

El significado y la fijación de la creencia

Hilary PUTNAM
Universidad de Harvard

Si hay un enfoque característico en la manera en que el siglo XX trata los temas de significado y referencia, éste es que estos problemas no pueden ser discutidos al margen de la idea de fijación de la creencia como un todo. Llamar a este enfoque del «siglo XX» no es reclamar una prioridad temporal para nuestros contemporáneos: Kant ciertamente vio estos problemas como profundamente interconectados, y también William James, la mayor parte de cuyo trabajo apareció en el siglo XIX. Pero el alcance de esta interconexión ha sido enfatizado en un grado sin precedentes por el filósofo americano del siglo XX Willard Van Quine y sus ideas son hoy incluso más centrales para la discusión que lo fueron cuando su famoso artículo «Dos Dogmas del Empirismo» apareció en 1950.

Tradicionalmente, la fijación de la creencia racional se piensa como un tema a ser dividido entre la lógica deductiva y algo así como un tenue asunto denominado optimistamente «lógica inductiva» (Por otra parte, C. S. Pierce, el creador del pragmatismo, separó la construcción de la teoría de la lógica inductiva bajo el nombre de «abducción»). Quine no supone que la lógica inductiva o la «inferencia abductiva» de Pierce realmente existan como métodos definitivos que pueden ser formalizados; lo que él acentúa es la informalidad de los procedimientos en base a los cuales se elige, en contextos diferentes, entre defender lo que creemos son nuestras observaciones y defender las teorías o principio teóricos que cultivamos. Las ideas centrales son que la «revisión puede aparecer en cualquier sitio» —un falibilismo que Quine puso honradamente en la tradición de Pierce y Dewey— y que las creencias de alguien sobre cualquier tema pueden llegar a ser relevantes en la fijación de las creencias sobre cualquier otro tema. (Es fácil dar ejemplos tomados de la historia de la ciencia: conside-

remos cómo el inesperado ennegrecimiento de una placa fotográfica en el laboratorio de Pierre Curie se convirtió en algo relevante para la posibilidad de la bomba atómica, o —después de que la existencia y naturaleza de la radioactividad hubiese sido investigada y aplicada para determinar la edad de los materiales orgánicos e inorgánicos *vía* la técnica de la fecha del carbón— para la determinación de la edad de los fósiles humanoides). La idea de que nuestras creencias son todas potencialmente interconectables y que la revisión puede aparecer en cualquier punto de la malla teórica, se ha considerado a menudo como «holismo», y Quine puede ser llamado el más sobresaliente abogado contemporáneo de las teorías holísticas tanto de la fijación de la creencia como del significado.

En el extremo opuesto al holismo de Quine están las visiones asociadas con el Positivismo Lógico, el cual sostuvo que cada frase significativa empíricamente tenía su propio área de experiencias verificantes y no verificantes (o, en el famoso eslogan, su propio «método de verificación») y que éste era fijado por los «significados empíricos» de los términos de la frase.

En la posición que acabamos de describir como el «extremo opuesto» del holismo de Quine, se aunan dos pretensiones: 1) que se podría, en principio, comprobar una teoría científica *frase por frase*; y 2) que esto es así porque cada frase empírica contiene términos con una cierta clase de *significados*. El reduccionismo positivista es una posición sobre la fijación de la creencia y simultáneamente una posición sobre el significado.

El célebre libro de Rudolf Carnap, *La construcción lógica del mundo* (*Der Logische Aufbau der Welt*), sostenía que todo término científico, debería, en principio, ser definible desde una «base autopsicológica»: desde términos que se referirían a las propias sensaciones de uno. (Así, cada científico interpretaría cada afirmación científica como una afirmación sobre sus experiencias privadas). Que Quine elaborara teorías holísticas de la fijación de la creencia y del significado constituye, en gran medida, una respuesta a esta posición de Carnap.

Hoy en día hay pocos filósofos, si los hay, que no estén de acuerdo en que la fijación de la creencia en *las ciencias maduras* es holística. Diferentes clases de consideraciones han jugado su papel para poder llegar a esta unanimidad tan poco usual. La primera de todas ellas es que se reconoce generalmente que la comprobación científica depende muy críticamente del uso de *teorías e hipótesis auxiliares*, incluidas las idealizaciones de todos los tipos. Por ejemplo, la teoría de la gravitación de Newton, por sí misma, es compatible con todas las órbitas, sean cuáles sean. Podemos incluso reconciliar la teoría de Newton con las órbitas cuadradas diciendo: «fuerzas no gravitacionales deben estar actuando en el sistema». Solamente en presencia de hipótesis auxiliares (las cuales están a menudo implícitas en los cálculos de los físicos más bien que formando parte explícitamente de la teoría) tales como «solamente las fuerzas gravitacio-

nales están actuando en el sistema», «las únicas fuentes gravitacionales significativas son el sol, la tierra...», etc., podemos derivar órbitas y otras predicciones comprobables de la teoría de Newton. En este sentido, no sólo no podemos hablar del «significado empírico» de la Ley de la Gravitación Universal por sí misma; sino que ni siquiera podemos hablar del «significado empírico» de la *mecánica newtoniana* como un todo, a menos que recurramos al expediente artificial de tomar las hipótesis auxiliares aceptadas corrientemente como parte de la «mecánica newtoniana». (Lo que la «mecánica newtoniana» es, cambia con cada problema que el físico aborda si aceptamos este recurso; y sigue siendo verdad que el «significado empírico» pertenece al sistema total bajo comprobación en cada caso y no puede adscribirse a ninguna frase aislada). En segundo lugar, incluso cuando un término científico posee una, así llamada, «definición», esa definición no tiene un privilegio epistémico especial. Por ejemplo, incluso si la frase «el momento es la masa por la velocidad» fuera una definición de hace dos o más siglos, mientras la frase «el momento se conserva» originariamente expresara una «consecuencia empírica», los científicos en el presente siglo no considerarían que el momento es la masa por la velocidad (considerada como una «verdad analítica») como algo más «seguro» que la Relatividad Especial. De hecho, Einstein mostró (suponiendo la Relatividad Especial) que si *hay* una cantidad vectorial conservada (en la dirección del movimiento de la partícula) que es aproximadamente igual a la masa por la velocidad, entonces, en el límite de velocidades pequeñas, *no puede* ser exactamente igual a la masa por la velocidad. La conclusión a la que llegó fue que, *estrictamente hablando se convierte en no verdadero* el que el momento sea el producto de la masa por la velocidad. La frase que era *originalmente* una definición se ha *venido abajo*.

El principio que jugó un papel crucial en el derrumbamiento de $p = m \cdot v$ (el momento es el producto de la masa por la velocidad), fue el Principio de la Relatividad Especial. $p = m \cdot v$ no fue derribado, como podría decirse, *directamente* —por un choque con una «consecuencia observacional»— sino por una colisión con un principio bien fundamentado que los científicos consideraron como mejor. Fue el atractivo de una construcción de la teoría física, de una manera de estructurar nuestra «red cognitiva» como opuesta a otra, lo que decidió esta cuestión.

Se observará que este argumento descansa fuertemente sobre el fenómeno de las revoluciones científicas. La creciente aceptación por los filósofos de la idea de que la «revisión puede aparecer en cualquier parte», y la aceptación de una idea que es en algunos sentidos incluso más importante, la idea de que algunas explicaciones pueden derrumbarse (como la caída del $p = m \cdot v$ de Einstein) *solamente si* se puede construir la totalidad de un esquema alternativo de teoría, y no sólo realizando experimentos, es ciertamente el resultado de las grandes revoluciones

científicas del siglo presente, y llevado a sus últimas consecuencias, explica la profunda aceptación de los enfoques «quinianos» de la fijación de la creencia en la ciencia teórica.

Donde todavía existe resistencia a aceptar la idea de que la fijación de la creencia es algo holístico, es en el área del «lenguaje observacional», como Carnap lo llamó, y en lo que algunos filósofos llaman «lenguaje ordinario». El mismo Carnap, mientras está de acuerdo con Quine en que la fijación de la creencia en la ciencia teórica es holística, parece vacilar entre conceder, por una parte, que el lenguaje observacional y el lenguaje teórico son interdependientes y querer mantener, por otra parte, que el lenguaje observacional es: 1) «completamente interpretado» (mientras que el lenguaje teórico es solamente «parcialmente interpretado»), y 2) que el lenguaje observacional recibe su interpretación de manera *previa* al lenguaje teórico. Michael Dummett, de forma similar, parece conceder que la ciencia teórica es «holística» pero insiste en que las frases del lenguaje cotidiano tienen condiciones de verificación canónicas que son fijadas (de manera recursiva) por sus propios significados. El holismo de la fijación de la creencia es algo que amenaza con volcar más de un carro filosófico que pueda de algún modo restringirse y hacerse inofensivo.

LENGUAJE ORDINARIO Y LENGUAJE CIENTÍFICO

La idea de que al lenguaje ordinario (o «lenguaje observacional») no le afecta la inestabilidad de la teoría científica, puede mostrarse atractiva para filósofos de más de una creencia, pero es totalmente falsa. Los anti-holistas parecen pensar que los «criterios operacionales» usados por el hombre lego agotan el significado de una palabra como *agua*. (En el caso del agua, éstos incluirían el ser transparente, incolora, inodora, calmar la sed, no ser tóxica, etc.). Pero no es difícil ver que esto es equivocado. Supongamos que hay (como podría haber, para todo *yo* que sepa química) un líquido, llamémosle «oxyhydroblahblah», tal que tuviera todas las propiedades «obvias» del agua excepto mantener la vida: si usted bebiera oxyhydroblahblah, le podría parecer agua (bien, quizás habría alguna pequeña peculiaridad en lo referente al sabor o al tacto), pero su sed no sería saciada. Una mezcla de H₂O y oxyhydroblahblah pasará, sin embargo, como agua (nunca sería peor que el agua de Philadelphia). Supongamos que usted da a un hombre lego en la materia, un vaso de cincuenta por ciento de H₂O y cincuenta por ciento de oxyhydroblahblah. Si le dice a este hombre que es un vaso de 99% de H₂O y uno por ciento de algo más completamente inofensivo, él quizás todavía desearía llamarlo «agua». Pero si le dice la verdad, no dirá que es «agua» (aunque

dirá que es *cincuenta por ciento* «agua»). El hecho de que la mitad de lo que se le ofrece sea una sustancia que no es constituyente del agua «normal», ciertamente lo excluye de ser clasificado como agua. En resumen, incluso el hombre lego incluye en su concepto de agua alguna noción sobre su «composición» normal o standard; y esto es suficiente para hacer que su uso del término *agua* dependa de las clases de hechos cuyo descubrimiento es asunto de la ciencia.

Alguna vez pensé que las extensiones lega y científica del término «agua» se relacionaban de una manera más simple de como ahora creo que lo hacen. Pensé que el «agua» del hombre lego denotaba H_2O químicamente pura (con alguna impureza más o menos). Pero esto es *demasiado* simple. Un vaso de café es más de un 99% de H_2O , pero el hombre lego no lo clasificará como un vaso de agua. Dirá que el «café consiste principalmente en agua» no que *es* un vaso de agua. Por otra parte, probablemente dirá que un vaso de H_2O con un 2% de lodo flotando en ella, es un «vaso de agua turbia». Nuestros intereses determinan qué clase de «impurezas» excluyen que una sustancia formada principalmente por H_2O sea «agua» *simpliciter*, y esto lo hacen de formas muy complicadas. Pero esto no ocurre en la sustancia que acabamos de fabricar con la ayuda del hipotético oxyhydroblahblah; la extensión científica y la extensión lega son diferentes pero interdependientes. Incluso el término «agua» del hombre lego no es un término puro y simple definido operacionalmente. Y esto significa que el holismo de la fijación de la creencia se realizará incluso en el uso del lego.

El concepto de agua del lego incluye el que tenga una clase normal de composición, como se ha dicho, pero el lego mismo no tiene los recursos o los criterios para determinar cuál sea realmente la composición normal en un nivel último, físico o químico. Debe ceder ante los expertos, pasados, presentes o futuros o, al menos, asumir que hay aspectos problemáticos sobre cuál sea la composición. Desde luego, la palabra «normal» por sí misma, presupone la existencia de un conjunto de paradigmas. Si todo el agua del Lago Michigan pasara a tener una composición «anormal» (relativa al paradigma de los organismos acuáticos), quizás seguiríamos considerando el líquido del Lago Michigan como «agua» en el sentido «ordinario» del término, así como seguiríamos considerando que las ballenas son «peces» en el sentido «ordinario» del término. Pero si una pequeña cantidad de agua putativa de procedencia incierta le diera a ese agua una composición «falsa» (como en el ejemplo del vaso de 50% de H_2O y 50% de oxyhydroblahblah), decimos «eso es una mezcla de agua y algo más»; el descubrimiento —que puede envolver la mayoría de las teorías de la investigación científica— de que la composición es errónea, muestra que el objeto no es (simplemente) agua. El «holismo moderado», como Quine lo denomina, es la línea correcta a seguir aquí. El fracaso ocasional del hombre lego al ceder ante los criterios científicos (como cuando sigue

llamando «peces» a las ballenas —si, como postulan los filósofos, *realmente* sigue llamando «peces» a las ballenas) muestra que el lenguaje ordinario y la fijación de la creencia no son *tan* holistas como el lenguaje científico y la fijación de la creencia, pero todavía permanece siendo cierto que las revoluciones científicas pueden derrocar las clasificaciones del lenguaje ordinario —derrocarlas para satisfacción del hablante *ordinario*, no sólo para el científico— en casos particulares, incluso si esto es una posibilidad que el hablante ordinario generalmente ignora justificadamente. La fijación de la creencia es (en principio) holística en cualquier parte.

EL TEMA DEL CAMBIO DE SIGNIFICADO

Aunque Carnap llegó a conceder que la fijación de la creencia en la ciencia teórica es holística, no abandonó la idea de que los *términos* en la ciencia teórica tienen sus propios «significados» individuales, como Quine insistía en que debería hacer. En lugar de ésto, eligió identificar el significado de un término en una teoría científica, con una función lógica (complicada) de la teoría misma. La «teoría», en esta concepción, no era el sistema meramente axiomático (visto como un cálculo no interpretado); más bien era el sistema axiomático junto con la interpretación dada de las palabras lógicas (*y, o, si, no, todos, algunos*, etc.) y los «términos observacionales». La teoría, tal como Carnap la concibe, es un «cálculo interpretado parcialmente» (los términos de observación serían —supuestamente— «interpretados completamente»; los términos teóricos serían «interpretados» solamente por estar conectados —conectados por la teoría misma— con el vocabulario observacional). Varios recursos de la lógica simbólica (la «frase de Ramsey» y el «operador épsilon de Hilbert») fueron empleados por Carnap en la formalización de esta idea.

El problema que plantea la propuesta de Carnap puede exponerse fácilmente: desde el punto de vista de Carnap, todo cambio en una teoría científica implica un cambio en el significado de cada uno de los términos teóricos que la teoría contiene. Además, la teoría no es sólo la teoría local o inmediata en la que el término se introduce; a causa del holismo de la fijación de la creencia en la ciencia teórica, es necesario que la «teoría» cuya «frase de Ramsey» se toma para fijar los significados de los términos teóricos, incluya todas las teorías auxiliares que se usen en la práctica en conjunción con la teoría local cuando se deriven predicciones comprobables. Así, en el caso del *electrón*, por ejemplo, la Relatividad Especial, la Relatividad General, la cromodinámica cuántica y los procedimientos usuales para la «gran unificación» son todas usadas en conjunción con la

teoría del «núcleo» del electrón y todas ellas modifican las predicciones que una sólo deriva de la teoría del núcleo; así todas ellas deben tomarse como parte de la teoría total que «implícitamente define» la palabra *electrón*, y cada uno de los grados en el desarrollo de este cuerpo de la teoría, incluyendo aquellos que normalmente tienen lugar, *cambia el significado* de *electrón*, según la teoría de Carnap.

Consideremos, por ejemplo, la cuestión «¿Por qué es H₂O un líquido a 20° C?». Antes de que esta pregunta pudiera ser contestada, la mecánica cuántica tuvo que desarrollarse. Pero incluso la mecánica cuántica no pudo aportar una respuesta a esta pregunta hasta el descubrimiento del Principio de Exclusión de Pauli. Esto significa que, desde el punto de vista de Carnap, «H₂O» cambio de significado al menos dos veces antes de que la pregunta original fuera contestada. O, por hablar con más precisión, la pregunta *original*, *nunca* fue contestada —a lo que se contestó fue a una pregunta expresada por palabras que eran homofónicas con las palabras de la pregunta original pero no sinónimas de ellas.

Esto significa que una distinción que es crucial en lógica y metodología, la distinción entre contestar realmente una pregunta y cometer una «falacia de equívoco», ha sido alterada hasta tornarse irreconocible. Porque no consideremos *una falacia* contestar así una pregunta como ésta —gracias al *desarrollo de teorías mejores*—, ¿de qué otra manera podríamos contestarla? Carnap podría replicar que hay dos clases de «cambio de significado», una buena y otra mala; y el equívoco es una falacia solamente cuando el cambio de significado es de la clase mala; esto significa reintroducir la distinción entre cambio de significado en sentido tradicional, cambio de significado en sentido propio, y cambio de significado en «sentido pickwickiano».

No solamente ocurre que el propósito de tomar el significado de un término o afirmación teórica como siendo la teoría, o una función satisfactoria de la teoría, implica una divergencia radical con nuestro uso «tradicional» o preanalítico de la noción de significado; tiene también el efecto de quitarle a esa noción todo su interés epistemológico. Si los significados no son invariables bajo los procesos normales de fijación de la creencia, la preocupación por el significado pierde su *raison d'être*.

MENTALISMO SOFISTICADO

Hasta aquí hemos argumentado que el carácter holístico de la fijación de la creencia no puede ser reducido a la «ciencia». Hay, sin embargo, otra forma en la que filósofo podría tratar de bloquear el argumento de Quine para el holismo del significado y esta forma ha sido propuesta insistente-

mente en una serie de publicaciones de Jerry Fodor. Fodor está completamente persuadido de que la fijación de la creencia —la fijación de toda creencia— es holística o, como recientemente ha dicho, «quiniana». Lo que Fodor modifica es el paso *desde* el holismo con respecto a la fijación de la creencia *al* holismo con respecto al significado.

Además, Fodor tiene buenas razones para pensar que esta opción es una salida posible. Tanto Quine como Carnap suponen que los «significados» se identifican en algún modo con sus condiciones de asertabilidad. Si la misma frase aparece en dos teorías diferentes, esto confirmará que la frase estará influida por la diferencia en las teorías circundantes. Una frase en una teoría es confirmada indirectamente por la confirmación de la teoría como un todo (al menos en muchos casos); si un fenómeno contradice la teoría A mientras es explicado por la teoría B, entonces una frase S puede decrecer en confirmación sobre la observación de ese fenómeno tanto tiempo cuanto la teoría A sea todavía la única teoría que contiene S, pero el fenómeno puede ser considerado más tarde (después de que ha aparecido la teoría B) como base de la teoría B, y, con eso, incrementar indirectamente el fundamento de la misma frase que se desechó previamente. Es por razones como ésta que Carnap considera el cambio desde una teoría A a una teoría B como un cambio en el «significado» de la frase S y que Quine desecha la noción de «significado» en conjunto (en conexión con frases simples). Pero la identificación de los «significados» con las condiciones de asertabilidad (o con las reglas o teorías que determinan las condiciones de asertabilidad) solamente está arraigada en aquellos filósofos que desean mantener la idea de que el *significado es el método de verificación*. Quine desea mantener esta idea y está dispuesto a pagar el precio: en la medida en que haya procesos que nos guíen en determinar cuándo aceptar y cuándo rechazar frases, estos procesos se asocian con el lenguaje entero, y así, desde su punto de vista, la «unidad de significado empírico» no puede ser más pequeña que la *totalidad del lenguaje*. Antes de Quine nos habíamos visto forzados a considerar la frase y no la palabra como unidad primaria. Después de Quine, hemos llegado a considerar el lenguaje entero como la «unidad de significado empírico». Al menos así es cómo le gustaría a Quine aparecer en las futuras historias de la filosofía.

Muchos filósofos, sin embargo, incluido Fodor, no aceptan ni siquiera la forma «holística» de la idea de que el significado es el método de verificación (yo mismo no puedo aceptarla). Si el significado —incluso en el caso de que el lenguaje entero o la totalidad del cuerpo de creencias de un hablante— no es una función del método de verificación o método de confirmación, entonces podría bien parecer que hay una manera fácil de preservar la idea de que las palabras y frases simples tienen significados determinados, significados que son invariables bajo los procesos normales de fijación de la creencia y cambio de teoría, sin encaminarse hacia una

discusión del holismo de los procesos de la fijación de creencia y del cambio de teoría. La idea general que define esta opción, la opción que llamaré «mentalismo sofisticado» es fácil de descubrir: 1) Postular «significados» como entidades psicológicas (pero no postular que necesitan ser, la mayoría de las veces, conscientes —ésta es la gran diferencia entre el mentalismo contemporáneo y el mentalismo ingenuo de los viejos empiristas. 2) Postular que estas entidades están, de alguna manera, asociadas con palabras individuales, morfemas y frases. 3) Postular que el paso desde el significado de una frase a sus condiciones de asertabilidad implica el uso de una inteligencia general y un fondo de conocimientos —y aquí es donde entra el «holismo».

La misión de este capítulo, y de este libro como un todo, es mostrar que esta opción no es, en un sentido, satisfactoria. Desde luego podemos decir que «conocemos el significado» de una pregunta como «¿Por qué es H₂O un líquido a 20° C?». Y podemos decir que esta pregunta significa precisamente lo que parece significar, que inquiriere por qué un cierto compuesto (cuya «fórmula», a un nivel de química de bachillerato, es H₂O) es un líquido a una temperatura ambiente.

No llamaríamos al descubrimiento de la mecánica cuántica o al descubrimiento del Principio de Exclusión de Pauli, un «cambio en el significado» de una pregunta como ésta. Pero algo extraño ocurre cuando empezamos a decir cosas como «la pregunta tiene un significado», «el significado no ha cambiado como resultado del descubrimiento de la mecánica cuántica», etc., en un tono de voz «teorizante». Tomando estas observaciones «cotidianas» como *explicativas* añadimos un peso a las nociones de «el significado» y «cambio de significado» que nada en su empleo ordinario les obliga a llevar. Y cuando incrementamos el peso añadiendo observaciones tales como «los significados son entidades teóricas», «los significados son reales psicológicamente», etc., entonces las nociones se colapsan totalmente. Porque *no hay* estados, procesos y objetos aislables —ésto es, independiente y científicamente aislables— que responden a los «significados» postulados por el mentalismo sofisticado. Al menos esto es lo que argumentaré.

TRES REQUISITOS EN CUALQUIER TEORÍA DE «CONTENIDOS»

Es parte de nuestra noción preanalítica de «significado» que frases diferentes y palabras diferentes puedan tener el mismo significado. Si hay *entidades* reales psicológicamente, llamémoslas «contenidos», las cuales se asocian con palabras y frases individuales de la forma postulada por la teoría del «mentalismo sofisticado» y fundamentan y explican nuestras

intuiciones sobre la «igualdad de significado» y la diferencia de significado, entonces sinónimos tales como la palabra inglesa «bachelor» (soltero) y la palabra francesa «celibataire» (soltero) deben tener las mismas entidades o muy parecidas de la clase en cuestión asociadas con ellos; los antónimos («casado»/«soltero») deben tener diferentes —pero relacionadas estructuralmente— entidades (contenidos) asociadas con ellos, y las palabras cuyos significados no son ni sinónimos ni están relacionadas en ninguna forma deben tener entidades (contenidos) diferentes y no relacionadas estructuralmente, asociadas con ellas. Que esto fuera así, sería nuestro primer requisito o coacción a cualquier teoría de «contenidos»:

I. *Diferentes contenidos deben, en general, estar asociados con frases que preanalíticamente suponemos difieren en significado, y los mismos o muy parecidos contenidos deben estar asociados con frases que preanalíticamente suponemos son las mismas en significado*

En resumen, los «contenidos», en el caso de que existan, deben tener los poderes correctos de *desambigüedad*.

En su célebre libro *Palabra y Objeto*, Quine intenta definir la noción de significado, al menos para las frases observacionales, en términos de la noción conductista de «movimiento de asentimiento/disentimiento» y de una noción neurológica de «estimulación de terminales nerviosas». El conjunto de todos los estímulos de las terminales nerviosas que incitarían al asentimiento a una frase S por parte de un hablante dado, es una conducta análoga a la noción positivista de «las sensaciones que confirmarían la frase»; el conjunto de todos los estímulos de las terminales nerviosas que incitarían al disentimiento de S es una conducta análoga a las «sensaciones que disconformarían S»; el par ordenado de estos dos conjuntos es llamado por Quine el *significado estimulativo* de S. Quine está claramente inclinado a identificar el significado estimulativo con el significado *simpliciter* y, al menos en estos días, a menudo se refiere al significado estimulativo de una frase observacional como su significado. Cuando es cuidadoso, sin embargo, distingue entre significado estimulativo y lo que él denomina «significado analítico». Es el significado analítico y no el significado estimulativo el que obedece a los requisitos que estoy tratando de enumerar; el significado estimulativo de una frase no puede determinar de hecho lo que una frase significa (en el sentido de determinar qué frase es sinónima con o puede ser parafraseada por) e incluso, no determina la referencia de los componentes de la frase. Además es Quine mismo quien ha argumentado esto. Su argumento se convierte así en el mismo requisito que acabo de enumerar.

Quine imagina un «lenguaje de selva» en el que la gente asiente a *gavagai* cuando ven un conejo y disienten de *gavagai* cuando no hay ningún conejo presente. De una manera más precisa, imagina que *gavagai* tiene exactamente el mismo significado estimulativo que nuestra frase castellana «¡un conejo!». (Es importante para la argumentación que *un*

conejo se piense como una frase completa —aunque no lo sea gramaticalmente— con la fuerza de «¡Mira, un conejo!» y no como una frase nominal). Aclara que no se sigue de esta suposición que *gavagai* deba significar «conejo». Ni siquiera si suponemos que *gavagai*, al igual que «un conejo», pueda utilizarse bien como una frase nominal o como una frase (elíptica). Por ejemplo —suponiendo que hubiera un «hecho» como qué significa *gavagai*, *gavagai* podría significar «la parte no fragmentada de un conejo» o «la conejeidad ejemplificada de nuevo» o «el instantáneo corte de tiempo de un conejo», en total conformidad con los hechos especificados. Si hablar de «conejeidad» llegara a ser habitual en castellano, la frase «¡Un conejo!» y «la conejeidad ejemplificada de nuevo» serían frases observacionales castellanas con exactamente el mismo significado estimulativo. Pero ningún conejo es idéntico con las partes no fragmentadas de un conejo o con el universal Conejeidad o con cualquier corte instantáneo del conejo (un «corte de conejo» es cualquier espacio tridimensional o sección cruzada de las cuatro dimensiones espacio-tiempo de un conejo).

Conejo, *conejeidad*, *parte no fragmentada de conejo* y *corte de conejo* no son idénticas en *denotación*. El número de «conejeidades» es exactamente uno; el número de cortes de conejo es infinito; el número de partes no fragmentadas de un conejo es probablemente indeterminado, pero ciertamente mucho más grande que el número de conejos. Como «frases ocasionales», «¡Un conejo!», «Conejeidad», «Parte no fragmentada de un conejo» y «Corte de conejo» podrían, repito, tener exactamente el mismo significado estimulativo. Pero como frases que están compuestas por partes que difieren en denotación no son, preanalíticamente, idénticas en significado; el requisito (I) requiere que a estas cuatro frases se les asignen diferentes «significados». Aquí significados (o «contenidos») no pueden ser idénticos con *significados estimulativos*.

He aquí un ejemplo diferente con el mismo efecto. He descubierto que mucha gente (erróneamente) cree que los tigres albinos no tienen rayas (realmente las rayas son tenues pero todavía bastante visibles). El hecho de que alguien pueda creer «los tigres albinos no tienen rayas» y «algunos tigres no tienen rayas» muestra que «tigre» y «tigre rayado» difieren en significado. Supongamos ahora que si hubiera un tigre sin rayas la gente dudaría en asentir a «¡Tigre!» —quizás porque no estarían seguros de que fuera un miembro de alguna otra especie de «grandes gatos». Entonces el significado estimulativo de «tigre» y «tigre rayado» podría ser exactamente el mismo incluso aunque todo hablante supiera que un tigre no es necesariamente un tigre rayado.

Mi próximo requisito se ha ido dibujando a lo largo de la discusión; es simplemente que:

II. *Los contenidos deben permanecer invariables bajo procesos normales de fijación de creencias.*

Para enlazar con el ejemplo anterior, imaginemos que todos los miem-

bros de una comunidad de hablantes aprenden muchísimas cosas sobre los tigres. Eventualmente, podemos suponer que todos llegan a ser diestros en el reconocimiento de tigres sin necesitar fijarse en la presencia o ausencia de rayas; si hubiera un tigre sin rayas (o si simuláramos uno) asentarían a «¡tigre!» en su presencia sin dudar excesivamente. En este caso el «significado estimulativo» de «¡tigre!» habría cambiado, pero esto no es lo que llamaríamos un cambio en el significado de la palabra *tigre*. Tomando «contenidos» como significados estimulativos violaríamos el requisito II (así como el requisito I); por este cambio en el significado estimulativo se nos habría llevado a lo que es ciertamente un «proceso normal de fijación de la creencia» —aprender más hechos sobre algo.

En un capítulo de su libro, Michael Dummett propone tomar el significado (esto es, el «contenido») de una frase como su método canónico de verificación. (Presupone que, en el lenguaje ordinario como opuesto al lenguaje científico, las frases tienen métodos canónicos estables de verificación). Sin embargo, esta propuesta se descarta también por el requisito II. No obstante, el método canónico de verificación de una frase puede realizarse por descubrimientos de varios tipos (nuevas pruebas para detectar oro, o ácido o cualquier otra cosa) o incluso por el cambio tecnológico (la invención de la escritura o el teléfono). Uno no puede desechar los métodos de verificación que implican los productos de la ciencia o la tecnología (consultar registros escritos, en el caso de afirmaciones sobre el pasado; o usar un cierto conocimiento científico para determinar cuánto lodo contiene un terreno) sobre la base de que estos métodos de verificación no tienen nada que ver con el *significado* de las frases en cuestión y son, *por tanto* «no canónicos», sin enfrentarse con la objeción de que las nociones de «significado» y de ser «canónico» se explican ahora una en función de otra de un *modo* tan *evidentemente circular* que no aporta ningún progreso en la comprensión de cada una de ellas. Además, hay muchas frases sobre el pasado que no pueden verificarse *excepto* consultando registros escritos; muchas frases sobre terrenos que tienen un alto contenido en lodo que no pueden verificarse *sin* ensayos científicos, etc. Si se considera que la totalidad de estas afirmaciones implica «usos extensos» de las palabras que contienen o «usos no ordinarios», o cualquier otra cosa, entonces nuestra teoría acabará postulando tantas «diferencias en el significado» que se perderá toda continuidad con nuestras intuiciones preanalíticas en torno a la semejanza y diferencia de significado. Se violarán tanto el requisito I como el II.

Finalmente, si estamos investigando la posibilidad de una teoría *mentalista* —una teoría en la cual el «contenido» de una palabra o frase es una entidad psicológica que fundamenta la práctica y las intuiciones de cada hablante que es completamente competente en el uso de una palabra o frase, requeriremos que:

III. *Sus contenidos deben estar «asociados» con las palabras y frases relevantes para cada hablante que se considere completamente competente en el uso de palabras y frases*

Este requisito excluye teorías del significado en las cuales los contenidos de los términos teóricos que son usados por todos los miembros de la comunidad de hablantes (*carga, electrón, ácido*) se identifican con teorías (o con funciones lógicas de las teorías como la frase de Ramsey) que son conocidas solamente por expertos. III es, así, un requisito de publicidad (un requisito de que los «contenidos», cualesquiera que sean, se compartan, implícitamente, de modo tan amplio como la competencia en el uso de palabras y expresiones que manifiestan esos contenidos) así como un requisito en el que los «contenidos» son entidades reales psicológicamente. Si el contenido de «carga» es una entidad real psicológicamente, presente en la cabeza de *todo* hablante que sea (para modelos normales) competente en el uso de la palabra, entonces ese «contenido» no puede ser la sofisticada teoría científica de la electricidad.

Ahora puedo manifestar más precisamente la propuesta que hice al final de la última sección, la propuesta que es el núcleo de este capítulo (y, en un sentido, del libro como un todo) y es establecer que: *No hay entidades que sean candidatos plausibles para desempeñar el papel de «contenidos» asociados con nuestras palabras y otras expresiones lingüísticas.* Todos los candidatos sugeridos hasta ahora por los escritores mentalistas como Fodor, violan al menos uno de los requisitos; y la sugerencia de que sólo tomemos «contenidos» como entidades *primitivas* en psicología, y de que *abandonemos* toda esperanza de identificarlos con algo neurológico o incluso con la descripción computacional del cerebro, coincide exactamente con mi argumentación. Porque la cuestión no es si es útil hablar de palabras «que tienen significado» y «no tienen significado», «que tienen el mismo significado» y «que no tienen el mismo significado», etc., en contextos en los cuales está claro por qué surge la pregunta y lo que vamos a hacer con la respuesta. La cuestión es si la idea de que los significados son sucesos, estados o procesos (o cualquier otra cosa) *aislables* con alguna clase de función explicativa en una genuina teoría científica, tiene el más mínimo fundamento.

UN EJEMPLO

Ya que la discusión ha discurrido a un nivel tan abstracto, quiero en este punto, proponer la discusión del ejemplo «ruritanio» de la «Psicología Computacional y de la Teoría de la Interpretación». En el ejemplo, una de las diferencias entre el dialecto del ruritanio que se habla en el norte y

el dialecto que se habla en el sur, es que en el ruritanio del norte *grug* significa «plata» mientras que en el ruritanio del sur, esta palabra significa aluminio. Imaginemos que la plata es algo tan común en el norte de Ruritania, que allí ollas y cacerolas están hechas de plata. Uno podría imaginar que en la Edad Media «grug» significaba «olla de metal» en Ruritania y que éste es el hecho de que las ollas del norte de Ruritania estén hechas de plata y las ollas del sur de aluminio, lo cual da cuenta de los cambios de significado que han tenido lugar. En cualquier caso, los niños del norte crecieron sabiendo que ollas y cacerolas estaban hechas normalmente de «grug» y los niños del sur crecieron sabiendo que ollas y cacerolas estaban hechas normalmente de «grug».

En el ejemplo, suponemos que Oscar y Elmer están en la misma condición psicológica (en el sentido «solipsista» de la «condición psicológica» de Fodor), esto es, la misma con respecto a todos los parámetros internos relevantes del lenguaje en t_0 .

Así, aunque en las comunidades adultas a las cuales pertenecen, «grug» tiene dos significados diferentes, tiene (en el sentido de Fodor) el mismo contenido para *Oscar* y para *Elmer* en t_0 . En t_1 (cuando se han convertido en adultos), las palabras deben diferir en contenido en los dos dialectos, tanto como «plata» y «aluminio» para los hablantes españoles. Por tanto, la palabra debe cambiar de contenido entre t_0 y t_1 (para al menos uno de los niños y, presumiblemente, para ambos).

	Elmer	Oscar
t_1	<i>grug</i> Elmer = «plata»	<i>grug</i> Oscar = «aluminio»
t_0	<i>la misma condición psicológica</i>	

Sin embargo, todo lo que ocurrió entre t_0 y t_1 fue la fijación normal de la creencia (en el modelo comúnmente empleado en lógica inductiva, condicionalización de las probabilidades previas a la nueva información a través de un fluido continuo). En t_0 ambos niños saben que «grug» es un metal, que es de un color gris brillante, que se empaña, que mamá tiene ollas y cacerolas de «grug», etc. En t_1 Oscar sabe que «grug» se llama «aluminum» en inglés americano y «aluminium» en inglés británico; que durante algún tiempo fue muy caro (Napoleón tenía joyas de aluminio y eran más costosas que el platino), pero se convirtió pronto en algo barato; que las ollas y cacerolas están hechas normalmente de «grug» o de acero o de cobre, excepto en el norte de Ruritania donde la plata (llamada «zlata» en el sur de Ruritania) se usa para hacer ollas y cacerolas; que el «grug» se saca de la bauxita, etc.; y Elmer sabe que «grug» es llamado «plata» en castellano; que el «grug» es un metal caro; que no se saca de la bauxita,

etc. Desde un punto de vista «interno», Oscar no sentiría que la adquisición de toda esta información fuera algo aparte de la adquisición de la información ordinaria que lleva consigo una noción que él ya tenía, la adquisición de unos pocos hechos más acerca del «grug». Una teoría de acuerdo con la cual la palabra «grug» cambiara su contenido en el curso del proceso perfectamente ordinario de fijación de la creencia, evita enfrentarse con las propiedades fundamentales de la noción de significado (requisito II). Además, en cualquier situación que decidamos estipular que la diferencia en «contenido» ha tenido lugar, será bastante arbitraria y no relacionada con las prácticas actuales de la paráfrasis e interpretación (lo que violará el requisito I).

Además, si postulamos que la palabra «grug» se asigna a una fórmula «de carácter mental» (una «representación semántica») en t_0 (en ambos cerebros, ya que los niños están en la misma condición psicológica), entonces si el proceso normal de fijación de la creencia no altera la «representación semántica» de las palabras, la fórmula mental —digamos XYZ— acabará teniendo *dos significados diferentes* en t_1 . Trasladando el problema del significado al «carácter mental» no habremos resuelto nada, puesto que necesitaríamos una teoría del significado de «carácter mental» tanto como necesitamos una teoría del significado para cualquier lenguaje natural.

MI TEORÍA EN «EL SIGNIFICADO DEL "SIGNIFICADO"»

En «El Significado del "significado"» propuse una teoría *no* mentalista de acuerdo con la cual los «significados» tienen varios componentes, p.e.,

Agua

MARCAS SINTACTICAS	MARCAS SEMANTICAS	ESTEREOTIPO	EXTENSION
nombre común concreto	clase natural líquido	incolora transparente sacia la sed, etc.	H ₂ O (aprox.)

En esta teoría, el significado de «grug» en el dialecto de Elmer es *plata desde t_0* en adelante y el significado de «grug» en el dialecto de Oscar es «aluminio» *desde t_0* en adelante. Mi teoría adscribe las *extensiones determinadas socialmente* a las palabras de los niños; si éstas fueran diferentes, diríamos que las palabras tienen diferentes significados, desatendiendo lo

que ocurriera en las *cabezas* de los niños. No esperamos que un niño pequeño tenga el conocimiento o la práctica que esperamos tendría un hablante adulto. No esperamos que el individuo medio que usa la palabra *oro* tenga el mismo conocimiento o la misma práctica que un joyero tendría, o el conocimiento o la práctica que tendría un químico. No obstante, tanto el niño, como el hablante medio, como el químico, dependen unos de otros y además dependen de todas las cosas que les rodean. Si un niño está «envuelto» en esta malla, atribuiremos a su palabra el significado y la referencia que la palabra tienen en la comunidad lingüística, incluso si el niño no está en posición de fijar la referencia por sí mismo o no está en posición de contestar las preguntas sobre el significado. Desde este punto de vista, el significado de *grug* es diferente para Oscar y para Elmer *incluso en el momento* t_0 . De hecho, es correcto decir que significa «plata» en el dialecto de Elmer y «aluminio» en el dialecto de Oscar en t_0 .

La teoría que he desarrollado en «El significado del "significado"» tiene un aspecto mentalista sin embargo y este aspecto es el que yo quisiera rechazar ahora. En aquel ensayo sugerí que tomamos los componentes psicológicos del significado (el «contenido») que son el estereotipo más las marcas sintácticas y semánticas. Me parece claro que no exigimos y no deberíamos exigir la *igualdad* del estereotipo antes de decir que dos formas tienen el mismo significado, sino más bien exigir una similitud suficiente, donde lo que consideramos como «suficiente» es algo altamente dependiente del contexto. En resumen, abandonaríamos la idea de que hay algo que es el «componente psicológico» de un «significado». Más bien que decir, con Fodor, que el contenido de *grug* es el mismo para Oscar y para Elmer en t_0 , rechazaríamos la noción de «contenido».

MÁS SOBRE «ESTEREOTIPOS»

Aunque no creo que cualquier filósofo o psicólogo defendería la idea de que los estereotipos pueden jugar el papel que Fodor asigna a los «contenidos» (el mismo Fodor ha argumentado que no pueden), creo que un examen de algunas de las razones por las que no pueden hacerlo aportará más claridad a la situación.

Empecemos con un caso muy simple. Nadie negaría que entre los hablantes españoles las rayas son un aspecto estereotípico de los tigres, o que las manchas son un aspecto estereotípico de los leopardos, o que una melena y una cabeza «leonina» son aspectos estereotípicos de los leones (machos). Pero la controversia empieza con las preguntas «¿Hay una conexión *semántica* entre la noción de tigre y la noción de tener rayas?», «¿Hay una conexión semántica entre la noción de leopardo y la de tener manchas?», etc., ¿Por qué ocurre esto?

Una razón para pensar que hay un problema aquí es ciertamente una mala razón. Los filósofos tendieron durante mucho tiempo a escribir como si «el tener "B" es parte del significado de "A"» significara que «Todo lo que sea A tiene B» es analítico. Ahora bien, «Todos los tigres tienen rayas» ciertamente no es analítico. Así, se seguiría de esta doctrina, que la característica de tener rayas no puede ser parte del significado de tigre, y, paralelamente, que la característica de tener manchas, no puede ser parte del significado de *leopardo*, etc. De hecho, si Quine tiene razón y no se puede establecer la distinción sintético/analítico (o si no hay frases analíticas), se seguiría que la noción de significado simplemente tiene que ser eliminada. Lo que la mayoría de nosotros —lo que, por ejemplo, tanto Fodor como yo— concluiríamos, es que la noción de significado tiene que ser separada de la doctrina de la analiticidad. Esta salida, al menos por lo que sé, fue propuesta por primera vez por Paul Ziff en *Análisis semántico*. De acuerdo con Ziff, la característica de tener rayas se asocia con la palabra *tigre*, tanto si es o no verdad de hecho que todos los tigres, o la mayoría de los tigres, tienen rayas. Incluso si ocurre que las rayas son pintadas por los nativos, o son un producto de una fantasía o ilusión colectiva, es un hecho que afecta a la lengua castellana el que haya una asociación semántica entre la palabra *tigre* y la característica de tener rayas.

Parte de la resistencia a este simple (y creo que muy plausible) paso dado por Ziff, surge del hecho de que los filósofos habían identificado durante mucho tiempo los significados con lo que ellos llamaban «conceptos». Ahora bien, los «conceptos» son cosas sobre las que los filósofos tienen teorías —*cantidad* de teorías, algunas de las cuales se remontan a la Edad Media e incluso más atrás. Decir «B es parte del concepto A» es una manera habitual que tiene el filósofo de decir «*Todo lo que sea A tiene B*» es una verdad necesaria. Identifiquen «significados» con conceptos y «verdad necesaria» con «frase analítica» y ¡voilà! obtendrán el principio al que aludí más arriba, el principio del cual se sigue que tener rayas no es parte del significado de *tigre*.

Incluso si se bloquea la argumentación de Quine (rechazando la noción de significado sobre la base de que tal noción presupone la noción de analiticidad) de una forma diferente a como lo hace Ziff, insistiendo en que Quine está realmente equivocado y que *hay* algunas verdades analíticas, nadie diría que «Todos los tigres tienen rayas» o «Todos los leopardos tienen manchas» o «Todos los leones machos tienen melenas» son frases analíticas. Así pues, todavía se seguiría de esto, que ninguna de las características por las cuales distinguimos *estereotípicamente* los leones de los leopardos y de los tigres, son parte del significado de estas palabras. Pero parece entonces como si *tigre*, *león* y *leopardo* fueran a tener exactamente el mismo significado.

Una vez Dretske propuso renunciar a la idea de que esta clase natural

de palabras tuviera significado. En efecto, propuso tratar estas palabras como los teóricos tradicionales trataron a menudo los nombres propios, como palabras con una referencia pero sin ningún contenido conceptual. Pero gran parte de nuestro lenguaje consiste en esta clase natural de palabras y renunciar a la noción de significado en relación a ellas podría ser solamente el primer paso hacia la posición de Quine, la posición en la que la noción de significado debe ser abandonada totalmente. Además dificultades similares surgen con otra clase de palabras. ¿Es *analítico* que los lápices son artefactos? (Quizás son realmente organismos).

Cualquier noción de significado que tenga continuidad con nuestra noción preanalítica debe poder «evitar la ambigüedad» existente entre *león, tigre y leopardo*. Y la única manera plausible de hacer esto es la manera en que «preanalíticamente» lo hacemos, la forma en que «intuitivamente» lo hacemos: decir que *tigre* tiene un significado diferente de *leopardo* porque *inter alia*, la característica de tener rayas está asociada con *tigre* mientras que la característica de tener manchas (más bien que rayas) está asociada con *leopardo*. Hasta aquí hemos de seguir a Ziff. O así me lo parece.

El error que cometí en *El significado del «significado»* fue tratar sin demasiada exactitud la noción de un «estereotipo». Incluso si estamos de acuerdo con Ziff en que los tigres son estereotípicamente rayados y los leopardos estereotípicamente moteados, no se sigue de ello que haya alguna cosa como «el» estereotipo de un tigre, leopardo, o de la casa de un gato; no se sigue que haya algún objeto invariablemente fijado (el mismo para cada hablante que sea competente en el uso de la palabra, como exigía nuestro requisito III) presente en la mente/cerebro de los hablantes que conocen la palabra o cualquier palabra con «el mismo significado».

He aquí un ejemplo en relación con la palabra *gato*, o más bien con la palabra tailandesa que los diccionarios traducen como *gato*, la palabra (¡créanlo o no!) *meew* (pronunciada como un francés pronunciaría *meu*): He aprendido que todos los gatos que uno ve en Tailandia (primeramente llamada Siam) son gatos siameses. Una persona que haya crecido en Tailandia tiene un estereotipo de gato bastante diferente al que nosotros tenemos. De hecho su estereotipo de un «meew» es nuestro estereotipo de «gato siamés».

¿Cómo, entonces, traduciríamos *meew*? Si lo traducimos como «gato siamés» comunicamos el estereotipo correcto pero la extensión equivoca, porque el siamés también se refiere a los gatos europeos y americanos como «meew». Así, sería equivocado decir sin más que *meew* significa «gato siamés» (como lo sería decir sin más que *tigre* y *tigre rayado* son sinónimos sobre la base de que el estereotipo es el mismo). El estereotipo puede ser un «componente» del significado, para usar la terminología que empleé en «El significado del "significado"», pero existen también otros componentes. ¿Traduciríamos *meew* como «gato»? Bien, eso es lo que

realmente hacemos. Lo traducimos así de una manera en que la extensión es correcta aunque no conserve el estereotipo. Sin embargo, hay casos en que decimos que términos con la misma extensión *no* tienen el mismo significado. Si un lenguaje tuviera términos, digamos, *cardiolate* y *lungolate*, para «criaturas con corazón» y «criaturas con pulmón», no diríamos que éstos fueran sinónimos incluso aunque pudieran tener exactamente la misma extensión. En resumen, no parece que sea una regla general.

Un problema ulterior (uno que fue mencionado en «El significado del "significado"») es que los estereotipos no son imágenes (o «prototipos perceptuales») en jerga psicológica sino creencias manifestadas en palabras («un átomo de hidrógeno contiene un electrón y un protón», «el aluminio es un metal»). Para saber cuál es el estereotipo de alguien o de algo se requiere poder interpretar el lenguaje de esa persona: una teoría que toma la noción de «estereotipo» como primitiva no puede *explicar* la interpretación.

Supongamos que para evitar esta dificultad restringimos la noción de estereotipo a aquellas características que pueden ser *descritas* (y supongamos que los psicólogos han descubierto un criterio razonable para determinar si una característica de algo es «estereotípica» para un objeto dado, en este sentido). Incluso así, tomar cada diferencia en el «estereotipo» (en tal sentido) como una diferencia en el significado, diferiría de la práctica actual. No solamente ocurriría que *meow* no tiene el mismo significado que gato, sino peor, que si Tom está junto a la puerta no incluye los bigotes de un gato o el recipiente de leche como parte de *su* estereotipo y el de Jane sí, entonces el significado de la palabra es diferente para Tom y Jane. Desde luego muchos «filósofos folk» dicen justo ésto: «ninguna palabra tiene el mismo significado para dos personas diferentes»; pero así *no* es como realmente *usamos* la noción de «significado» cuando *no* estamos siendo filósofos folk. De hecho todos los mentalistas sofisticados que conozco rechazan la idea de que los significados *son* estereotipos (y creo que correctamente; ¿podríamos dar algún sentido a la noción de «estereotipo» en el caso de palabras tales como «mente», «esprit», «Geist»?); pero si alguien estuviera tan poco atento como para mantener este punto de vista, deberíamos aconsejarle que *abandonara* la palabra «significado» y hablara del papel de los estereotipos en la comunicación. (Desde luego, no son invariables en la fijación de la creencia y excepto en las revoluciones científicas, no determinan relaciones de paráfrasis).

Así, las dos direcciones extremas —es decir, negar que los estereotipos juegan *algún* papel en la fijación de los significados de las palabras y considerar *toda* diferencia en el estereotipo como una diferencia en los significados de las palabras—, conducen a teorías del significado que violan desastrosamente nuestros tres requisitos. El curso que seguimos ahora en nuestra interpretación es un curso intermedio entre estos dos extremos. En primer lugar, cuando las palabras cortas tienen exactamen-

te la misma extensión, tendemos a tratarlas como sinónimos independientemente de lo que haya «en la cabeza». Esta es la razón por la cual las teorías mentalistas (del tipo de las hechas en el Instituto Tecnológico de Massachusetts) tienen problemas desde el principio: tratan de *excluir* la extensión; pero la extensión es lo que más firmemente nos guía en la interpretación, especialmente cuando no hay un contexto específico para guiarnos. (Sí, sé que «criaturas con corazón» y «criaturas con pulmón» tienen un significado diferente pero la misma extensión; pero las *partes* de estas frases no tienen las mismas *extensiones*).

En segundo lugar, cuando consideramos otros factores además de la extensión, consideramos *de hecho* estereotipos (aquellas rayas *son* algo conectado con el significado de «tigre»), pero a lo que nos estamos refiriendo no es a la *identidad* del estereotipo (aunque podría definirse), sino a una *semejanza suficiente*. Y no hay una regla general para decidir cuándo dos estereotipos son suficientemente similares; depende del contexto particular, incluyendo las razones por las que alguien quiere saber lo que una palabra significa y lo que va a hacer con la respuesta.

Para resumir: No he argumentado que el significado es indeterminado (aunque Quine argüiría así). Lo que he argüido es que *cuando* el significado es determinado *no es una cosa* lo que hace ésto. El «significado normal» de *meew* en tailandés es *gato*; no porque la *misma* entidad esté en las cabezas de los hablantes tailandeses y españoles, ya que ésto sería olvidar el hecho de que hay diferencias psicológicas. Decir que los significados son los «mismos» es decir que es razonable, en un contexto dado, ignorar las diferencias en la creencia, actitud o estereotipo que tienen los hablantes o escritores en cuestión. La «igualdad de los significados no es la «igualdad» de los objetos (dos protones o incluso dos corbatas), sino sólo la racionalidad de ignorar las diferencias, incluidas las diferencias en los «procesos psicológicos».

MI DESCRIPCIÓN PRESENTE

Acabará mi conferencia describiendo la relación que todo esto guarda con la cuestión del «funcionalismo». La descripción funcionalista era una descripción a dos niveles:

F ₁ F ₂ F ₃ F ₄ F ₅ F ₆	Estados computacionales (software) = Estados psicológicos Estados físico-químicos
S ₁ S ₂ S ₃ S ₄ S ₅	

en la cual cada estado psicológico se identificaba con un estado computacional que podría verificarse por un gran número (potencialmente un número infinito) de estados físico-químicos. Más tarde se comprendió que no solamente un estado computacional dado podría verificarse de muchas (infinitas) maneras diferentes cuya relación (entre cada una de ellas) no podría verse a un nivel físico-químico, sino que también un estado físico podría ser la verificación de un número infinito de diferentes estados computacionales; es decir, que un estado físico-químico no tiene una descripción computacional «intrínseca». Así pues, la relación de los estados en los dos niveles diferentes es completamente plural.

Esto era una descripción *anti-reduccionista* llevada a tal punto que se negaba cualquier reducción directa de las propiedades psicológicas a las propiedades físico-químicas. Pero todavía se suponía que existía una reducción de las propiedades psicológicas a las computacionales. Hoy, en vista del tipo de consideraciones que acabamos de analizar, me movería en una descripción a tres niveles:

- _____ Estados psicológicos (creer,
desear, pensar que, etc.)
- _____ Estados computacionales
- _____ Estados físico-químicos

con muchas y diferentes relaciones entre cada uno de los niveles. En todo esto, lo que estamos viendo gradualmente es un desmoronamiento de la descripción reduccionista.