

**ESTRUCTURA Y CONTENIDO
DE LA INTERPRETACION.
UNA INTRODUCCION A LA TEORIA DE
LA RACIONALIDAD DE D. DAVIDSON [II]**

Ramón del CASTILLO
Dpto. de Filosofía, UNED

VI. El significado, las creencias y la contrastación de la teoría.

Ahora habría que detenerse en las constricciones empíricas de la teoría que hacen verosímil la solución holística al problema de su adecuación material. En el centro del método de interpretación tenemos, según hemos visto hasta ahora, una teoría formal: una teoría de la verdad que impone una estructura composicional a las oraciones y que hace uso de una noción primitiva de verdad que se aplica a través de la forma de la teoría. Ahora se trata de determinar la naturaleza de la evidencia relevante para su adecuación¹. Y debemos hacerlo sin presuponer aquello que pretende describir la teoría²; debemos disponer de una evidencia que no sea descrita en términos semánticos o lingüísticos³. O mejor dicho: necesitamos una evidencia para la verdad de las oraciones T que,

¹ «Radical Interpretation», ITI, p. 137.

² ITI, Introduction, p. xvi.

³ «Belief and the Basis of Meaning», ITI, p. 143.

Éndoxa: Series Filosóficas, nº 6, 1995, UNED, Madrid:

Ramón del Castillo: Estructura y contenido de la interpretación. Una introducción a la teoría de la racionalidad de D. Davidson. II.

pp. 133-166.

sea cual sea, no pueda describirse en términos que la relacionen de antemano con algún lenguaje particular⁴. No es que la evidencia tenga que reducirse o definirse explícitamente con términos no semánticos. Reducción y definición —dice Davidson— son expectativas demasiado grandes. La teoría y la evidencia tienen una relación mucho más laxa⁵. Trataremos de ver cómo la totalidad de las oraciones T se puede ajustar óptimamente con la evidencia disponible⁶.

Si, como hemos dicho, la teoría se contrasta a nivel macroscópico, entonces tiene sentido la afirmación de Davidson de que es más fácil conectar a la verdad con evidencia no-lingüística que a la referencia. (Hemos visto que no es necesaria la referencia como *el* lugar donde hay contacto directo entre la teoría lingüística y los sucesos, acciones u objetos describibles en términos no-lingüísticos⁷). La verdad —dice Davidson— no es sólo más fácil de conectar con evidencia no semántica, sino también con las actitudes más básicas y simples de los hablantes⁸ que pueda distinguir un intérprete radical. (Recordemos que un principio metodológico de Davidson es que una emisión tiene cierto significado sólo si el hablante tiene la intención de que se la interprete con ese significado. Un uso exótico, un desliz verbal, una variante idioléctica, una palabra nunca oída, si significan algo —dijimos—, significan lo que el hablante quiere que signifiquen). Ahora podríamos decir: una emisión tiene ciertas condiciones de verdad sólo si el hablante tiene la intención de que se la interprete con esas condiciones de verdad. Las condiciones de verdad no pueden ser independientes de las intenciones del hablante (no dependen ni de la mayoría, ni de los expertos-bien-informados)⁹.

⁴ «Reality Without Reference», ITI, p. 223.

⁵ *Ibidem*.

⁶ «Radical Interpretation», ITI, p. 139; «Belief and The Basis of Meaning», ITI, p. 149.

⁷ «Reality Without Reference», ITI, p. 223.

⁸ «Radical Interpretation», ITI, p. 134.

⁹ «The Structure and Content of Truth», p. 310.

Para comprender la verdad y el significado, debemos «pegarnos tanto como nos sea posible a la actitud que el hablante pone a disposición de su audiencia. Lo que importa para la comunicación con éxito es, por un lado, la intención del hablante de que se le interprete de cierta forma y, por otro, la interpretación real de las palabras del hablante en la línea de sus intenciones a través de su reconocimiento por parte del intérprete»¹⁰. En consecuencia: hay que mostrar que podemos aceptar sin circularidad actitudes hacia las oraciones que sirvan como evidencia primera para la teoría¹¹. El proceso de contrastación tendría, como dije ya, algún parecido con el que Quine describió en *Word and Object*: El intérprete que quiere interpretar no tiene otra opción que intentar ver en qué condiciones un hablante asiente a (o disiente de) una serie de oraciones que se le ofrecen¹². Al principio, el intérprete no podrá poner palabras en boca de su informador, simplemente porque aún no las tiene. Pero, en cuanto pueda, suministrará a su informador oraciones de la lengua de éste para que las apruebe. En ese punto, el intérprete corre el riesgo de deformar los datos por sugestión, pero no tiene ninguna otra opción (aparte del hecho de que tiene que conjeturar qué signos se deben tomar como signos de aprobación o negación en la lengua nativa). Davidson lo dice así: intentaremos ver bajo qué condiciones el hablante asiente o disiente a una serie de oraciones y consideraremos que las condiciones relevantes son las condiciones de verdad de sus oraciones¹³.

El problema es que la teoría empírica del lenguaje no puede ser algo aislado: necesita apoyarse en una teoría de las creencias y de

¹⁰ *Ibid.*, p. 311.

¹¹ No debemos perder de vista que estamos intentando dar una interpretación de un lenguaje que es que forma parte de la vida de los hablantes, de sus conductas. La tesis de Davidson es que, al igual que los rasgos semánticos de las palabras se abstraen de los rasgos de las oraciones, los rasgos semánticos de éstas han de abstraerse de su contribución a las formas en que los hablantes tratan de alcanzar objetivos o realizar deseos («Reality Without Reference», ITI, p. 220-222).

¹² «A Coherence Theory of Truth and Knowledge», PTI, p. 315. «The Inscrutability of Reference», ITI, p. 230.

¹³ «Semantics for Natural Languages», ITI, p. 62.

las intenciones del hablante¹⁴. La totalidad de la evidencia disponible puede adecuarse a dos teorías distintas siempre que efectuemos ajustes compensatorios en las creencias e intenciones del hablante¹⁵. Lo que hay que mostrar es lo que se ha sugerido en la última sección: que la indeterminación «no es tan grande como para que cualquier teoría que pase el test nos sirva para interpretar»¹⁶. La evidencia de que las oraciones T son verdaderas *no* puede consistir en descripciones detalladas de las creencias e intenciones del hablante, porque las atribuciones de actitudes exigirían una teoría que debería descansar en una evidencia muy parecida a la de la interpretación¹⁷. No hay evidencia lingüística que sea independiente de las atribuciones de creencias, deseos, reproches, apetencias (todas estas actitudes tienen —dice Davidson— un componente de creencia¹⁸). Incluso habría que ver las cosas así: los rasgos semánticos de las palabras se abstraen de las oraciones, y los de las oraciones de su contribución a los sistemas de creencias de los hablantes y a sus patrones para alcanzar fines o realizar deseos¹⁹. En suma, no es posible determinar la corrección de una teoría de la interpretación sin conocer bastantes cosas sobre las creencias y deseos de los hablantes; pero tampoco es posible derivar una teoría de la interpretación a partir de información detallada (de índole no-lingüística) sobre las intenciones, los deseos y las creencias de los hablantes, pues ¿podríamos verificar su existencia cuando aún somos incapaces de determinar lo que significan las palabras del hablante? «Hay un obstáculo de principio, y no meramente práctico, para verificar creencias e intenciones particulares, generales y abstractas mientras no seamos capaces de decir lo que significan las palabras del hablante. Todos

¹⁴ «The Structure and Content of Truth», p. 295.

¹⁵ «Semantics for Natural Languages», ITI, p. 62, «The Inscrutability of Reference», ITI, p. 240.

¹⁶ «Radical Interpretation», ITI, p. 139.

¹⁷ «Radical Interpretation», ITI, p. 134.

¹⁸ «Belief and the Basis of Meaning», ITI, pp. 142-143.

¹⁹ «Reality without Reference», ITI, p. 220.

entendemos bien lo absurdo que es tratar de aprender sin preguntar al hablante si alguien cree que existe el mayor número primo o si, haciendo ciertos ruidos, tiene la intención de hacer que alguien deje de fumar cuando esa persona se dé cuenta de que los ruidos se hicieron con esa intención. El absurdo no reside en el hecho de que sería muy difícil dar con esas cosas sin lenguaje, sino en el hecho de que no tenemos ni idea de cómo se verificaría la existencia de esas cosas cuando no es posible la comunicación»²⁰. Aun restringiendo la atención a la conducta que revela cuándo y bajo qué condiciones un hablante da crédito a una oración, no hay forma directa de diferenciar los papeles de la creencia y el significado en la explicación de esa actitud²¹. Un hablante que considera que una oración es verdadera en una ocasión lo hace en parte por lo que quiere, o querría decir, con una emisión de esa oración y, en parte, por lo que cree. Si sólo podemos proseguir en base a la emisión honesta, no podemos inferir la creencia sin conocimiento del significado y no podemos tener esperanza de inferir el significado sin la creencia²². La aportación de datos puede resolver algunos casos, pero siempre hay otros en los que todos los datos disponibles dejan abierta la elección entre atribuir un significado corriente y una creencia exótica, o un significado extraño y una creencia común²³. Necesitamos seguir un procedimiento que nos permita hacer efectiva una distinción entre

²⁰ «Belief and the Basis of Meaning», ITI, p. 144.

²¹ ITI, Introduction, p. xvi.

²² «Belief and The Basis of Meaning», ITI, p. 142; «Thought and Talk», ITI, p. 162. Davidson utiliza ejemplos sencillos como éstos: Si ves un yate navegando y tu interlocutor te dice: «Mira qué barca tan bonita», una posibilidad natural es que tu amigo haya confundido un yate con una barca y se haya formado una creencia falsa. Pero si tiene bien la vista y su punto de observación es apropiado, es más plausible que no use la palabra "barca" tal como tú lo haces y no haya cometido error alguno sobre la posición del mástil al paso de una embarcación. «On the Very Idea of a Conceptual Scheme», ITI, p. 196. Si alguien dice «aquello es una estrella fugaz», ¿debo considerar que realmente habla de una estrella, pero cree que algunas estrellas son muy pequeñas y frías? ¿O, más bien, debo pensar que no se está refiriendo a una estrella sino a un meteorito y cree que las estrellas siempre son grandes y calientes? («Psychology as Philosophy», EAE, p. 257).

²³ «Psychology as Philosophy», EAE, p. 257.

contribuciones de significado y de creencias en un grado suficiente para la comunicación: «como se depende de una evidencia que, sin ayuda de una teoría, no distingue entre la contribución de la creencia y la del significado para la conducta lingüística, se necesita un método para separarlas en un grado suficiente para la comunicación»²⁴, un procedimiento «que mantenga fijo un factor mientras se determina el otro»²⁵. Con ese método, la interpretación puede empezar a dar sus primeros pasos sin necesidad de forzar una decisión²⁶, sino derivando simultáneamente una teoría de la creencia y una teoría del significado²⁷. (Además, si la creencia y el significado están relacionados de esta forma, entonces la teoría adecuada no puede partir de la idea de que cada creencia tiene un objeto definido y cada oración un significado definido²⁸).

VII. Verdad, evidencia e intenciones.

La tesis de Davidson es que es posible saber que el hablante asiente a una oración sin saber qué significa la oración para él ni cuál es la creencia expresada por ella²⁹. El reconocimiento del asentimiento a la oración no tiene por qué presuponer ni el conocimiento de las creencias del hablante ni el de las relaciones gramaticales y lógicas que la oración guarda con otras oraciones que tenga por verdaderas. Por tanto, un principio básico de la teoría es que el hecho de que los hablantes consideran que una oración es verdadera (bajo circunstancias observables) es, *prima facie*, evidencia de que la oración es verdadera³⁰. Entonces, la

²⁴ ITI, Introduction, p. xvii.

²⁵ «Thought and Talk», ITI, p. 167.

²⁶ *Ibid.*, p. 168.

²⁷ «Belief and the Basis of Meaning», ITI, p. 144.

²⁸ *Ibid.*, p. 163.

²⁹ «Thought and Talk», ITI, p. 162.

³⁰ «Belief and the Basis of Meaning», ITI, p. 152.

evidencia empírica de la verdad de oraciones T, de teoremas de la teoría de la verdad como

«A cloud is passing before the sun» es verdadera si y sólo si una nube está pasando delante del sol

serán hechos como que

William considera verdadera a «a cloud is passing before the sun» el sábado por la mañana si y sólo si una nube pasa por delante del sol cerca de William el sábado por la mañana.

Desde luego, la actitud de considerar verdadera a la oración es una clase de creencia, pero es la única actitud aplicable a todas las oraciones, y, por tanto, *no* requiere que seamos capaces de hacer distinciones finas entre las creencias del hablante. Podemos identificar esa actitud *antes* de poder interpretar, porque podemos saber que una persona prefiere expresar una verdad sin tener la menor idea de *qué* verdad se trata³¹. Esto satisface el requisito de la disponibilidad independiente que requiere la base evidencial de la teoría³². Esta actitud que reconocemos directamente en el

³¹ «Radical Interpretation», ITI, p. 134-135; «Belief and the Basis of Meaning», ITI, p. 144.

³² «Radical Interpretation», ITI, p. 135; «On the Very Idea of a Conceptual Scheme», ITI, pp. 195-196. «Thought and Talk», ITI, p. 161. Además, Davidson sugiere que esta actitud no es la única que nos puede hacer suponer que una persona considera verdaderas a ciertas oraciones. Órdenes, historias, ironías, mentiras si se identifican como actitudes, podrían revelar lo mismo. Ni siquiera hay que presumir que la mayoría de las emisiones de un hablante corresponden a emisiones que considera verdaderas. Si fuera así, ciertamente aseguraríamos la disponibilidad independiente de la evidencia. Pero, afirma Davidson, «presunciones más débiles también funcionarían, pues incluso el mentiroso compulsivo o el eterno bromista pueden ser descubiertos». También actitudes como desear la verdad de la oración, querer probar su verdad o creer que uno va a probar su verdad, o imaginarse si es verdadera, o pretender hacerla verdadera podrían ser otras actitudes reveladoras, pero para Davidson toda la evidencia de esta clase podría verse en términos de la actitud de tomar por verdadera a una oración.

Una posible crítica a su planteamiento es que su método de interpretación hace depender demasiadas cosas de la actitud básica de considerar verdaderas a las oraciones. ¿Implica esto la negación de la *autonomía del significado*? ¿No se pasa por alto que tener un lenguaje es poder usar una oración *para cualquier fin*?

hablante es el vector de dos fuerzas, significado y creencias, y, por eso, tenemos que abstraer de ella, al mismo tiempo, una teoría del significado adecuada y un teoría aceptable de la creencia³³.

El problema que se presenta si tratamos de obtener evidencia de esta actitud del hablante es que siempre podría estar equivocado³⁴. Parece que esto nos cerraría el paso para obtener más evidencia que respaldara generalizaciones. El hecho de que una oración se considere verdadera no parece que pueda garantizar, en sí mismo, la inferencia de que la oración *es* verdadera. Pero ¿y si la ascripción de creencias que necesita la teoría evita necesariamente la posibilidad de un error generalizado? Además de satisfacer las restricciones formales, una teoría de la verdad debe estar guiada por una ineludible maximización del acuerdo: no podemos contrastar las oraciones-T sin un principio constitutivo de la ascripción de creencias de los hablantes según el cual se deben asignar condiciones de verdad a las oraciones que se quieren interpretar de forma que la mayoría de las oraciones que el hablante acepta como verdaderas *sean* verdaderas (desde el punto de vista del constructor de la interpretación). Es una presunción que no atañe a creencias específicas, sino a las creencias en general, y que «ayuda a resolver el problema de la interacción entre

En realidad, podría aceptarse la simplicidad del método de Davidson y sostener que la explicación de la conducta general de las personas es más compleja que la de su conducta verbal. Pero no, Davidson dice que la autonomía del significado es esencial al lenguaje y que explica por qué el significado lingüístico no puede explicarse en base a creencias e intenciones extralingüísticas. Hay que distinguir, dice Davidson, entre un lenguaje que *podiera* usarse con un solo fin y uno que *se usa* para un solo fin. Un instrumento que pudiera usarse para un único fin no sería lenguaje. Pero la aserción honesta sola podría producir una teoría de la interpretación, y, de esa forma, un lenguaje que, aunque fuera capaz de más, podría no utilizarse nunca para otros usos. (En la práctica, esta situación es impensable. Alguien que conoce las condiciones bajo las cuales sus oraciones son socialmente verdaderas no puede dejar de entender ni privarse de las posibilidades de la aserción deshonestas, o de las bromas, de los cuentos, los incordios, las exageraciones, insultos, y todo ese tipo de cosas» («Thought and Talk», ITI, p. 164-165).

³³ «On the Very Idea of a Conceptual Scheme», ITI, p. 196; «Belief and the Basis of meaning», ITI, pp. 144-145.

³⁴ «Radical Interpretation», ITI, p. 136.

significado y creencia al restringir los grados de libertad concedidos a la creencia mientras se determina la forma de interpretar las palabras»³⁵. Según Davidson, la mayoría de los medios que ponemos en acción para disponer de evidencia independiente descansaría sobre este *principio de caridad*. Davidson toma este término de Quine³⁶, pero le da un campo de aplicación mucho más amplio: desde la forma lógica hasta las oraciones ocasionales.

VIII. El principio de caridad: coherencia y causalidad.

El principio de caridad determinaría, al menos, la obtención de evidencia para la interpretación de constantes lógicas de un modo parecido a como la describió Quine. La evidencia consistiría en clases de oraciones que siempre se tienen por verdaderas o siempre se consideran falsas (verdades lógicas) y en patrones de inferencia³⁷. Davidson está de acuerdo con Quine en que no

³⁵ «A Coherence Theory of Truth and Knowledge», PTI, p. 316.

³⁶ Quine, a su vez, lo tomó de Neil Wilson que lo enunció en «Substances without Substrata», *The Review of Metaphysics*, 12, 1959). Wilson lo introdujo en una discusión sobre las propiedades semánticas de los nombres propios y las descripciones. Según él, la interpretación del significado que los hablantes dan a nombres propios está guiada por la asunción de que debemos elegir como *designatum* de nombres propios los individuos que hagan verdaderas a más creencias del hablante. Este principio fue muy criticado por H. Putnam quien opone a Wilson lo que llama «principio del beneficio de la duda» («Language and Reality», *Mind, Language and Reality*, Cambridge, Cambridge University Press, 1975, p. 274 y ss.)

³⁷ Según Quine, podemos traducir las conectivas veritativo-funcionales apoyán donos en evidencias como las siguientes: Como una conjunción lógica es verdadera si y sólo si sus dos miembros son verdaderos, podemos traducir una expresión "X" de un lenguaje por "y" si, y sólo si, para cualesquiera dos oraciones *p* y *q* de ese lenguaje, el hablante afirma "*p X q*" si, y solo si, afirma *p* y afirma *q*. Si otra palabra *Y* es tal que para cualquier oración *p* que afirma el nativo niega "*Y-p*", y para cualquier oración *p* que niega, afirma "*Y-p*", entonces podemos traducir "*Y*" por "no".

De este modo, imponemos al hablante nuestra lógica ortodoxa: «La lógica está presente *constructivamente* en nuestro manual de traducción. Y no hay razón para avergonzarse de ello: hemos de basar la traducción en alguna evidencia, y no disponemos de ninguna otra mejor que esa *combinación* del indígena y nuestra lógica» (*Filosofía de la Lógica*, Madrid, Alianza, 1973, p. 142). La lógica se inserta

tenemos más opción que leer nuestra propia lógica en el hablante: «buscamos la mejor manera de ajustar nuestra lógica hasta el punto de obtener una teoría que satisfaga la convención T en el nuevo lenguaje». Para ello, no bastaría con leer en él las conectivas, sino que habría que atender también a la estructura lógica de la teoría de cuantificación de primer orden y la identidad pero es un proceso que tiene lugar no tomando las constantes lógicas una a

en el dispositivo de traducción de un modo elemental por la importancia que tiene «la preservación de lo obvio» en la construcción de un manual de traducción. Si hemos identificado las conectivas de negación y conjunción y después le preguntamos al hablante: «¿p X Y-p?» y para nuestra sorpresa contesta: “Sí”, tenemos un dato que se puede acomodar a teorías muy diferentes. Una podría ser que hemos traducido mal las palabras nativas, que es poco razonable aceptar una oración y su negación y nadie razonable lo haría. Otra teoría diría que hemos traducido bien, pero que el hablante tiene diferentes creencias acerca de la conjunción y la negación, o, más concretamente, que no acepta la ley de no-contradicción. (Esto implica, obviamente, hasta qué punto admitiremos si la notación sigue siendo reconocible como negación: si los hablantes nativos consideraran verdaderas algunas conjunciones de la forma “p y no-p” pero no las consideraran oraciones de las que se sigue lógicamente cualquier otra oración, entonces la notación ha dejado de ser reconocible como negación). La sugerencia de Quine es que sigamos la costumbre de preservar en lo posible la racionalidad del hablante cuyo lenguaje tratamos de comprender: «la estupidez de tu interlocutor, a partir de cierto punto, es menos probable que una mala traducción».

Quine cree que el principio basta para determinar, al menos en el ámbito de los valores veritativos, la interpretación de *algunas* oraciones de casi cualquier parte del cuerpo de conocimiento y del discurso, por pequeñas que sean esas partes. Siempre puede haber partes, por pequeñas que sean, con oraciones cuya verdad no estará garantizada por la interpretación. Quine lo dice así: la verdad lógica queda garantizada por la traducción aunque en un sentido negativo. El canon de preservar lo obvio condena toda interpretación que se aparte de nuestra lógica, y «lo negativo de esta garantía es que no nos asegura que todas nuestras oraciones lógicamente verdaderas se conviertan en verdades del lenguaje extranjero; podría ser que *algunas* de ellas resultaran intraducibles» (*Ibid.*, p. 143).

Para Quine, la construcción de la teoría depende, además, de un conjunto de presunciones que permiten poner en relación de igualdad trozos de lenguaje extranjero con palabras y oraciones del propio lenguaje y que no abarcan sólo la traducción de funciones veritativas, sino también de oraciones ocasionales. Esas «hipótesis analíticas» rebasan lo implícito en las disposiciones del comportamiento lingüístico de cualquier hablante. Por eso «el hecho de que esas traducciones inverificables procedan sin contratiempos no debe entenderse como una evidencia pragmática de la buena lexicografía: el contratiempo, simplemente, es imposible» (*Palabra y Objeto*, Barcelona, Labor, 1968, pp. 81-82 y ss.).

una, sino tratando a esta parte de la lógica como una red que se instala sobre el lenguaje de golpe³⁸.

El segundo paso se concentra en oraciones con términos indicativos, oraciones que se consideran unas veces verdaderas y otras falsas de acuerdo con cambios de acontecimientos en el mundo³⁹. Aquí Davidson se aparta de Quine de una forma muy importante, porque, como dijimos arriba, según él debemos pegarnos tanto como sea posible a lo que el hablante pone a disposición de su audiencia, o sea, a la actitud que tiene hacia la oración. Lo dice así: «El punto crucial en el que coincido con Quine podría expresarse así: toda la evidencia a favor o en contra de una teoría de la verdad (o de la interpretación o de la traducción) se presenta en forma de hechos acerca de qué sucesos o situaciones del mundo causan, o causarían, que los hablantes asintieran o disintieran respecto de cada oración de su repertorio. [Pero] Quine describe los sucesos o situaciones en términos de patrones de estimulación, mientras que yo prefiero una descripción en términos más parecidos a los que usa la oración que se analiza. Quine daría más peso que yo a una gradación de las oraciones en términos de la observacionalidad; y donde él se inclina por el asenso y el disenso por su susceptibilidad de recibir un test conductista, yo reniego del conductismo y acepto actitudes francamente intensionales hacia las oraciones, como la de “conside-

³⁸ «A Coherence Theory of Truth and Knowledge», PTI, p. 315; «Radical Interpretation», ITI, p. 136. Nos llevaría muy lejos analizar la idea de Davidson de que las teorías de cuantificación de primer orden tienen las suficientes virtudes como para hacer de ellas todo lo que se necesita para elaborar una semántica de un lenguaje natural. Las razones que Davidson da para excluir nociones como la cuantificación sustitucional, verdad en un modelo, en un mundo posible, es que sustituyen un predicado monádico, el de verdad, por una relación, y esto dificultaría la recursión sobre la verdad, algo esencial al carácter desentrecomillador de las oraciones T. («In Defense of the Convention T», ITI, p. 68). Davidson entiende que se considera insuficiente su esqueleto lógico porque se cae en el error de pensar que, como los teoremas que demostrarían que la teoría es correcta son triviales, la teoría o el concepto de verdad que ella caracteriza también son triviales.

³⁹ «Radical Interpretation», ITI, p. 147.

rarlas verdaderas"»⁴⁰. O como dice en su réplica a Foster, «Mi forma de intentar explicar el lenguaje y el significado da un uso principal a conceptos como los de creencia e intención y no creo que sea posible reducirlos a conceptos más científicos o conductistas. Trato de dar una explicación del significado (interpretación) que no dé un uso principal a conceptos lingüísticos no explicados... No estropea mis planes el que para decir lo que sabe un intérprete sea necesario usar una de esas nociones llamadas intensionales, que coexiste con las de intención, creencia y otras parecidas»⁴¹.

Si no entiendo mal, lo que Davidson sugiere es esto: La actitud de asentir a una oración, de considerarla o aceptarla como verdadera es, ciertamente, una actitud intencional, una actitud que probablemente sólo se sabe adoptar bajo la asunción de la presencia de muchos factores psicológicos. Pero —como él dice— su objetivo nunca ha sido eliminar actitudes intencionales, sino actitudes o estados intencionales *individualizadores*, o sea, actitudes *intensionales* que tengan un objeto (un objeto proposicional —se dice—). La teoría, según todo lo que hemos visto, *necesita* que la interpretación de oraciones dependa de la atribución de estados intencionales, pero se puede «reducir conceptos complejos y relativamente teóricos a otros intensionales que al aplicarlos estén más próximos a la conducta públicamente observable»⁴². Y su proximidad a la conducta observable no tiene que ver con su reducción a algo más conductista. Incluso podríamos decir que, en la medida en que podemos detectar esa actitud sin saber qué significan las oraciones, podemos concebirla como una relación extensional entre el hablante, una oración y un acontecimiento en el espacio y el tiempo que nos permite acceder a una interpretación *del hablante*.

⁴⁰ «The Inscrutability of Reference», ITI, p. 230-231.

⁴¹ «Reply to Foster», ITI, p. 176.

⁴² «Towards a Unified Theory of Meaning and Action», *Grazer Philosophical Studies*, 1980, p. 4.

¿Qué ocurriría —según Davidson— si en vez de adoptar este método adoptáramos el de Quine? Quine haría depender la interpretación de oraciones ocasionales (a las que se asiente por causas presentes o ausentes en distintos tiempos y lugares) de «patrones de estimulación» sensorial que funcionan como evidencia para su traducción. Esto quiere decir que, para Quine, los patrones de estimulación no son *parte* de la cadena causal que lleva a la creencia (algo que Davidson acepta sin ninguna objeción), sino que son evidencia y fuente de *justificación* de las creencias que estimulan⁴³. El problema es que, si lo vemos así, entonces las interpretaciones no preservarían la verdad todo lo necesario para construir la teoría. «A pesar de abjurar de los datos de los sentidos y de la reducción de términos teóricos a observacionales, la explicación que da Quine del significado y de la evidencia conduce al escepticismo»⁴⁴.

Para Quine, los sucesos y objetos que determinan el significado y proporcionan evidencia para la teoría no son los propios objetos sobre los que versa la oración⁴⁵. Davidson lo dice así: la evidencia para la teoría no pueden ser ni las causas sensoriales de *Word and Object* y *Ontological Relativity*, ni las “observaciones” en *Roots of Reference* ni oraciones de observación, pues éstas pueden desempeñar el papel de evidencia a menos que tengamos *razones* para creer que son verdaderas⁴⁶. Davidson pone este ejemplo: tener la

⁴³ «A Coherence Theory of Truth and Knowledge», PTI, p. 317.

⁴⁴ «Meaning, Truth and Evidence», *Perspectives on Quine*, Barret, R & Gibson, R., (eds.), Oxford, Basil Blackwell, 1990, p. 74; «A Coherence Theory of Truth and Knowledge», PTI, pp. 316-318. Véase también «Reply to Quine on Events», *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, Lepore, E., McLaughlin, B., (eds.), Oxford, Basil Blackwell, 1985, pp. 162-171.

⁴⁵ Davidson muestra cómo el propio Quine dio a entender esto a veces, pero finalmente otorgó más peso a los patrones de estimulación como evidencia última. Todo depende de la localización de los estímulos en la cadena causal que va desde afuera hasta distintas partes del sistema nervioso. Quine los pone en los receptores sensoriales, Davidson los pondría en los propios objetos y sucesos sobre los que versan nuestras oraciones de observación. Por eso llama a la teoría de Quine una *proximal theory* y a la suya una *distal theory* del significado y de la evidencia. Véase: «Meaning, Truth and Evidence», *Op. Cit.*, p. 77.

⁴⁶ «A Coherence Theory of Truth and Knowledge», PTI, p. 311. Para Quine el

sensación de una luz blanca relampagueante puede justificar la creencia de que una luz blanca está relampagueando. Si alguien tiene esa sensación, es probable que haya tal luz. Pero supongamos que esa persona cree que no tuvo la sensación. Entonces la sensación no justifica su creencia de que hay una luz. O sea, no se puede estar justificado en la creencia de que hay una luz si no se tiene la creencia de que se tuvo la sensación. Además, aun si las sensaciones justificaran la creencia en la sensación, ¿cómo justificaría la creencia en acontecimientos y objetos reales? Si los patrones de estimulación son causas, no justifican las creencias que causan, y si suministran alguna información pueden estar engañándonos. Los patrones de estimulación son intermediarios entre las creencias y los objetos usuales sobre los que éstas versan. Pero, así, las oraciones que tenemos por verdaderas podrían ser sistemáticamente falsas. Por tanto, la razón por la que el planteamiento de Quine conduce al escepticismo no es que éste se apoye tanto en la idea de datos sensoriales o cosas por el estilo, sino en la asunción más general de que el conocimiento empírico requiere un intermediario epistemológico entre el mundo tal como lo concebimos y nuestra concepción de él⁴⁷. Desde luego que el significado y la creencia dependen de la experiencia, pero este "depende" es el de la causalidad y no el de la evidencia o la justificación.

La política de Davidson, en cambio, es así de simple: decimos que se comparte un significado cuando idénticos sucesos, objetos

«significado estimulativo» o de «patrón de estímulos» de una oración comprende un conjunto de estimulaciones que incitarían su asentimiento rápido y el conjunto de estimulaciones que incitarían su desaprobación rápida. Esta noción permite aislar una especie de alcance empírico neto de cada oración considerada individualmente, sin apelación al sistema entero de oraciones que la contenga y, al mismo tiempo, sin perder lo que la oración deba a dicho sistema de oraciones. O sea, la noción es sólo un expediente para explorar el edificio de las oraciones encadenadas procediendo una a una. (*Palabra y Objeto*, p. 47). Quine no dice que el significado estimulativo sea el significado *tout court*, porque ninguna oración está completamente desconectada de la red entera de oraciones que forman un lenguaje. Pero, según él, tenemos que atender al significado-estimulativo para darle algún contenido empírico al manual (*Ibid.*, p. 38).

⁴⁷ «Meaning, Evidence and Truth», *Op. Cit.*, p. 74.

o situaciones causan o causarían el asentimiento o disentimiento de la oración. «Como un intérprete radical correlaciono respuestas verbales del hablante con cambios en el contexto circundante. Infiriendo una relación causal, traduzco esas respuestas verbales por una oración mía que me haga aceptar o rechazar los mismos cambios en el contexto»⁴⁸. Así no se niega que ciertos acontecimientos y objetos en el mundo externo causen que creamos cosas sobre el mundo externo y que parte de la causalidad se oriente a través de nuestros órganos sensoriales, pero consideramos directamente a los objetos de una creencia como las causas de esa creencia.

Esto maximiza la caridad: para el intérprete, *no hay forma* de descubrir que el hablante está ampliamente equivocado acerca del mundo, porque él interpretará las oraciones que se tienen por verdaderas según los eventos y objetos externos que causan que la oración se tenga por verdadera⁴⁹. Dicho de otra forma: la distinción entre objetos y estimulaciones no tiene uso en la construcción de la teoría, porque el intérprete radical no puede empezar traduciendo oraciones de observación individuadas por los patrones de estimulación que causan su asentimiento. Si asumimos la interpretación radical —diría Davidson—, no podemos identificar primero creencias y significados y luego preguntar qué las causó. Buscamos una primera comunicación: cuando la obtenemos es posible saber lo que se cree con independencia de lo que causa la creencia. Por tanto, *en el orden de la interpretación* hemos de considerar los objetos de una creencia del hablante como sus causas; y lo que consideremos que son, eso serán de hecho. Conectaremos sistemáticamente la causa de una creencia con su objeto, y una oración significará lo mismo si la creencia en su verdad está causada sistemáticamente por los mismos eventos y objetos⁵⁰. Para identificar o individuar creencias

⁴⁸ *Ibid.*, p. 73.

⁴⁹ «A Coherence Theory of Truth and Knowledge», PTI, p. 317.

⁵⁰ *Ibid.*, p. 318.

hay que suponer que tienen la misma causa que las nuestras. El intérprete radical se ve empujado a interpretar las oraciones que el hablante considera verdaderas según los sucesos y objetos del mundo que causan el asentimiento a la oración como verdadera y, así, identifica las condiciones de verdad con esas causas.

Pero, en realidad, las creencias se identifican por su relación causal con las circunstancias de emisión y con información sobre las relaciones inferenciales que las creencias guardan entre sí (relaciones deductivas pero también de inferencia no-demostrativa). La ascripción de contenido depende también de la posición de la creencia en un patrón o conjunto suficientemente denso de otras creencias coherentes entre sí⁵¹.

IX. La presunción de verdad y la explicación del error

La coherencia, dice Davidson, no garantiza que todo lo que se cree sea verdad. Pero justifica que la mayoría de las creencias sean verdaderas. La coherencia es una *prueba* de la verdad en ese sentido: justifica una presunción a favor de la verdad de las creencias coherentes con un conjunto de otras creencias⁵². Sin el

⁵¹ Obsérvese que la creencia es una actitud causada por cosas y sucesos que, bajo ciertas descripciones (las de un patrón de creencias) son su materia, su objeto. Las acciones intencionales, entonces, serían sucesos causados por acontecimientos que, bajo ciertas descripciones (un patrón de acción—dirigida—a fines), son las *razones* para esas acciones. Esto nos sienta de bruces en la filosofía de lo mental de Davidson y no seguiré por aquí. -

⁵² Esta es la tesis a la que se dedica buena parte de «A Coherence Theory of Truth and Knowledge». El argumento dice más exactamente que, si tenemos razones para pensar que muchas creencias son coherentes, entonces tenemos razones para pensar que muchas son verdaderas. En este único sentido, la coherencia genera correspondencia, aunque «correspondencia» aquí no implica ninguna confrontación o comparación con los hechos. El significado sigue dado por condiciones objetivas de verdad, pero el conocimiento de que las condiciones se satisfacen no requiere una confrontación entre lo que creemos y la realidad. (O como dice en «The Myth of the Subjective»: Las creencias son verdaderas o falsas pero no representan nada). Podemos hablar de la verdad de las oraciones sin mencionar los hechos que representan o los trozos de sensación que les

patrón atribuido tampoco se determina la materia sobre la que versa una creencia particular⁵³. Aquí, la situación hacia el hablante es análoga a la del propio intérprete cuando se pregunta por la verdad de sus creencias: puede creer que una nube está pasando delante del sol, pero sólo si cree que hay un sol, que las nubes están hechas de vapor de agua, que el agua puede existir en forma líquida o gaseosa... La lista de creencias sería interminable: no hay una lista particular que se requiera para dar sustancia a la creencia de que la nube está pasando por delante del sol, pero siempre se necesita un conjunto de ellas que goce de una presunción de verdad⁵⁴. Antes de que un objeto del mundo o un aspecto de él o un suceso puedan formar parte de la materia de la que trata una creencia (verdadera o falsa), debe haber innumerables creencias acerca de esa materia con una presunción de verdad. Demasiadas creencias falsas socavarían la simple identificación de esa materia y, por tanto, la adecuación de una asignación que la vincule con esa materia. Si el intérprete es capaz de entender la oración «a cloud is passing before the sun», debe tener muchas creencias y éstas no pueden ser muy distintas de las que el hablante debe tener si cree que una nube está pasando por delante del sol. ¿Podemos estar seguros de que el hablante cree que una nube está pasando por delante del sol? Y si no cree *nada* de lo que el intérprete cree sobre las nubes, el sol, el agua... ¿seguro que está pensando en nubes? *No* necesitamos respuesta: «No se trata de que cualquier creencia falsa necesariamente destruya la capacidad para identificar otras creencias, sino que la inteligibilidad de éstas debe depender de un trasfondo de creencias mayoritariamente verdade-

corresponden (sin *explicar* la verdad). La idea de “ajustarse” a esas cosas no agrega nada inteligible al simple concepto de verdad y, en suma, no hay que confundir la pregunta «¿qué significa que una oración sea verdadera?» con «¿qué hace verdadera a una oración?» Véase «In Defense of Convention T»; «On the very Idea of a Conceptual Scheme», ITI, y «Representation and Interpretation», *Modelling the Mind*, Newton-Smith, W., (ed.), Oxford, 1990, pp. 13-26.

⁵³ «Thought and Talk», ITI, p. 168.

⁵⁴ «The Method of Truth in Metaphysics», ITI, p. 200. Véase también la primera parte de «Thought and Talk».

ras. O, para decirlo de otra manera: cuanto más acertado se presume que es el hablante, más definidos serán sus errores. Demasiado error simplemente empaña el foco»⁵⁵.

No parece posible, además, establecer una distinción clara entre un error según los patrones del propio hablante y un error según los criterios del intérprete, porque la coherencia interna es en sí misma una norma fundamental. Cuanto más extraña parezca la incoherencia del hablante para el intérprete, más inútil le resultará la distinción entre sus propias normas y las del hablante para explicar el error. Las diferencias pequeñas —digámoslo una vez más— tomarán forma y se podrán explicar sobre el fondo de una semejanza, y, por lo que se refiere a desviaciones importantes, tenderemos a atribuirnoslas como intérpretes. Si podemos entender las creencias del otro, entonces podemos asignar oraciones a esas creencias. Pero una creencia no puede mantener su identidad cuando pierde las relaciones con otras, y, en consecuencia, no es posible que la misma oración sirva para interpretar creencias de dos personas y guarde al mismo tiempo con las demás actitudes de una de ellas relaciones muy diferentes de las que guarda con las de la otra⁵⁶.

Si se tiene en cuenta todo esto, puede entenderse mejor la naturaleza del principio de caridad. Si es un método irrecusable, lo es —dice Davidson— por ser el único disponible al principio de una comunicación⁵⁷. Si no es una opción, sino una condición para tener una teoría practicable, carece de sentido sugerir que, en caso

⁵⁵ Parece que esto también afecta a la interpretación diacrónica. «¿Estamos seguros —pregunta Davidson— de que algunos antiguos creían que la tierra era plana? ¿Esta tierra? Bueno, esta tierra, la nuestra, es parte del sistema solar, un sistema identificado parcialmente por el hecho de ser un racimo de cuerpos sólidos, grandes y fríos, que gira alrededor de una estrella muy grande y caliente. Si alguien no cree *nada* de esto acerca de la tierra, ¿seguro que está pensando en la tierra? Tampoco se necesita una respuesta —dice Davidson—. («Thought and Talk», ITI, p. 168).

⁵⁶ «Deception and Division», en *Actions and Events. Perspectives on the Philosophy of Donald Davidson*. Lepore, E., y McLaughlin, B., (eds.), Oxford, Basil Blackwell, pp. 138-148.

⁵⁷ «A Coherence Theory of Truth and Knowledge», PTI, p. 316.

de respaldarla, podríamos caer en un error generalizado. Hasta que no establezcamos adecuadamente una correlación sistemática de oraciones que se consideran verdaderas con oraciones que se consideran verdaderas, no hay forma de cometer errores. La caridad nos es impuesta, nos guste o no; si queremos comprender a los demás, debemos darles por acertados en la mayor parte de los asuntos. Si podemos producir una teoría que concilie caridad y condiciones formales, hemos hecho todo lo que puede hacerse para asegurar la comunicación. No es posible nada más, y no se necesita nada más⁵⁸. O sea, no es impuesta porque no sólo proporciona al intérprete intuiciones sobre qué teoría de la verdad puede tratar de ajustar a los datos; también es una parte del corpus de constricciones empíricas a las que una teoría de la verdad que se haya elegido tiene que adecuarse⁵⁹. Tampoco puede tomarse literalmente como un principio de "maximización". Por dos razones: (1) porque las oraciones son infinitas en número y no hay ninguna forma útil de contar creencias; y (2) porque, una vez que la teoría va adquiriendo forma, tiene sentido aceptar errores inteligibles y conceder la relativa probabilidad de distintas clases de error⁶⁰.

Sería absurdo suponer que el principio siempre nos obliga a eliminar desacuerdos o que el principio nos impide atribuir creencias falsas incluso cuando eso contribuyera a la comprensión. Se trata de maximizar la comprensión, no el acuerdo⁶¹. En «Thought and Talk», Davidson dice que «lo que hace posible la interpretación es el hecho de que podemos *a priori* descartar la posibilidad de un error completo»⁶², pero esto sólo significa que *a priori* tratamos de hacer posible el *desacuerdo significativo* y esto depende de *alguna* fundamentación en el acuerdo. Sólo así pueden

⁵⁸ «On the Very Idea of a Conceptual Scheme», ITI, p. 197.

⁵⁹ *Ibid.*, p. 197.

⁶⁰ «A Coherence Theory of Truth and Knowledge», PTI, p. 308; «Radical interpretation», ITI, p. 136.

⁶¹ ITI, Introduction, p. xvii.

⁶² «Thought and Talk», ITI, pp. 168-169. Sub. mío.

aparecer errores⁶³. El exceso de error atribuido al hablante le priva de objeto, igual que un exceso de errores en el intérprete le privaría de algo de lo que equivocarse⁶⁴. En resumen: una teoría de la interpretación practicable no persigue el acuerdo y de vez en cuando debe atribuir errores al hablante o al intérprete. Es imposible especificar todas las consideraciones necesarias que conduzcan al tipo correcto de acuerdo. El "tipo correcto" es tan difícil de especificar como una buena razón para sostener una creencia particular⁶⁵. «No se pueden especificar las consideraciones relevantes, porque todo lo que conocemos y creemos sobre la forma como la evidencia apoya a la creencia puede hacerse funcionar para decidir en qué punto la teoría puede tolerar mejor el error, y qué errores son menos destructivos para la comprensión»⁶⁶. La interpretación de una oración depende de que acepte-

⁶³ «On the Very Idea of a Conceptual Scheme», ITI, p. 197.

⁶⁴ «The Method of Truth in Metaphysics», ITI, p. 200. ¿Cómo sabe el intérprete que sus creencias son sistemáticamente o mayoritariamente verdaderas? La respuesta —dice Davidson— está contenida en la pregunta: para poder dudar o preguntarse sobre el origen de sus creencias, una persona debe saber qué es la creencia. Sólo tenemos que reflexionar sobre la naturaleza de la creencia para darnos cuenta de que la mayoría de las creencias son verdaderas. Sabemos que nuestras creencias son por lo general verdaderas porque las creencias son en general verdaderas por naturaleza. Y sabemos que las creencias que por lo general son verdaderas son en general verdaderas. («A Coherence Theory of Truth and Knowledge», PTI, pp. 315-316). No podemos suponer que sabemos dónde descansa la verdad: «no podemos interpretar sobre la base de verdades conocidas, pero no porque no conozcamos ninguna, sino porque no siempre sabemos cuáles son» («The Method of Truth in Methaphysics», ITI, pp. 200 y ss.). De aquí Davidson extrae un corolario: el escepticismo es absurdo, porque del hecho de que no haya nada sobre lo que no pudiéramos estar equivocados extrae la conclusión de que podríamos estar equivocados acerca de todo. No es necesario que haya algo de lo que tengamos que estar absolutamente seguros para poder presumir que generalmente estamos en lo cierto. («The Myth of the Subjective», *Op. Cit.*, p. 165).

⁶⁵ ITI, Introduction, p. xvii.

⁶⁶ «Thought and Talk», ITI, p. 169. A este respecto —sigue diciendo Davidson— «la metodología de la interpretación no es otra cosa que epistemología reflejada en el significado». Y esto podría significar: el acto de rechazar una interpretación o de aceptarla introduciendo modificaciones en otra parte de la teoría depende del grado de "bondad" de esa teoría en comparación con otras, tomando en cuenta no sólo su ajuste a la evidencia y su coherencia con el resto de teorías sino también su simplicidad. Aunque se pudiera formular una regla para equilibrar

mos el hecho de que un hablante considera verdadera a una oración en ciertas circunstancias como evidencia de que la oración es verdadera en esas circunstancias, pero una vez que la teoría empieza a cobrar forma tiene sentido atribuir creencias falsas. Además, no puede esperarse que toda la evidencia se oriente de la misma forma, o sea, puede haber, como dije al principio, diferencias de hablante a hablante, y de un momento a otro para un mismo hablante.

La política general de interpretación es elegir condiciones de verdad tales que los hablantes consideren verdaderas a oraciones cuando esas oraciones sean verdaderas. Ésta es la estrategia general, pero tiene que ser modificada de diferentes maneras. Los hablantes pueden diferir más a menudo y más radicalmente respecto a unas oraciones que respecto a otras, y no hay razón para no tomar en consideración las diferencias individuales observadas o inferidas. La construcción de la teoría no puede depender de decidir qué oración T es la apropiada para una oración del hablante en un momento; más bien, construimos un patrón que preserva las constricciones formales mientras se ajusta la evidencia lo mejor posible. El hecho de que la teoría no erija a cada hablante como acertado no es una inadecuación suya. Queremos hablar de algo nuevo, sorprendente, discutible o incluso desviado. Una teoría para un único hablante también debería partir de sus actitudes hacia las oraciones, pero entonces tendría muchas teorías igualmente elegibles. Las diferencias podrían neutralizarse o aumentarse. Por un lado, «dada una comunidad de hablantes que *aparentemente* tuvieran el mismo repertorio, el intérprete hará lo posible para hallar una única teoría de la interpretación. Así, reducirá mucho la elección práctica de teorías

los factores, sólo evaluaría el carácter general de una teoría muy amplia y por tanto sólo se aplicaría indirectamente a una decisión sobre una interpretación particular y sólo después de que una larga cadena de razonamientos tuviera en cuenta las posibilidades para todas las demás partes. Aunque una regla pueda determinar algún resultado particular no hay garantías de que en un número finito de pasos y de tiempo se pudiera aplicar el criterio para averiguar qué resultado era ése.

preliminares para cada hablante individual»; comenzaría con una teoría con fuerte aplicación a una comunidad para luego refinarla a medida que acumula evidencia particular de cada hablante⁶⁷. Pero una serie de hablantes —dice Davidson— pertenece a la misma comunidad de habla si las mismas teorías de la interpretación funcionan para ellos. Por eso, la apelación a una comunidad de habla en la obtención de evidencia sólo puede servir para simplificar la tarea interpretativa, pero no supone nada de antemano⁶⁸. No es posible determinar antes si preservar la coherencia del individuo o la de la comunidad. En aras de la comunicación, podemos atribuir errores, pero debemos hacerlo sin privilegiar las interpretaciones de la comunidad.

X. Una analogía entre la teoría de la interpretación y la teoría de la decisión.

Hasta aquí he tratado de dejar más o menos claro el papel de las constricciones que tiene una teoría de la verdad cuando se aplica empíricamente. Quizá debería insistir ahora en la naturaleza de una teoría de la verdad, y para eso es bueno volver a la analogía de Davidson entre una teoría de la verdad y una teoría de la medición. En el mundo físico —dije ya— una teoría de la medición de longitudes nos permite asignar números a los objetos, porque contiene un conjunto de restricciones formales para las asignaciones de números. Los números asignados no están determinados sólo por las restricciones, pero lo importante es el patrón entero de asignaciones. Ahora hemos visto que el patrón que importa en la construcción de una teoría de la interpretación

⁶⁷ «Belief and the Basis of Meaning», ITI, pp. 152-153.

⁶⁸ «Radical Interpretation», ITI, pp. 135-136. La condición de miembro de una comunidad de habla depende de la capacidad de interpretar las emisiones de los miembros del grupo construyendo interpretaciones. «Thought and Talk», ITI, pp. 167-168.

está formado por relaciones entre oraciones y creencias. Una teoría de la verdad con restricciones formales, más las constricciones holísticas y caritativas, forma, por así decirlo, un paquete compacto de condiciones que nos permitirían aplicar nociones teóricas como los conceptos de significado y creencia. En ese sentido, igual que la asignación de números para medir magnitudes básicas en física presupone que se cumple el conjunto de condiciones, y si, en definitiva, la satisfacción de éstas puede considerarse constitutiva de la esfera de aplicación de las ciencias que usan esas medidas⁶⁹, entonces, el marco de restricciones que impone una teoría de la verdad sería *constitutivo* de los conceptos asumidos por otras teorías lingüísticas satisfactorias, y, en consecuencia, deberíamos «esforzarnos en ajustar nuestros resultados o nuestras interpretaciones para preservar nuestra teoría»⁷⁰. Por eso, Davidson compara las dificultades que tendríamos para llevar adelante una interpretación sin el trasfondo de patrones inferenciales entre oraciones (forma lógica) y de creencias coherentes, con las dificultades que tendríamos en el mundo físico para descubrir un caso en el que se incumple la transitividad de la relación *más pesado que*⁷¹. La causa de la dificultad es que en ninguno de los dos casos podríamos dar sentido a la atribución de números y significados: no podemos dar un sentido adecuado a la atribución de significados, creencias e intenciones sin suficiente coherencia.

De hecho, hay otra analogía aún más interesante que la de la teoría de la medición y a la que Davidson da mucha importancia: toda creencia está justificada en un conjunto coherente de otras creencias del mismo modo a como lo está una acción intencional de un agente racional, cuyas creencias, elecciones y deseos fueran coherentes en el sentido que tiene ese concepto en una teoría formal de la decisión⁷².

⁶⁹ «Psychology as Philosophy», EAE, pp. 236-237.

⁷⁰ *Ibid.*, p. 273.

⁷¹ *Ibid.*, p. 237.

⁷² «Belief and the Basis of Meaning», ITI, pp. 145-148; «Thought and Talk», ITI, pp. 160-163. El significado de esta analogía no es accidental: A finales de los años

Como hemos visto, en la teoría de la interpretación el problema era cómo determinar creencias y significados dadas las oraciones que se tienen por verdaderas. En el caso de la teoría de la decisión, el problema es parecido: cómo determinar creencias y deseos dadas las preferencias entre cursos alternativos de acción. Las elecciones entre apuestas son resultado de dos factores: valores relativos que el agente adjudica a los resultados y la probabilidad que asigna a esos resultados. Dadas las creencias del agente (probabilidad subjetiva) es fácil determinar sus valores relativos a partir de sus elecciones. Dados sus valores, podemos inferir sus creencias. Pero dadas sólo sus creencias, ¿cómo podemos obtener sus creencias y sus valores? Si se supone que un agente juzga probabilidades de acuerdo con la probabilidad objetiva, se pueden calcular sus valores a partir de sus elecciones. Si sus valores son expresables linealmente (p.e. en dinero), entonces es fácil calcular sus creencias. Pero ¿cómo puede derivarse de la elección una de las dos cosas (valor o creencia) hasta que no se conoce la otra?⁷³. Podemos decir que un agente considera tan improbable que ocurra como que no ocurra un acontecimiento (llover), si no le importa el resultado que implica:

cincuenta Davidson estudió psicología experimental y trató de contrastar la teoría de la decisión de Ramsey que luego fue redescubierta por Von Neumann y Morgenstern. Véase: Davidson, D., y Marschack, J. «Experimental Tests of a Stochastic Decision Theory» en *Measurement, Definition and Theories*, Churchman, C., y Ratoosh, P., (eds.), New York, John Wiley, 1959; Davidson, D., Suppes, P., Siegel, S., *Decision-Making: An Experimental Approach*. Stanford, Stanford University Press, 1957. (Reimpresión en Chicago, Chicago University Press, 1977); «Towards a Unified Theory of Meaning and Action», *Grazer Philosophische Studien*, 2, 1980, pp. 1-12.

⁷³ Davidson pone el siguiente ejemplo: supongamos que un agente se muestra indiferente entre obtener \$5 y una apuesta que le ofrece \$11 si sale cara y \$0 si sale cruz. Podemos explicar, interpretar su indiferencia suponiendo que \$5 está a medio camino en su escala de valores subjetiva entre 0 y 11. Inferimos esto bajo la suposición de que la apuesta vale la suma de los valores de los posibles resultados estimados por sus probabilidades, y asumimos que cara y cruz son igual de probables. Pero también hay otra explicación: puesto que \$5 no está en la mitad de 0 y 11, el agente debe creer que es más probable que salga cruz que cara. Si pensara que cara y cruz tienen la misma probabilidad, debería preferir la apuesta que entonces sería equivalente a una oferta directa de \$5,5. («Belief and the Basis of Meaning», ITI, p. 145).

OpciónAB

Si llueve, ganas10

Si no llueve, ganas01

La indiferencia hacia A y B muestra que juzga tan probable que llueva como que no llueva. Si pensara que es más probable que llueva preferiría la opción A. Si pensara que no es probable que llueva, preferiría B. Así *se separa probabilidad subjetiva y utilidad subjetiva (expected utility)*. Una vez que se descubre un suceso (como la lluvia), es posible confeccionar una escala de valores y luego se puede determinar la probabilidad subjetiva de los sucesos. En esta teoría, la base evidencial son las preferencias entre alternativas (que corresponderían, en la teoría de la interpretación, a la actitud básica de considerar verdadera a una oración). La explicación de una preferencia particular involucra la asignación de una clasificación comparativa de valores y una evaluación de probabilidades. Se trata de observar otras preferencias del mismo tipo que el que hay que explicar. No hay forma de probar independientemente de la teoría las atribuciones de valores y probabilidades subjetivas⁷⁴: «Mi idea —comenta Davidson— es que debemos ver los significados y creencias [en teoría de la interpretación] como constructos interrelacionados de una sola teoría, igual que vemos a los valores y la probabilidad subjetiva en la teoría de la decisión»⁷⁵.

⁷⁴ Además se asignan números para medir los valores subjetivos de los resultados, pero la teoría refleja comparaciones de diferencias (no comparaciones absolutas). Lo importante es que en esta teoría no se juzgan directamente los grados de creencia ni se hacen directamente comparaciones numéricas de valor. Más bien se postula un patrón razonable de preferencia entre cursos de acción y se muestra cómo construir un sistema de creencias y deseos cuantificados que expliquen las elecciones. Por eso, dadas las condiciones ideales que postula la teoría, este método, antes que nada, permite identificar creencias y deseos («The Structure and Content of Truth», p. 156).

⁷⁵ «Belief and the Basis of Meaning», ITI, p. 147.

La teoría se ocupa directamente de relaciones entre acciones (elecciones reales entre distintas apuestas que corresponden a las emisiones en la teoría de la interpretación) y trata a deseos y creencias como constructos teóricos⁷⁶. La cuestión es que, para atribuir inteligiblemente actitudes y creencias, estamos *obligados* a encontrar en las pautas de conducta, creencia y deseo un alto grado de racionalidad y coherencia. La situación, de nuevo, es análoga a la medición de magnitudes básicas en física. Una teoría de la medición establece en forma axiomática las propiedades de las relaciones entre x e y que tienen lugar cuando x es al menos tan pesado, o tan largo, o tan caliente como y . Esta relación debe, entre otras cosas, ser transitiva, reflexiva y simétrica. E igual que es difícil imaginar un experimento capaz de persuadirnos del incumplimiento de la transitividad de la relación *más pesado que*, parece imposible diseñar un experimento que muestre el incumplimiento de la transitividad de la relación *preferir a*. ¿Con qué argumento se nos habría de persuadir de que una persona, *sin cambiar de ideas*, prefiere A a B , B a C y C a A ? «Si la longitud no es transitiva, ¿entonces qué significa utilizar un número para medir longitudes? Podríamos encontrar o inventar una respuesta, pero a pesar de que lo hagamos o hasta que lo hagamos, deberemos interpretar “más largo que” de tal forma que resulte transitivo. Lo mismo ocurre con “preferir a”»⁷⁷.

En ambos casos, los axiomas *no* definen la relación principal (x es tan pesado como y , x se prefiere a y) ni nos enseñan cómo determinar cuándo tiene lugar la relación. «Antes de que se pueda usar o contrastar la teoría se tiene que poder decir algo sobre la interpretación de los conceptos no definidos. Lo mismo se aplica al concepto de verdad»⁷⁸. La teoría de la verdad proporciona una teoría para cada individuo o grupo análoga a la que proporciona

⁷⁶ «Psychology as Philosophy», EAE, p. 234.

⁷⁷ «The Structure and Content of Truth», p. 313. «Hempel on Explaining Action», EAE, p. 273.

⁷⁸ «The Structure and Content of Truth», p. 314.

la teoría de la decisión para cada agente. La teoría de la decisión *no* dice cuándo se aplica a un agente particular. La teoría de la verdad no nos dice cuándo se aplica a un individuo ó a un grupo. En el primer caso tratamos de especificar las condiciones que un agente *debe satisfacer* para poder decir que (poder darle sentido a la idea de que) prefiere una opción a otras. En el caso de la teoría de la verdad, queremos saber cómo decir cuándo las oraciones T (y la teoría considerada como un todo) describen el lenguaje de un grupo o individuo⁷⁹.

No podemos dar sentido adecuado a la atribución de preferencias si no es sobre la base de actitudes coherentes⁸⁰. La interpretación del lenguaje, o del comportamiento verbal, puede, por eso, mostrar los rasgos más destacados de la explicación de la conducta en general. Por eso no podíamos tratar de interpretar las oraciones aisladamente: siempre que nos ponemos a interpretar experimentamos la necesidad de enmarcar la evidencia en las pautas generales de la racionalidad. «La fuerza constitutiva en el ámbito de la conducta deriva de la necesidad de ver a los otros lo suficientemente semejantes a nosotros mismos», y «debemos hacer que la evidencia se adecúe como sea a ese marco»⁸¹. Por eso, la máxima de optimizar el acuerdo no es una asunción que se pueda volver falsa. Si no podemos encontrar una forma de interpretar las emisiones de una criatura y sus actos de elección como manifestaciones de la conducta de una criatura que tenga un conjunto amplio de creencias consistentes, no tenemos razones para considerar que esa criatura es racional, tiene creencias o siquiera «dice algo»⁸². La caridad gobierna la aplicación de conceptos como los de significado y creencia, pero si lo tomamos como el esqueleto de una “psicología popular”, si lo entendemos como una teoría empírica sobre la gente cuyas creencias y preferencias pudieran de-

⁷⁹ *Ibid.*, p. 297.

⁸⁰ «Psychology as Philosophy», EAE, p. 237.

⁸¹ *Ibid.*, p. 239.

⁸² «Radical Interpretation», ITI, p. 137.

terminarse independientemente, entonces cualquier criterio que eligiéramos para determinar creencias podría hacer falso al principio.

De igual forma, las presunciones para la conducta racional no son psicología científica. No se puede, por ejemplo, dar un fundamento conductista: si diéramos una interpretación conductista —dice Davidson— de «alguien prefiere x a y », entonces la teoría de la decisión (cuyos teoremas contienen la palabra “prefiere” *sin interpretar*) se convertiría en una teoría empírica potente, contrastable y manifiestamente falsa. Si la preferencia se interpreta así y buscamos un tipo específico de conducta que sirva de evidencia para la preferencia, las constricciones que la gobiernan se volverían falsas y los experimentos mostrarían que la gente prefiere a veces A a B , B a C y C a A ⁸³: «estaríamos satisfechos con una interpretación que hiciera parecer a la teoría más o menos verdadera sobre un conjunto limitado de alternativas. Si las cosas no funcionaran bien, habría varias explicaciones posibles. Para cada cosa siempre hay que preguntarse si la alternativa al suceso fue la misma que en una jugada anterior. Un sujeto al que se le pide que elija entre A y B y luego entre B y C , podría no considerar a B en compañía de A como la misma alternativa que formara B en compañía de C . O un suceso descrito como una moneda que se tira al aire podría verse como un suceso con distinta probabilidad en la primera tirada que en la sexta. O quizá el sujeto podría no entender nuestro idioma, y la interpretación de su elección dependería de la suposición de que podría interpretar nuestro lenguaje»⁸⁴.

⁸³ «The Material Mind», EAE, p. 272.

⁸⁴ ⁸⁴ 12
bid., p. 270.

XI. La integración de la teoría de la racionalidad.

Esta tercera explicación de fallos en la aplicación de la teoría es muy interesante. Muestra que no sólo hay una analogía entre la teoría de la interpretación lingüística y la interpretación de la conducta: también hay una conexión, y además muy importante. La conexión se manifiesta, como acabo de decir, en el problema de la «presentación de opciones». Para entender las preferencias de un agente, es necesario describir las opciones con palabras. Y no podemos dar por supuesto que las palabras usadas por el sujeto y el experimentador tienen la misma interpretación. Supongamos que ofrezco a alguien una manzana y una pera. Señala manzana y tomo nota de que ha elegido manzana. Al describir sus acción de este modo, doy por supuesto que él quería señalar la manzana y que al señalarla quería indicar cuál era su elección. También doy por supuesto que él creía que estaba eligiendo la manzana. Al atribuir creencias podemos hacer distinciones muy sutiles, tan sutiles como nuestra lengua nos lo permita. No sólo hay una diferencia entre que la persona crea que está eligiendo una manzana y que crea que está eligiendo una pera: Hay incluso una diferencia entre que crea que está eligiendo la mejor manzana de la caja y que crea que está eligiendo la manzana más grande, cosa que puede ocurrir cuando la más grande es la mejor. Todas las distinciones posibles en nuestra lengua se usan en la atribución de creencias, deseos e intenciones⁸⁵. Esto quiere decir que la teoría de la decisión presupone que podemos identificar e individuar los términos de los pares de opciones, y la habilidad para identificar y distinguir esas cosas depende de la capacidad para la interpretación verbal. Sabemos qué se quiere, qué se prefiere, y qué se cree sólo si podemos interpretar el habla. Por tanto, una teoría que deduce las actitudes y creencias que explican las preferencias debería incluir una teoría de la interpretación radical y, claro, ésta debería poder construirse, como dice Davidson, en ausencia de

⁸⁵ «Psychology as Philosophy», EAE, pp. 237-238.

información detallada sobre los contenidos de los deseos, las creencias y las intenciones. Esto nos devuelve a una de las cuestiones con las que empezamos.

Dijimos que había varios medios para explicar el error atribuible a los hablantes: uno de ellos era la relación de las oraciones, el grado o medida en que el hablante acepta que la verdad de una oración está apoyada por la verdad de otras. Esto vale tanto para la interpretación de constantes lógicas, como para los términos observacionales y teóricos. El apoyo evidencial depende de probabilidades condicionales. Las relaciones son holísticas y probabilísticas. Y, por tanto, si queremos identificar creencias y significados debemos poder detectar el grado con que un agente considera verdadera una oración: sus probabilidades subjetivas. El simple asentimiento o disentimiento se encuentran «en los extremos opuestos de una escala y necesitamos ser capaces de localizar actitudes con peso intermedio»⁸⁶, pero el grado de creencia, como acabamos de ver, no puede ser directamente determinado por el intérprete.

Por tanto, nos vemos conducidos a esta dependencia: la teoría de la decisión debe estar libre de la suposición de un acceso independiente a los significados. Pero la teoría de la interpretación (del significado) exige una teoría del grado de creencias para poder hacer un uso serio de las relaciones de apoyo evidencial⁸⁷. La conclusión es ésta: deberíamos integrar la teoría que abstrae las funciones de creencia y significado (interpretación verbal) con la que abstrae creencia y deseo⁸⁸. Pero ¿cómo abstraer grados de creencia y deseabilidad en una escala de intervalos y una interpretación del habla sin asumir que los deseos o las creencias han sido individuados de antemano (o, peor aún, sin suponer que han sido cuantificados de antemano)? Está claro que, en esa teoría integra-

⁸⁶ «The Structure and Content of Truth», p. 322.

⁸⁷ *Ibid.*, p. 322.

⁸⁸ «Towards a Unified Theory of Meaning and Action», *Grazer Philosophical Studien*, 2, 1980; «A New Basis for Decision Theory», *Theory and Decision*, 18, 1985, pp. 87-98.

da, significado, creencia y deseo serían nociones teóricas coordinadas de la interpretación global de la acción humana. Pero la disponibilidad de evidencia para esta teoría tendría que satisfacer la siguiente condición: ya no nos bastaría como evidencia el hecho de que un hablante considera verdadera una oración. Más bien, la evidencia serían hechos sobre las circunstancias o condiciones en las que un agente expresa una preferencia por la verdad de una oración más que por otra, o situaciones y episodios que causan en el hablante la preferencia por una oración. Por decirlo de otro modo, habría que usar elecciones de los agentes entre pares de oraciones, y, por supuesto, el método invocaría las constricciones integrales de caridad. La teoría, por tanto, se basaría en una actitud simple que el intérprete pudiera reconocer en un agente antes de que tuviera conocimiento detallado de cualquiera de las actitudes del agente. Un intérprete puede reconocer esa actitud sin saber qué significan las oraciones, qué estados de cosas valora o qué creencias sostiene (aunque la preferencia sea una función de todas esas cosas)⁸⁹.

⁸⁹ El problema es: ¿cómo podemos decir que un agente ve una oración como algo que presenta una apuesta o jugada hasta que no nos metemos en el proceso de interpretar su lenguaje? Una apuesta, por ejemplo, especifica una conexión, presumiblemente causal, entre la ocurrencia de un determinado suceso (se tira al aire una moneda) y un resultado específico (ganas un caballo). Incluso asumiendo que pudiéramos decir cuándo acepta el agente esa conexión, la directa aplicación de la teoría depende también de que el suceso causante (la moneda) no tenga valor, negativo o positivo, en sí mismo. Y también es necesario asumir que la probabilidad que el agente asigna a la moneda que se lanza no está contaminada por el pensamiento sobre la probabilidad de ganar un caballo. En los tests experimentales de teoría de la decisión, intentamos encontrar contextos en los que esas asunciones tengan posibilidades de ser verdaderas, pero la aplicación general que ahora tendríamos en mente no puede ser tan fácil de determinar. Sobre la forma de desarrollar un modelo en el que las preferencias no tengan como objeto oraciones ya interpretadas, véase la parte final de «The Structure and Content of Truth».

XII. La consistencia y la irracionalidad.

Éste es, pues, en líneas generales, el camino que lleva desde la interpretación lingüística a la explicación global de las creencias y las preferencias de los hablantes. Con la teoría de la decisión determinamos las condiciones que debe satisfacer el agente para poder decir que tiene preferencias; con la teoría de la verdad, lo que queremos saber es cuándo las oraciones T (o la teoría como un todo) describen el lenguaje de un individuo o de un grupo. Una teoría axiomatizada de la verdad y una teoría de la decisión producen, juntas, una teoría distinta para cada hablante o agente individual⁹⁰. Y al integrar las dos teorías, cada interpretación y cada atribución de actitudes será un movimiento en una teoría holística gobernada por principios de consistencia generales para la explicación de la conducta.

Naturalmente, el asunto más importante al que conduce esto es el de la explicación de la irracionalidad. De hecho, Davidson dice que la teoría de la racionalidad que sustituya a la epistemología tradicional no tiene que encarar el problema del relativismo, ni el del escepticismo, pero le surgen «nuevos problemas relacionados con la naturaleza del error, pues es difícil identificarlo y explicarlo si no se restringe de algún modo el holismo que implica una concepción no fundamentalista»⁹¹. La razón puede ser ésta: si queremos dar sentido a la atribución de incoherencias, no podemos concebir sus causas como fuerzas ciegas que operan en la cabeza de un agente sin poder ser identificadas como parte de ella. Las fuerzas ciegas no son irracionales, son no-rationales. Para poder explicar, o más aún, para poder describir la irracionalidad,

⁹⁰ «The Structure and Content of Truth», p. 297.

⁹¹ The Myth of the Subjective», *Op. cit.*, pp. 166-167. Sobre este punto: «Actions, Reasons, and Causes», EAE, pp. 3-19; «How is Weakness of the Will Possible», EAE, pp. 21-42; «Paradoxes of Irrationality», en Wolheim, R, y Hopkins, J., (eds.), *Philosophical Essays on Freud*, Cambridge, Cambridge University Press, pp. 289-305; «Incoherence and Irrationality», *Dialectica*, 39, 1985, pp. 345-354; «Problems in the Explanation of Action», en *Metaphysics and Morality*, Pettit, P., (ed.), Oxford, 1987, pp. 35-49.

necesitamos concebir sus causas como algo próximo a las actitudes del agente. Las preferencias inconsistentes, por ejemplo, las preferencias intransitivas, son intenciones inherentemente inconcebibles. Pero la estructura de la acción no parece estar determinada sólo por el patrón que proporciona la teoría de la decisión. Las elecciones y deseos de un agente racional son coherentes en el mismo sentido que lo son sus creencias, y dijimos que la coherencia, por muy plausible que sea, no asegura que lo que se cree sea efectivamente así. Debe haber una presunción de consistencia a favor de toda acción intencional que sea coherente con un patrón de acciones (tal como la describe la teoría formal de la decisión). Pero —deberíamos suponer— tampoco todas las acciones consistentes tienen por qué ser racionales. El patrón de explicación implica que los aparentes casos de inconsistencia se vean como casos en los que el agente ha cambiado sus fines últimos. (Otra condición *a priori* de la explicación de la decisión individual sería que el abandono de un fin implica la existencia de otro de orden superior. Y la indeterminación que tenga su especificación es —creo que diría Davidson— tan poco preocupante como la de la interpretación de las oraciones).

En el caso de las creencias, la verdad da un patrón; en el caso de la acción, podríamos decir que el interés de la voluntad por la satisfacción, por “el bien”, impone otra estructura unificada. Por eso, la irracionalidad choca con el axioma de que lo que se desea se desea como un bien (o mejor: que todo lo que se desea, se desea en la medida que lo juzgamos bueno). O también: la maximización de la consistencia nos empuja a pensar que nadie puede actuar contra su propio juicio. Igual que el principio de caridad hace posible la aplicación de una teoría de la verdad a un hablante individual, el principio de consistencia hace posible la aplicación de una teoría de la decisión a un agente individual: excluye *a priori* la posibilidad de irracionalidad global. Pero, según Davidson, este modo de explicación no convierte necesariamente en paradojas los casos de debilidad de la voluntad. El holismo no puede hacer imposible hechos como que las personas decidan,

sobre la base de lo que creen y valoran, seguir un curso de acción (preferir x a y) y que, al mismo tiempo, esas creencias y valores les empujen a rechazar la acción (prefiriendo y a x). Davidson se toma mucho trabajo para explicar esto, para explicar las relaciones entre la vida causal y la intencional de las personas, pero su respuesta se complica con demasiados detalles de su teoría de lo mental y, me temo, eso nos llevaría mucho más lejos de lo que me propuse aquí.