplicar ahora a Cisco Maioranos -lo que es cierto para (b), (d) y (g)-. Lo realmente importante para mi argumentación es que, además de eso, el universo del mundo nocional resultante se ha reducido: Terry Lennox y Cisco Maioranos son la misma persona. Es decir, tras la adición de (1) el MUNDO NOCIONAL I se transforma en el

MUNDO NOCIONAL II:

Universo: Terry Lennox [= Cisco Maioranos], Randy Starr, etc.

Hechos: (a), (b), (c), (d), (e), (g), (h)

(i) Terry Lennox es Cisco Maioranos

(j) Terry Lennox está vivo

(k) La policía americana no persigue a Cisco Maioranos

Este ejemplo ilustra los que son, a mi modo de ver, los dos tipos de efectos no triviales que puede tener sobre un sistema de creencias la adición de un enunciado de identidad no-trivial. Puede permitirnos inferir *nuevas* creencias no-tautológicas de algunas que ya se poseían, como sucede con el paso de (h) a (j) por la acción intermedia de (i). Cuando esto sucede, el mundo nocional cambia. Antes de la adquisición de (i), el mundo nocional de Marlow lo podíamos considerar formado por el conjunto de todos los mundos posibles (parciales) compatibles con el MUNDO NOCIONAL I. Y aquí habría mundos en los que Terry Lennox estuviese muerto y mundos en los que estuviese vivo. Este conjunto se reduce sustancialmente después: los únicos mundos relevantes serán aquellos en los que Terry Lennox está vivo.

El segundo tipo de efecto es también importante: el universo del mundo nocional ha disminuido. Cuando construimos el nuevo sistema de creencias de Marlowe a partir del antiguo, no distinguimos tres individuos -Terry Lennox, Randy Starr y Cisco Maioranos- entre otros, sino uno menos. Cambia el mundo nocional porque cambia el universo. En semejante respecto es la identidad útil.

Valor cognitivo como rol conceptual

En un sentido bastante inmediato, la explicación que he dado de las diferencias de valor cognitivo entre (1) y (2) es anti-fregeana. No es necesario postular que los términos singulares tengan necesariamente un sentido, además de poseer contingentemente un referente. La explicación sería igualmente válida si sostuviésemos que un nombre propio se limita a actuar de representante de un objeto (de su referente), pues la capacidad de un enunciado de identidad para promover cambios sustanciales en un sistema de creencias es independiente de qué especie de término singular pueda ir a un lado de la forma verbal «es». Todo lo que la explicación requiere es que la forma verbal vaya flanqueada por términos singulares diferentes¹³. (Más abajo doy el porqué de esta condición). El argumento fregeano que conduce desde la diferencia de valor cognitivo de (1) y (2) a la conclusión de la existencia de dos tipos de valores semánticos es un argumento con una fuerza inferior a lo que quizás podría parecer.

¹³ Ni siquiera esto es suficiente. El mismo término singular, entendido como término-tipo, puede dar lugar a un cambio sustancial en un sistema de creencias. A este respecto tiene gran interés el ejemplo de «Paderewski» que expone S. Kripke en «A Puzzle About Belief», en A. Margalit, ed.: *Meaning and Use*, Dordrecht: D. Reidel, 1979, pp. 265 y ss.

El misterio del valor cognitivo no reside de forma exclusiva, por consiguiente, en los términos singulares. También la identidad importa. La identidad es un concepto cuyo contenido, como diría Harman¹⁴, se agota en el peculiar papel que desempeña en la argumentación deductiva y en el razonamiento. Para determinar su contenido, todo lo que hay que hacer es saber en qué inferencias deductivas y en qué pautas de razonamiento sirve de puente hacia estadios posteriores de la demostración o del razonamiento: cómo interviene, especialmente, en un cambio en nuestro sistema de creencias o en un cambio en nuestros objetivos o planes. Una teoría sobre el rol del concepto de identidad es, así pues, una teoría del rol conceptual de dicho concepto, una teoría acerca del modo en que la identidad contribuye a cambiar un mundo nocional. Ciertos enunciados de identidad no pueden introducir cambio alguno, pero otros sí que pueden hacerlo. Lo que vimos hacer al nuevo teórico de la referencia es negar que una teoría del rol conceptual sea una teoría semántica. Es una teoría de la perspectiva cognitiva, pero no una teoría semántica.

El rompecabezas de Frege y la hipótesis de la Representación del Conocimiento

Si la solución que Frege halló al rompecabezas que lleva su nombre *no* es vinculante, eso se debe a que hay otra forma de hacer encajar las piezas que no obliga a aceptar su distinción entre el sentido de una expresión y su referente. Esa otra explicación de cómo es que (1) tiene un valor cognitivo no-nulo, mientras que (2) carece de uno tal, es aceptable siempre que la consideremos inseparable de los dos siguientes principios: (i) a cada nombre propio del sistema de creencias le corresponderá un individuo -su referente- en el universo del mundo nocional; (ii) a cada dos *nombres* diferentes del sistema corresponderán dos *individuos* diferentes en el universo del mundo nocional. Es porque aceptamos algo como (i) y (ii) *en nuestra práctica heterofenomenológica* por lo que el mundo nocional de Marlowe sufre una modificación sustancial. Es decir, es porque tácitamente asumimos que una creencia sobre *a* y una creencia sobre *b* son creencias sobre individuos diferentes por lo que (1) es, como se diría, en una jerga filosóficamente más clásica, un enunciado *sintético*: el mundo nocional de Marlowe distinguía dos individuos, cada uno con su propio acopio de propiedades y relaciones, pero luego resultaban ser el mismo. Avanzo, por lo tanto, la hipótesis de que los cambios en los mundos nocionales que hayamos de postular, llevados de nuestra actividad hermenéutica, se rigen por una máxima (entre tantas otras) que dice lo siguiente: Todo individuo es distinto de los demás hasta que no se demuestre lo contrario.

Siendo esto así -como yo creo que es-, lo sorprendente del caso es lo bien que encaja este diagnóstico de las razones por las que el rompecabezas de Frege es, efectivamente, un rompecabezas con la hipótesis central que guía los estudios de IA,

¹⁴ Esta es la opción denominada por Harman *Semántica no solipsista del rol conceptual*. Cf. G. Harman, «(Non-Solipsistic) Conceptual Role Semantics», en L. Lepore, ed.: *New Directions in Semantics*, New York: Academic Press, 1987.

la *Hipótesis de la Representación del Conocimiento* (RC): cualquier sistema que pueda ejecutar algún proceso inteligente (como un proceso de razonamiento, búsqueda, etc.) es un sistema que hará uso de representaciones, las cuales desempeñan un papel causal esencial en la conducta del citado sistema¹⁵. Esta hipótesis asigna a las representaciones dos características que son especialmente importantes. Las representaciones se conciben como los elementos de un sistema que guarda una conexión sistemática con aquello acerca de lo cual el sistema infiere, busca, etc. Si la mente es un sistema así, entonces la mente trabaja con representaciones (parciales, fragmentarias) de su mundo, y es en virtud de esa conexión sistemática entre las representaciones y lo que representan que podemos decir que Marlowe reconoce a Terry Lennox cuando tiene ante sí a Cisco Maioranos o por lo que le podemos atribuir que recuerda que acompañó a Terry Lennox hasta el avión que le permitió escapar de la persecución policial. Porque existe una conexión sistemática es por lo que, sigue diciendo RC, podríamos apuntar a una representación del sistema «y decir cómo tendría que ser el mundo para que fuese verdadera»¹⁶.

El segundo rasgo de las representaciones es el que aquí más nos interesa. Un sistema inteligente trabaja con representaciones, pero los cómputos que hace con ellas son completamente independientes de los valores semánticos que puedan tener aquellas por mor de la conexión sistemática a la que me he referido en el párrafo precedente. Las representaciones desempeñan la función causal que desempeñan al margen de cualquier tipo de relación semántica entre ellas y lo representado por ellas. Las representaciones son expresiones de un lenguaje *formalizado* (como diría Putnam). Esta condición de formalidad caracteriza la concepción de la mente del solipsismo metodológico, la Teoría Computacional de la mente. Aplicada al caso de los seres humanos, esa teoría postula que «dos pensamientos pueden tener distinto contenido sólo si se los puede identificar con relaciones con representaciones formalmente distintas»¹⁷. Dos pensamientos son el mismo pensamiento si sus relaciones con cualesquiera representaciones son idénticas.

La salida al rompecabezas de Frege que he expuesto es solidaria del todo de la hipótesis RC. Al menos dos razones avalan la conclusión de que esa salida queda favorablemente sancionada por la condición de formalidad. Una de esas razones es que el hecho de que dos nombres del sistema de creencias se correspondan con dos individuos diferentes del mundo nocional, si los nombres son ellos mismos diferentes

¹⁵ La formulación clásica de esta hipótesis dentro de los estudios de AI se debe a Brian C. Smith. Cf. «Prologue to *Reflection and Semantics in a Procedural Language*», en R. J. Brachman & H. J. Lebesque, eds.: *Readings in Knowledge Representation*, Los Altos, California: Morgan Kaufmann Publishers, 1985. En filosofía, la Hipótesis de la Representación del Conocimiento aparece dentro de la llamada Teoría Representacional de la mente y, más en particular, en la de esa variante de la teoría representacional conocida como Teoría Computacional de la mente. El *locus classicus* es J. Fodor, «Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology», en *RePresentations*, véase más arriba. Una interesante exposición y discusión de la hipótesis de la representación del conocimiento es la que hace H. Putnam en su ensayo «Computational Psychology and Interpretation Theory», en R. Born: *Artificial Intelligence: The Case Against*, London: Croom Helm, 1987.

¹⁶ H.J. Lebesque & R. J. Brachman: «A Fundamental Tradeoff in Knowledge Representation», en R.J. Brachman & H.J. Lebesque, eds.: *Readings in Knowledge Representation*, Los Altos, California, Morgan Kaufmann Publishers, 1985, p. 44.

¹⁷ J. Fodor, *op.cit.*, p. 227.

significa que lo decisivo del caso no estriba en el valor semántico del nombre, sino en que los nombres no sean el mismo nombre. En efecto, lo que hace de (1) una oración dotada de un valor cognitivo no-nulo es su *forma*: diferentes nombres, en un caso; el mismo nombre en el otro. La segunda razón es que creencias como las que expresan (1) y (2) tienen contenido diferente porque sus relaciones con el resto de los miembros del sistema serán potencialmente diferentes. En la medida, por lo tanto, en que sea lícito dar cuenta de las diferencias de valor cognitivo entre (1) y (2), en tanto que diferencias en los sistemas de creencias que su aceptación pueda originar, la evolución de estos sistemas se rige por una condición de formalidad.

Si un sistema de creencias *es* un conjunto de expresiones (o de representaciones), entonces he argüido en estas páginas que su dinámica se rige (en parte al menos) por la condición de formalidad. Y si un sistema de creencias es el texto que el heterofenomenólogo trata de interpretar construyendo para él un mundo nocional, entonces la interpretación también se rige por una condición de formalidad como la citada.

Predicados

Hasta el momento me he ocupado de examinar el rompecabezas de Frege y de diagnosticar su origen. Obsérvese, sin embargo, que el rompecabezas en cuestión no surge de forma exclusiva para los nombres propios (o de sus análogos representacionales). Una explicación paralela a la ofrecida para el caso de (1) y (2) puede también darse para predicados como «húngaro» y «magiar». Así como

(5) Los húngaros y sólo ellos son magiares

tiene valor cognitivo efectivo,

(6) Los húngaros y sólo ellos son húngaros

no tiene uno no-trivial. También aquí la clave del análisis reside en la hipótesis RC y, de forma más específica, en la condición de formalidad.

Conocimiento incompleto

Si la condición de formalidad se aplica a los nombres propios, entonces, desde el punto de vista de la perspectiva cognitiva, en un sentido muy literal un nombre propio se comporta como una frase de cuantificación existencial de una lengua natural o como una variable ligada por un cuantificador existencial en una teoría clásica de primer orden. En efecto, si ante expresiones de la forma

(7) X -algún Y que Z- W

(8) (Ex) (... x ...)

se nos pide que supongamos la existencia de algún individuo que satisfaga las condiciones del caso, lo que haremos será asumir que

(9) X -c- W, c es un Y y cZ

(10) (... a...)

donde «c» y «a» son parámetros o nombres nuevos (en algún sentido en el que ahora

no entraré). Las razones técnicas para imponer estas exigencias son bien conocidas. Una frase de cuantificación o una variable ligada por un cuantificador siempre introducen individuos nuevos por hipótesis. Y si, finalmente, no resultaran serlo, eso se debería al hecho de que podemos identificar ese presunto nuevo individuo con otro ya mencionado o descrito. De suceder eso, se debería a otra información disponible independientemente.

Esta analogía es, a mi juicio, lo suficientemente fuerte como para preguntarnos si hay algo más por debajo de ella. Sabemos por los estudios de IA concernientes a las condiciones de representación del conocimiento que el uso de la lógica de primer orden resulta especialmente oportuno por su flexibilidad para analizar *conocimiento incompleto*; que la capacidad expresiva de un lenguaje de primer orden no sólo determina qué cosas pueden decirse con él «sino lo que puede quedar sin decir»¹⁸. Esto sugiere algo que, por otro lado, era de esperar: que nuestros sistemas de creencias se forjan y evolucionan bajo condiciones en las que nuestro conocimiento es (algo, muy) incompleto. Esto es, al final, lo que hace del rompecabezas de Frege un problema tan interesante. Dentro de una particular perspectiva cognitiva de Marlowe, Terry Lennox y Cisco Maioranos son sujetos distintos. Marlowe ignora que Randy Starr ha acordado con Lennox fingir el asesinato de éste y cambiar luego su apariencia física y su identidad legal. Su perspectiva cognitiva es la que es por lo que Marlowe ignora.

Si queremos estudiar la perspectiva cognitiva de una persona, es decir, si deseamos saber cuál es su mundo nocional, habremos de regirnos, es mi hipótesis, por una condición de formalidad y por el principio de que a nombres (y predicados) distintos corresponden individuos (y propiedades o relaciones) diferentes del mundo nocional. Si dispusiéramos de un programa de ordenador que simulara cómo cambia el sistema de creencias de alguien tras la incorporación de enunciados de la forma de (1) o de (6), habríamos dado un paso importante para saber más cosas acerca de lo que una perspectiva cognitiva es en su sustancia.

Juan José ACERO
Universidad de Granada

¹⁸ Cf. R. J. Brachman & H.J. Lebesque, *op.cit.*, sección 4.1. y ss.; cf. también R.C. Moore: «The Role of Logic in Knowledge Representation and Commonsense Reasoning», en R.J. Brachman & H.J. Lebesque, eds.: *Readings in Knowledge Representation*, *loc. cit.*, especialmente p. 338.