

El internalismo de los estados mentales en J. Searle*

PATRICIA MOYA CAÑAS**

Sommario: 1. Introducción. 2. Intencionalidad intrínseca, representación y autorreferencialidad. 2.1. Análisis del contenido intencional. 2.2. Percepción y causalidad intrínseca. 2.3. Consideraciones en diálogo con otros filósofos. 3. Crítica a la ciencia cognitiva. 4. Intencionalidad intrínseca y relativa al observador. 5. La habitación china. 6. La sintaxis no es intrínseca a la física y el olvido de la conciencia. 7. La hipótesis del lenguaje del pensamiento. 8. El caso del robot. 9. Las reglas y los procedimientos formales. 10. El dualismo de la Inteligencia Artificial. 11. Algunas observaciones críticas. 11.1. El biologicismo de Searle. 11.2. La emergencia de los estados mentales a partir de propiedades biológicas. 11.3. La perspectiva de la primera y tercera persona.



1. Introducción

El objetivo de este artículo es explicitar el internalismo que suscribe Searle (S.) en sus obras. Se puede, desde ya, adelantar que no es el internalismo explicado desde la perspectiva de la teoría de la justificación, pues la postura de S. es más bien de confianza y aceptación del conocimiento en un contexto del sentido común; no sería posible enmarcarlo, por ejemplo, en una corriente fundacionalista. El internalismo, en su sentido fuerte, ha sido definido como la teoría que sostiene que los estados mentales de una persona dependen solamente de lo que obtiene u ocurre dentro del propio cuerpo, o más particularmente, dentro de la propia cabeza, suponiendo que lo que ocurre en el interior de una persona es siempre algo de naturaleza física. Pero el internalismo de S. es más complejo que esta descripción porque no suscribe el materialismo eliminativo ni de identidad, sino un biologicismo naturalista. Además su comprensión de la intencionalidad cognoscitiva le permite compaginar el internalismo con la referencia a un estado de cosas.

* Este artículo se ha desarrollado dentro del marco del proyecto Fondecyt n° 1000486, año 2000, titulado «La teoría de la intencionalidad en John Searle».

** Instituto de Filosofía, Universidad de los Andes, Chile

Si bien es cierto que su intento es demostrar que la intencionalidad de los estados mentales no es derivada, sino intrínseca a ellos (cfr. *Intentionality*¹, Introducción, p. 14), no desconoce del todo la importancia del entorno físico y social.

Hay dos temas clave que ayudan a responder a la cuestión central que me he planteado, que es determinar cuál es el internalismo de S.; éstos son: 1º La comprensión de lo que él llama intencionalidad intrínseca, que incluye la explicación de la representación y de la causación autorreferencial que se ve de una manera privilegiada en el análisis de la percepción (ver n. 2). 2º La crítica a la ciencia cognitiva, que S. vincula muy estrechamente al desarrollo de teorías acerca de la Inteligencia Artificial y que se vincula con temas como la diferencia entre intencionalidad intrínseca y relativa al observador, las diferencias entre procesos sintácticos y semánticos a través del famoso experimento mental de la habitación china, y la refutación de otras teorías que desconocen el carácter interno de los estados mentales (ver nn. 3-10). El tema de la conciencia y la posibilidad de una desvinculación entre la conducta y los estados mentales son temas que tenía la intención de tratar, pero en el desarrollo del artículo me di cuenta que merecían un trabajo aparte².

Procederé en primer lugar por el estudio de la intencionalidad intrínseca, como marco esquemático para la comprensión del internalismo, para luego hacer un análisis de la crítica de S. a la ciencia cognitiva que pone de manifiesto más claramente cuál es su pensamiento.

Es muy interesante detenerse en este aspecto de la filosofía de S., porque en ambos temas se abre una perspectiva absolutamente nueva para considerar los estados mentales que se opone a la reducción formalista propia de la Inteligencia Artificial (IA).

Al iniciar el estudio de S. barajaba la hipótesis de que su teoría de la intencionalidad podría explicarse al margen de su naturalismo biologicista, pero hasta donde ha llegado esta investigación, parece que es muy difícil trazar una línea demarcatoria entre estos dos temas y que es necesario enfrentar más a fondo este presupuesto del cual parte S. Algunas ideas relativas a este punto están en la parte final del artículo en la que hago algunas observaciones críticas. Sin embargo, es necesario analizar más detalladamente los problemas que genera esta posición biologicista y ver si son o no solubles.

2. Intencionalidad intrínseca, representación y autorreferencialidad

En *Int.* el internalismo (i.) está directamente ligado con los conceptos de representación y con el tema de la causación autorreferencial. El término «repre-

¹ J. SEARLE, *Intentionality, an Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge 1983. De ahora en adelante citaré esta obra abreviadamente: *Int.*

² Cfr. J. SEARLE, *The Rediscovery of the Mind (TRM)*, MIT Press, Cambridge (MA) - London, England ⁶1995=1992, chapter III.

sentación» lo entiende S. en esta obra de la misma manera que lo ha interpretado en *Actos de Habla*³: significa que una creencia representa sus condiciones de satisfacción en el mismo sentido en el que lo hace un enunciado, esto es, dando a conocer en su contenido intencional⁴ estas condiciones de satisfacción. La representación se distingue de los elementos sintácticos de la proposición, pues se refiere a su significado o aspecto semántico. La autorreferencialidad de la causalización quiere decir que lo que especifica las condiciones de satisfacción de la experiencia perceptiva está dado por el contenido perceptivo mismo y no por la experiencia perceptiva, pues, como dice S., lo que vemos es la cosa y no la experiencia de ver la cosa (cfr. *Int.* p. 37).

En *The Rediscovery of the Mind (TRM)* el i. se vincula con el enfoque de la primera persona para el estudio de los estados mentales; el subjetivismo de estos mismos estados; la admisión de la conciencia; la refutación, en esta línea, del funcionalismo, la Inteligencia Artificial Fuerte (IAF) y el conductismo (cfr. cap. 1-3). Conviene tener en cuenta que en *Int.* el tratamiento del i. es más explicativo que polémico, es en esta obra donde S. traza los fundamentos de su posición. En *TRM*, en cambio, hay una polémica abierta contra posturas externalistas que amplía la visión de *Int.* y también se introduce explícitamente el tema de la conciencia. Comenzaré por la explicación de *Int.* que es básica para comprender la obra posterior.

2.1. Análisis del contenido intencional

Ya en la Introducción de *Int.* S. hace afirmaciones que son relevantes para la posterior comprensión de la intencionalidad. Parte del supuesto de que tenemos en el lenguaje una demostración de la capacidad de representar objetos y estados de cosas del mundo. Esta capacidad es una extensión de las capacidades biológicas más fundamentales de la mente para relacionar el organismo con el mundo, la cual se establece por medio de estados mentales, como son la creencia o el deseo y más particularmente a través de la acción y la percepción. En esta obra se trata de dar cuenta de cómo se establece la relación intencional.

En *Actos de Habla* se preguntaba S. por qué podemos considerar las oraciones, que desde un punto de vista son objetos compuestos de sonidos que surgen de la boca de alguna persona o marcas que se hacen sobre el papel, como realidades significativas, es decir, que representan objetos o estados de cosas. La respuesta es que la mente hace que estos «objetos del mundo» sean representativos, porque los considera desde su punto de vista propio, es decir, mental, el

³ J. SEARLE, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge 1969.

⁴ Searle escribe Intencionalidad con mayúscula para diferenciarla de la intencionalidad de la voluntad, pero teniendo claro que siempre me referiré a la intencionalidad de los actos mentales, y más propiamente del conocimiento, no seguiré ese uso.

cual es intrínseco a ella y que coincide también con una intencionalidad propia (cfr. cap. 1)⁵.

En los dos primeros capítulos de *Int.* se afirma explícitamente que la teoría de la intencionalidad que se sustenta es internalista, en oposición a la mayoría de las teorías contemporáneas que son externalistas, porque según S., hablar de la intencionalidad de la mente equivale a hablar de una característica que es intrínseca a los estados mentales, que no se deriva de alguna forma previa de intencionalidad. Son externalistas las teorías que explican los estados mentales por causas extrínsecas, como pueden ser aspectos culturales, estructurales, lingüísticos, etc. Estas teorías no reconocen la existencia, o restan importancia, a los rasgos que pueden ser explicados por la misma operación mental. El creer y el desear son de por sí intencionales, de tal modo que en una oración se puede distinguir un aspecto sintáctico y otro intencional que se impone al aspecto sintáctico. Este último es, por así decirlo, el que confiere a la oración su carácter de objeto, mientras que el intencional es el que le da un carácter representativo (cfr. pp. vii-viii).

S. describe los estados intencionales como compuestos de un contenido intencional en un modo psicológico. Cuando ese contenido es una proposición completa y hay una dirección de ajuste, el contenido intencional determina las condiciones de satisfacción. La especificación del contenido intencional es ya una especificación de las condiciones de satisfacción. Por ejemplo, si tengo la creencia de que está lloviendo, el contenido de mi creencia es que está lloviendo y las condiciones de satisfacción son que efectivamente esté lloviendo. Una clave para entender la representación son las condiciones de satisfacción. Todo estado intencional con una dirección de ajuste es una representación de sus condiciones de satisfacción (cfr. p. 13).

Para entender que las condiciones de satisfacción de la intencionalidad están en el interior del contenido de la proposición, no hay que considerar a esta última como objeto de un enunciado o creencia, sino atender a su propio contenido. Esto es importante por el papel que juega el sujeto en este análisis ya que en el estado intencional la relación no se establece entre la persona y la proposición, sino que es más bien una relación de representación entre el estado intencional que se da en la persona y la cosa representada por este estado (cfr. p. 19). «En mi explicación, el contenido Intencional, que determina las condiciones de satisfacción, es interno al estado Intencional: no hay manera que el agente pueda tener una creencia o un deseo, sin que éstos tengan sus condiciones de satisfacción. Por ejemplo, parte de lo que es tener la creencia consciente de que está lloviendo, es

⁵ Tanto el lenguaje como la intencionalidad son esencialmente formas sociales, esto quiere decir que están, en último término, anclados en un trasfondo que no es un inconsciente colectivo, sino más bien un «suelo» común en el cual se da la existencia humana. Esta perspectiva está más ampliamente desarrollada en *The Construction of Social Reality*, The Free Press, New York - London - Toronto - Sydney - Singapore 1995, cap. VI, especialmente p. 147. Hago desde ya esta advertencia acerca del pensamiento de S. para no considerar su i. como una suerte de solipsismo.

ser consciente de que la creencia es satisfecha si está lloviendo e insatisfecha si no lo está» (p. 22)⁶. Se elimina así la necesidad de postular un tercer juicio, pues basta con el del agente, porque no es posible que éste pueda tener creencias o deseos sin sus propias e internas condiciones de satisfacción.

De este modo queda claro en qué sentido una creencia es intrínsecamente una representación: lo es en cuanto que consiste simplemente en un contenido intencional y un modo psicológico. El contenido determina sus condiciones de satisfacción y el modo determina que esas condiciones de satisfacción se representen con una cierta dirección de ajuste establecida por el modo. En el ejemplo nombrado, el de la creencia, la dirección de ajuste es de la mente al mundo, pero en el modo psicológico del deseo, ésta es del mundo a la mente. La creencia y el deseo no se pueden separar del contenido representativo, y si el agente es consciente de las condiciones de satisfacción de éstas, es porque están implícitas en las mismas creencias y deseos. Es, insisto, en este sentido en el que el contenido intencional es interno a los estados intencionales tal como lo señala el ejemplo citado (cfr. pp. 22-23). Esta cuestión tiene importancia, porque sitúa la intencionalidad, con sus contenidos y modos psicológicos, en un nivel que podríamos llamar autorreferencial, en la medida que no necesita de un elemento externo de explicación. Da cuenta de sí misma porque expresa sus condiciones de satisfacción. Esta característica le permitirá a S. mantener una postura realista que se hará más evidente al analizar la intencionalidad de la percepción.

Antes de continuar con la explicación de la autorreferencialidad de la intencionalidad, conviene exponer otro acercamiento al internalismo, posterior al de *Int.*, y que recoge una distinción que será frecuente en los escritos de S.

En un artículo en el que replica a las críticas suscitadas por otro artículo anterior⁷, S. distingue entre intencionalidad intrínseca e intencionalidad adscripta desde fuera (*intrinsic intentionality and observer-relative ascriptions of intentionality*). La primera se refiere a estados mentales actuales y la segunda son modos de hablar que tenemos acerca de entidades que están en nuestras actividades, pero que no tienen una real intencionalidad intrínseca. Por ejemplo, si digo que es-

⁶ Esta teoría coincide en algunos aspectos con la de Tarski; por ejemplo, en el valor semántico de la verdad y también está en la línea de Davidson, aunque este último no coincide plenamente con Tarski, pues da a la verdad, y especialmente al significado, un sentido más pragmático. El acuerdo con Searle en este punto concreto que se trata en el artículo estaría en la afirmación de Davidson de que una teoría del significado puede ponerse a prueba empíricamente, comparando alguna de sus consecuencias con los hechos. Cfr. A. TARSKI, *The Semantic Conception of Truth and the Foundations of Semantics*, en Feigl and Sellars 1949, pp. 52-84. Ver también, D. DAVIDSON, *Truth and Meaning*, «Synthese», 17 (1967), pp. 304-323. Para una discusión que hace Davidson de la teoría de Tarski, ver: M. DEVITT, *Realism and Truth*, Princeton University Press, Princeton, New Jersey 1997² y R. RORTY, *Truth and Progress, Philosophical Papers*, Volume III, chapter I, Cambridge University Press, Cambridge 1998.

⁷ *Intrinsic Intentionality: Reply to Criticisms of Minds, Brains and Programs*, «Behavioral and Brain Sciences», III (1980), pp. 450-456.

toy enojada o que el candidato presidencial cree que puede ganar las elecciones, me refiero a dos formas de intencionalidad intrínseca. En cambio, si digo que X es el candidato presidencial o utilizo la sentencia: “«Es regnet» significa está lloviendo”, tanto X como la frase alemana «Es regnet» no tienen una intencionalidad intrínseca, sino que son dos casos de adscripción de intencionalidad hechas a dos entidades que no tienen ningún estado mental. Esta adscripción es necesaria para el lenguaje y la comprensión de las sentencias. Así sabemos que el nombre se refiere al candidato y que la frase alemana tiene ese significado determinado. La adscripción de intencionalidad relativa a un observador depende siempre de la intencionalidad intrínseca del observador, porque sólo se da la intencionalidad en el sujeto.

En este sentido sólo es auténtica intencionalidad la intrínseca, porque aunque le adscribamos intencionalidad a una entidad, ésta es siempre relativa al observador y la razón de otorgarle intencionalidad es más bien de orden práctico, para entendernos en el lenguaje ordinario (cfr. pp. 451-452). La misma cuestión la expone más claramente en un artículo de 1990⁸, a través de cuatro sentencias en las que se distinguen los diversos tipos de intencionalidad:

A. Robert cree que Ronald Reagan es el presidente.

B. Bill ve que está nevando.

C. «Es regnet» significa está lloviendo.

D. El termostato de mi auto percibe los cambios de temperatura del motor.

En esta ocasión S. se refiere a diferentes *status* de adscripción de intencionalidad. El caso A simplemente adscribe a una persona un estado mental intencional, una creencia; el caso B hace más que el anterior, pues no sólo dice que una persona tiene una forma de intencionalidad, sino también que se trata de una intencionalidad satisfecha. Los verbos «ver» como «conocer», a diferencia de «crear» son verbos de «éxito». B da cuenta de este éxito, es decir de la intencionalidad satisfecha. A y B son fenómenos intencionales intrínsecos en la mente o cerebro de los agentes; la adscripción de intencionalidad en ambos casos ha de ser considerada literalmente, no como una forma de expresión. En el caso C la intencionalidad también es literal, pero *no intrínseca*, sino *derivada*. Literalmente es verdadero lo que se afirma, pero la intencionalidad no es intrínseca a la sentencia, pues como ya se ha expuesto a propósito del artículo anterior, se deriva de la intencionalidad intrínseca de los hablantes del alemán. Finalmente, D no es una intencionalidad literal ni intrínseca, sino sólo metafórica, porque no es una expresión que usa la intencionalidad propia e intrínseca de algún agente. Resumiendo, sólo son estados intencionales A y B; aunque más sutilmente se pueda decir que A es un estado intencional y B un evento intencional. C ilustra la distinción entre intencionalidad intrínseca y derivada y D se diferencia de las otras tres en que no

⁸ J. SEARLE, *Intentionality and its Place in Nature*, en Cole, David J. - Fetzer, James H. - Rankin, Terry L. (eds.), *Philosophy, Mind, and Cognitive Inquiry*, Kluwer Academic Publishers, Dordrecht – Boston – London 1990, pp. 267-280.

es literal, ni tampoco intrínseca, sino metafórica⁹. Desde esta perspectiva se mantiene la afirmación hecha en *Int.*, pero destacando que los estados mentales son siempre intrínsecos, en este sentido, ontológicamente subjetivos y no relativos al observador.

En *The Construction...*, S. da varias vueltas a esta diferencia, utilizándola especialmente para separar los hechos brutos y los hechos institucionales. El ejemplo del destornillador que pone ahí nos puede ser útil para explicar lo que quiere decir. Las moléculas y elementos físicos que componen este objeto son una característica intrínseca y el que sea un destornillador, es decir, el uso que el hombre le ha dado a ese objeto, es una característica relativa al observador, pues este objeto es solamente tal en cuanto que un agente consciente lo trata como destornillador; ahora bien, el hecho por el cual el agente tiene esta actitud es una característica intrínseca del agente: su estado de ser consciente que le permite tratarlo así. S. quiere mostrar con esto que dentro de las características intrínsecas de la realidad se incluyen los estados mentales o de conciencia, así se distancia de la postura heredada del cartesianismo, en la que toda inclusión de la conciencia supone, *ipso facto*, la exclusión de toda realidad distinta de ella y con un modo de ser intrínseco. S. quiere reconocer tanto los estados mentales que llevan consigo un modo de considerar o ver las cosas, como lo que las cosas son intrínsecamente en sí mismas. De acuerdo con esto, lo propio de los estados mentales, la característica que los define como tales es su carácter intrínseco¹⁰.

2.2. Percepción y causalidad intrínseca

En el segundo capítulo de *Int.*, que se refiere a la percepción, S. advierte acerca de la equívocidad que propone la metáfora espacial externo-interno, ya que nuestro cuerpo forma parte del mundo externo y lo mental no difiere esencialmente de lo corpóreo¹¹. ¿Dónde se supone que está el mundo inter-

⁹ Cfr. *art. cit.*, pp. 267-269. Esta misma distinción la había ya hecho en *TRM*, capítulo III, V.

¹⁰ Cfr. J. SEARLE, *The Construction of Social Reality*, pp. 9-11.

¹¹ Es interesante para la comprensión de la postura de Searle aclarar en qué sentido para él lo mental no difiere de lo corpóreo. En prácticamente todos sus escritos afirma que la mente es un fenómeno biológico natural del mundo como cualquier otro. El argumento que se repite a lo largo de su obra es el siguiente: los fenómenos mentales tienen una base biológica: están causados por las operaciones del cerebro al mismo tiempo que realizados en su estructura. De acuerdo con esto la conciencia y la intencionalidad son parte de la biología humana, como la digestión o la circulación de la sangre. Con esto no se niega la realidad de los fenómenos mentales, sino que se afirma su naturaleza biológica, sin pretender buscar otras explicaciones causales para ellos. Cfr. entre otros escritos, *Int.*, p. IX; *TRM*, pp. 86-93; Blakemore, Colin - Greenfield, Susan (eds.), *Minds and Brains Without Programs in Mind-waves*, Basil Blackwell 1987, pp. 209-233. La cuestión que se plantea en este punto es si pueden explicarse de una manera satisfactoria los fenómenos mentales desde un punto de vista exclusivamente biológico.

no? ¿En qué sentido están mis experiencias perceptivas ‘aquí dentro’ y el mundo ‘allí fuera’? (cfr. p. 37). Para responder estas cuestiones es necesario apelar a la intencionalidad intrínseca de la percepción y entenderla adecuadamente. La afirmación del internalismo de los eventos mentales puede llevar a la pregunta: ¿por qué puede ser relevante la presencia o ausencia, por ejemplo, de un auto en la percepción visual del mismo, puesto que después de todo la experiencia visual es simplemente un evento mental y, por lo mismo, interno? La respuesta es que la experiencia visual es un fenómeno intrínsecamente intencional. El contenido intencional es interno a cualquier fenómeno mental y determina por sí mismo sus condiciones de satisfacción, esto quiere decir que no hay, por una parte, un fenómeno visual considerado como un evento mental y, por otra, el objeto real, porque entonces habría que conectar lo externo con lo interno (cfr. pp. 38-39). La experiencia visual no «representa» solamente el objeto, sino que proporciona el acceso directo a él. Sucede en la percepción algo diferente a la creencia: la experiencia perceptiva tiene una especie de claridad, inmediatez e involuntariedad, que no es compartida por la creencia, que se podría tener sobre el objeto en su ausencia. Por eso, no cabe describir las experiencias visuales como representaciones: «Más bien, a causa de los rasgos especiales de las experiencias perceptivas, propongo llamarlas ‘presentaciones’» (p. 46). Porque la experiencia visual no representa solamente el estado de cosas percibido, sino que más bien, cuando está satisfecha, nos da acceso directo a él y en este sentido es una presentación de ese estado de cosas. La ‘presentación’ vendría a ser una subclase de la ‘representación’ (cfr. *ibidem*).

El contenido intencional de la experiencia visual requiere que ésta sea causada por el estado de cosas percibido. El contenido intencional no supone solamente que haya un estado de cosas en el mundo, sino más bien que el estado de cosas en el mundo deba causar la misma experiencia visual, que es la materialización o realización del contenido intencional. ¿Cómo explicar esta causalidad más allá de la «teoría causal de la percepción» que dice que la presencia y rasgos del objeto causen la experiencia del agente? Lo que S. quiere es explicar la causalidad como un hecho que entra dentro del contenido intencional, es decir, situar la causalidad en un ámbito interno de la intencionalidad. Para esto, pone el siguiente ejemplo, o fórmula: «Tengo una experiencia visual (que hay un *station wagon* amarillo allí y que hay un *station wagon* amarillo allí está causando esta experiencia visual)» (p. 49). Con el ejemplo quiere dejar claro que es el mismo contenido intencional de la experiencia visual el que determina bajo qué condiciones se satisface o no, lo que la hace o no ‘verídica’. Es importante advertir que para que la escena del *station wagon* sea verídica, se debe dar un presupuesto que haga válidas las condiciones de satisfacción, a saber: «el mundo debe ser como visualmente me parece que es y por consiguiente su ser de esa manera debe ser lo que me cause el tener la experiencia visual que constituye el que me parezca que es de esa manera» (pp. 48-49). Y al decir visual, se incluyen todas las demás for-

mas de percepción. Este presupuesto no es obstáculo para admitir errores perceptivos, pero en un marco general, se admite que el mundo es como normalmente se percibe.

En este sentido, la experiencia visual es autorreferencial también en cuanto a la causalidad, porque ésta figura en sus propias condiciones de satisfacción. Lo que se ve son objetos y parte de las condiciones de satisfacción de la experiencia visual de verlos es que la experiencia en sí misma debe ser causada por lo que se ve (p. 42), porque no hay experiencia perceptiva sin objeto percibido. La percepción sólo se da o se entiende en su vinculación a un objeto, o mejor, a una cosa. De acuerdo con esta explicación, la percepción es una transacción intencional y causal entre la mente y el mundo. La dirección de ajuste es mente-a-mundo. La dirección de causación es mundo-a-mente. Estas dos direcciones no son independientes. Por eso afirma S.: «La introducción de la noción de autorreferencialidad causal de ciertas clases de Intencionalidad -una autorreferencialidad que se muestra, pero no se dice- es una adición crucial al aparato conceptual de este libro» (p. 49). Y lo es, porque con esta noción S. mantiene, a la vez, el carácter interno de la intencionalidad y la vinculación de los estados mentales con un mundo externo. Si bien es cierto que el núcleo explicativo de la intencionalidad es intrínseco y por eso autorreferencial, la causalidad que se requiere en un tipo de intencionalidad como es la percepción, nos muestra que no toda intencionalidad puede desligarse del mundo.

Lo que se ha explicado en este primer apartado será útil para comprender por qué S. distingue el uso que hace del término «representación» la Inteligencia Artificial contemporánea y la psicología cognitiva y el que él ha propuesto. S. define la representación por su contenido y su modo y no por la estructura formal —que sería característica de la representación en la IA—, ya que no comparte la convicción de que tenga tal estructura formal, sino que tiene, como se ha visto, un contenido y unos modos de ese contenido.

2.3. Consideraciones en diálogo con otros filósofos

Antes de pasar al apartado siguiente, en el que se expone la crítica de S. a las ciencias cognitivas y muy particularmente a la Inteligencia Artificial, que está más desarrollada en *TRM* y otros textos posteriores, quisiera plantear algunas dificultades que surgen del análisis que hasta aquí ha discurrido de una manera más bien descriptiva.

El primer punto es la distinción que establece S. entre el aspecto sintáctico y el semántico de las proposiciones. En tradiciones filosóficas diferentes, que también se ocupan del tema de la intencionalidad, como es, por ejemplo, la de Husserl, este divorcio no se da, sino que el aspecto sintáctico es reconstruido como originado *en y a partir* de los actos de carácter intencio-

nal¹². Esta diferencia no es sólo una discrepancia de puntos de vista, sino que encierra o plantea un problema en S., a saber, el de concebir el aspecto semántico como algo en cierto modo montado sobre el sintáctico y como exclusivo de lo mental. Esto se debe ciertamente, como se verá más claramente en el apartado II, a su rechazo a una visión meramente formal del lenguaje y de las proposiciones, que es lo propio de IAF. Pero tiene el peligro de distanciar excesivamente la semántica de la sintaxis y comprender a esta última como si estuviera ya dada fuera del acto de habla, lo que termina empobreciendo la misma semántica o dejándola sin explicación, porque no se sabría bien de qué manera se la considera algo más que el correcto uso de unas reglas. S. dice que el aspecto semántico es característico de lo mental, pero es difícil comprender en qué consiste realmente la particularidad de lo mental, además de ser un rasgo de nivel superior del cerebro, si se lo disocia de la explicación de cómo se alcanza este uso adecuado de las reglas de una manera que no sea exclusivamente formal, es decir, si no reconstruye, de alguna manera, el proceso o la vía por la que el hombre adquiere un nivel semántico. Aunque S. da importancia a la percepción, no afina en el modo como ésta es el dato originario de la proposición intelectual, y de qué manera el dato perceptivo adquiere una significación diferente, sin que signifique

¹²Para Husserl es fundamental en una explicación de la intencionalidad, partir del nivel de la receptividad sensible y establecer la relación entre este nivel y el correspondiente al de la espontaneidad intelectual. Esto se ve ya en una fase temprana de su pensamiento como es la de las *Investigaciones Lógicas*, particularmente la Sexta Investigación, «dedicada a la elucidación fenomenológica del conocimiento, en la cual el fenómeno del conocimiento es tratado en términos de lo que Husserl denomina la síntesis de ‘repleción’ o ‘cumplimiento’ (*Synthesis der Erfüllung*)». Estos conceptos apuntan «en su nivel primario y más básico de aplicación al mismo tipo de convergencia entre actos intencionales de receptividad sensible y de espontaneidad intelectual que también ocupa el centro de atención en el tratamiento de *Experiencia y Juicio*. (...) De hecho, Husserl dedica toda la primera sección del desarrollo de *Experiencia y Juicio* (cfr. parágrafos 15-46) a un análisis detallado de aquellas estructuras básicas de la experiencia receptiva antepredicativa, a partir de las cuales puede explicarse luego, en la parte central de la segunda sección (cfr. parágrafos 47-65), la génesis de las formas lógico-categoriales básicas, en particular, la de la forma básica del juicio predicativo». Para la reconstrucción de la génesis del juicio predicativo, que Husserl lleva a cabo fundamentalmente en el parágrafo 50 de *Experiencia y Juicio*, son muy importantes las correspondientes «preestructuraciones» en el nivel de la síntesis pasiva de la receptividad, porque «de acuerdo con las premisas básicas del modelo genético husserliano, todo ‘paso’ (*Schritt*) en el nivel de la predicación presupone un ‘paso’ correspondiente en el nivel antepredicativo de la experiencia receptiva y su explicitación. En tal sentido, Husserl establece a modo de principio que «sólo puede ser predicado originariamente lo que está originariamente dado, aprehendido y explicitado en la intuición». Más concretamente, la forma básica y elemental del juicio predicativo, representada aquí por un juicio perceptivo de la forma «*S es p*», se deriva de modo inmediato a partir de su origen en el despliegue explicitante de un objeto dado en la percepción (cf. parágrafo 49). Cfr. A. VIGO, *La concepción husserliana del origen del juicio predicativo en Erfahrung und Urteil*, «Escritos de Filosofía», Buenos Aires, 37 (2000), pp. 235-272.

que esté desconectado de la información de los sentidos. Esta dificultad hace posible aplicar la crítica de desconexión entre sintaxis y semántica que él hace a la IAF a su propia perspectiva, lógicamente en sentido inverso y menos agudo.

El otro aspecto en el cual S. discrepa, por ejemplo, con la visión de Heidegger es en el de la distinción entre aquello que es intrínseco y lo que es relativo al observador en la consideración ontológica de las cosas. La dificultad que se presenta en S. se da especialmente en los objetos artificiales, en los que es difícil decir cuál es su ser en sí o su sentido intrínseco, pues han sido hechos con un propósito, para un uso que siempre es relativo a un agente, y en ese sentido a un observador. S. señala que poder usarlo como instrumento es lo propio del sujeto dotado de intencionalidad intrínseca. Sin embargo, se le puede objetar que propiamente no podríamos hablar en el caso del destornillador de un aspecto intrínseco, o que éste sólo podría ser «su condición de objeto de metal y plástico», pero no podríamos llamarlo intrínsecamente destornillador, porque esta es una característica que le damos con el uso. Es difícil comprender cómo se podría captar el aspecto intrínseco del destornillador en cuanto tal, desligado de su uso. En ese sentido, para Heidegger, el uso es muy importante dentro de la categorización ontológica, porque fenomenológicamente es lo primero con lo que nos encontramos en nuestro conocimiento del objeto. Es el ser a la mano, del cual parte en sus análisis¹³.

También se puede analizar más detalladamente el tratamiento que S. hace de la causalidad perceptiva. Para él tiene mucha importancia introducir la noción de autorreferencialidad causal para explicar la intencionalidad, pues tanto el contenido intencional como la causalidad, se requieren mutuamente (cfr. *Int.*, p. 49). Esta es su manera de explicar la intencionalidad perceptiva sin necesidad de una especie de deducción o inferencia a partir de premisas, sino en nuestro contacto directo con el mundo. Este aspecto tiene gran relevancia, pues refuta una concepción representacionista, necesariamente deductivista. Sin embargo, habría que ver a qué tipo de causalidad se está refiriendo S., pues la causalidad perceptiva se puede entender en un nivel semántico, pero no en el nivel en el que habitualmente hablamos de causalidad, es decir, como un efecto producido por un agente. Si entendemos la causalidad desde una perspectiva física, por ejemplo si se explica la visión como la incidencia del color rojo en la pupila, inmediatamente caemos en un análisis de la intencionalidad desde la tercera persona y, por lo tanto, desde un plano exclusivamente físico.

S. insiste además en que la dependencia causal no es parte de la experiencia perceptiva, o mejor dicho, que esta dependencia causal no es percibida, pero sí es intrínseca al contenido intencional de la percepción, como lo son sus condi-

¹³Cfr. M. HEIDEGGER, *Ser y Tiempo*, párrafos 14-18, en los que analiza el mundo a partir de la determinación ontológica del ente intramundano como 'a la mano'; éste *es* el ser *en sí* del ente intramundano *qua* mundano.

ciones de satisfacción. Surge la cuestión: ¿en qué plano nos estamos moviendo? En sentido estricto, el que la percepción sea verdadera, porque efectivamente nos vincula con el mundo en el que vivimos, es más bien una aplicación semántica de la causalidad, pero que no se puede explicar en términos causales de uso normal. De hacerlo así, nada separa esta explicación de una explicación física, pero lo que se quiere es explicar la intencionalidad, y más concretamente el carácter interno de esta. Si se quiere ver desde un punto de vista interno, habría que partir de la experiencia misma de la intencionalidad perceptiva en la que lo percibido se identifica con el que percibe y, por lo tanto, lo percibido se identifica con la causa. Sólo teóricamente, es decir, desde una análisis de la tercera persona, se puede tratar acerca de lo que causa la percepción, pero siempre perdiendo en este caso el carácter directo e interno de la percepción. Esto es así, porque la relación entre el perceptor y lo percibido tiene una dirección que va del perceptor a la cosa. No hay una conciencia temática de cómo se da esta relación, sino que sólo se percibe y esta experiencia supone, de una manera atemática, la causa. Esto también lo ve, en cierto modo Searle, cuando habla de la percepción como ‘presentación’, pero no lo lleva hasta sus últimas consecuencias. El error perceptivo puede ser la ocasión de cambiar del nivel de la experiencia al nivel temático, porque se comprueba que no hay ajuste entre lo percibido y el mundo, como señala S., pero se trataría de un cambio de punto de vista, pues ahí ya pasaríamos a una consideración temática o teórica, en la que comprobamos que nuestra experiencia perceptiva no se ajusta con el mundo. Es por esto que no se ve con claridad desde qué punto de vista el exponente causal puede ser parte de la experiencia. Es más bien la familiaridad con el mundo y no la causalidad, la que explica esta relación directa con las cosas percibidas; para la familiaridad es irrelevante lo causal. Es pertinente notar que S. se refiere a este aspecto de la familiaridad (cfr. *The Construction of Social Reality*, p. 133), especialmente cuando habla del trasfondo, pero no lo incluye en esta explicación de la percepción.

3. Crítica a la ciencia cognitiva

Una característica de las teorías de la ciencia cognitiva es que asimilan, de diferentes maneras, el conocimiento humano con los sistemas de procesamiento de la información; en este sentido conciben la mente como un sistema representacional formal y consideran la psicología como el estudio de los diversos procesos computacionales según los cuales se construyen, organizan, interpretan y transforman las representaciones mentales¹⁴.

Dentro de la psicología o ciencia cognitiva cabe distinguir dos grandes ramas: la psicología computacional clásica y la psicología computacional conexionista.

¹⁴Cfr. P. MARTÍNEZ-FREIRE, *La nueva filosofía de la mente*, Gedisa, Barcelona 1995, cap. X.

Un representante de la primera postura es el psicólogo Zenon Pylyshyn¹⁵. Señala este autor que la posibilidad de imitar la vida mediante un artefacto ha intrigado a la gente a lo largo de la historia, pero sólo en la segunda mitad del siglo XX se ha considerado seriamente la posibilidad de usar este especial tipo de artefacto llamado computador como un medio de entender los fenómenos mentales, ya sea en cuanto procesos internos o públicamente observables. La extrema plasticidad de la conducta de los computadores es una razón por la que estos artefactos han sido considerados como capaces de exhibir inteligencia. Pero el fracaso en distinguir entre la computación como un tipo de proceso y la forma física particular que adopta en las máquinas de computación actuales ha impedido a mucha gente tomar la computación como una explicación literal de los procesos mentales. Si comprendemos la computación, concluye Pylyshyn, en un nivel completamente general (como de hecho se comprende en teoría de la computación) podemos ver que la idea de que el procesamiento mental es computacional es una hipótesis empírica seria y no una metáfora.

S. también dialoga con J. Fodor¹⁶, representante de la primera postura, que considera que ésta es la única psicología científica que puede llegar a realizarse. Para Fodor el mejor modelo explicativo del conocimiento es el computacional y es coherente con esta afirmación en el desarrollo de su obra *El lenguaje del Pensamiento*, donde expone la teoría de que así es como se realiza el proceso o el sistema de representaciones internas, propio de la computación y de la mente humana. Los procesos computacionales para este autor también presuponen un sistema representacional. Aunque el texto es largo, vale la pena exponer qué entiende Fodor por representación y cómo la incluye al interior del pensamiento: «Por eso, tener una actitud proposicional es estar en cierta relación con una representación interna. En especial, tener una actitud proposicional es estar en cierta relación *computacional* con una representación interna (...). Más exactamente: los estados mentales son relaciones entre organismos y representaciones internas, y los estados mentales causalmente interrelacionados se suceden entre sí según unos principios computacionales que se aplican formalmente *a las representaciones*. Este es el sentido de que las representaciones internas constituyen el dominio de los procesos de datos que informan la vida mental. En resumen, un as-

¹⁵Searle discute con este autor a lo largo del capítulo IX de *The Rediscovery of the Mind*. Se refiere particularmente a la obra de Pylyshyn, *Computation and Cognition: Toward a Foundation for Cognitive Science*, MIT Press, Cambridge (MA) 1984. Este autor es un representante de la ciencia cognitiva que iguala lo que se puede decir sobre el cerebro a lo que se conoce o puede llegar a conocer sobre la computación.

¹⁶Es también en *TRM* (cfr. cap. II y IX) donde se encuentran más alusiones a este autor y principalmente una crítica a lo que S. considera una reducción de los procesos semánticos a los exclusivamente sintácticos. Hace referencias a la teoría de Fodor de la naturalización del contenido intencional. Estas teorías están expuestas fundamentalmente en *The Language of Thought*, Thomas Y. CROWELL, New York 1975 y *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge University Press, Cambridge 1986.

pecto esencial de las teorías cognitivas es que tratan de interpretar las transformaciones físicas (causales) en cuanto transformaciones de la información, teniendo como resultado el de mostrar la racionalidad de los procesos mentales»¹⁷.

La psicología cognitiva conexionista, por su parte, utiliza las redes neuronales o modelos conexionistas (o con un término más reciente, sistemas neuromórficos) como modelos de computación contruidos expresamente para reflejar los principios de organización y funcionamiento del cerebro humano. El modelo conexionista trata de reflejar el funcionamiento de las neuronas, que no siempre son digitales o discretas, sino que también tienen valores continuos en sus *outputs*, aunque a veces puedan ser analizadas como si fueran digitales. El grupo de investigación más importante en esta línea es PDP (*Parallel Distributed Processing*), con investigadores de la Universidad de San Diego (California), Stanford y Carnegie -Mellon¹⁸.

John L. Tienson, en su artículo *An Introduction to Connectionism*¹⁹, expone las principales características de ambas teorías cognitivas, lo que se puede llamar la IA tradicional (IAT) y el conexionismo. Resumidamente se puede decir que la IAT sostiene que la cognición es lo que hace el computador clásico: manipulación de símbolos gobernados por reglas. El objetivo es escribir programas que resulten en un comportamiento inteligente, y finalmente escribir programas como aquellos que subyacen a nuestro comportamiento inteligente. Se asume que tales programas pueden ser descubiertos. De acuerdo con esto, el punto de vista tradicional de la mente se denomina a menudo el punto de vista de «reglas y representaciones», caracterizado por 1) Representaciones estructuradas sintácticamente y 2) Reglas formales precisas que refieren a la estructura sintáctica de esas representaciones. Esta teoría ha decepcionado las expectativas que el mundo científico había puesto en ellas, ya que no ha sido posible igualar computacionalmente operaciones cognitivas humanas, aunque en otras el computador aventaje al hombre. Esta decepción ha llevado a pensar que los cerebros humanos no actúan de la manera convencional en que los computadores lo hacen, o que nuestra inteligencia es de una clase diferente. Los conexionistas han intentado, entonces, buscar una aproximación a la cognición que sea más análoga a la manera en que el cerebro trabaja. El sistema conexionista trabaja con redes neuronales, que consisten en una red de procesadores simples similares a neuronas (*neuron-like processors*), llamadas *nodos* o *unidades*. Los nodos están interconectados recibiendo y enviándose mutuamente señales. Las conexiones entre los nodos son análogas a las conexiones de los cables eléctricos o sinápticos²⁰.

S. vincula los trabajos provenientes de la ciencia cognitiva con aquellos de la llamada Inteligencia Artificial (IA), puesto que la primera aúna los resultados de

¹⁷ *El lenguaje del pensamiento*, p. 214.

¹⁸ Cfr. P. MARTÍNEZ-FREIRE, o.c., cap. X.

¹⁹ En «*The Southern Journal of Philosophy*», 126 (supl.), 1987, pp. 1-16.

²⁰ Fodor y Pylyshyn realizan un análisis crítico del conexionismo en *Connectionist and Cognitive Architecture: A Critical Analysis*, «*Cognition*», 28 (1988), pp. 2-71.

la IA con los propios y contempla el computador como la representación correcta de la mente y no sólo como una metáfora. Es habitual distinguir entre la IA débil y la que S. llama fuerte. De acuerdo con la primera, el principal valor del computador en el estudio de la mente es servir como un instrumento poderoso que nos permite formular y comprobar hipótesis de una manera más precisa y rigurosa. En cambio, la IAF considera que un computador programado de forma apropiada es realmente la mente, en el sentido de que de los computadores, dado programas correctos, se puede decir que entienden y tienen estados cognitivos. En esta forma de entender la IA, los programas no son instrumentos o herramientas para comprobar explicaciones psicológicas, sino que más bien ellos mismos son la explicación. S. no está en contra de la IA débil, lo que directamente ataca es la IA fuerte, como una postura importante en la ciencia cognitiva²¹. Según nuestro autor, el cognitivismo iguala el pensar al procesar información, pero considerando este procesamiento exclusivamente como una manipulación de signos: «La mejor manera de estudiar el pensar (...) es estudiar los programas computacionales de manipulación de símbolos, ya estén en los computadores, o en los cerebros. Según este punto de vista la tarea de la ciencia cognitiva es, entonces, caracterizar el cerebro no al nivel de las células nerviosas ni al nivel de los estados mentales conscientes, sino más bien al nivel de su funcionamiento como un sistema de procesamiento de la información»²².

S. no ve grandes diferencias entre los planteamientos de la IAT y la conexionista, porque ambas eliminan todo elemento esencialmente biológico en lo que respecta a la mente humana, pues en ambos casos se piensa que la mente es al cerebro lo que el programa es al *hardware* del computador, sólo que se concibe de una manera diferente este *hardware*: «Según este punto de vista, cualquier sistema físico que tuviese el programa correcto con los *inputs* y *outputs* correctos tendría una mente en exactamente el mismo sentido que tú y yo tenemos mentes. (...) Y el punto no es que, dado todo lo que sabemos, podría tener pensamientos y sensaciones, sino más bien que tiene que tener pensamientos y sensaciones, puesto que todo aquello en lo que consiste tener pensamientos y sensaciones es esto: llevar a cabo el programa correcto»²³. El punto esencial de discrepancia de S. con toda esta postura cognitivista y de la IA fuerte es que en el computador digital sus operaciones se especifican de una manera completamente formal, en término de símbolos abstractos que no tienen ningún significado.

El problema que ha llevado a la investigación acerca de la mente a propiciar los modelos cognitivistas es el afán por superar lo que S. llama el hiato (*gap*) en-

²¹ Cfr. J. SEARLE, *Minds, Brains, and Programs*, en M.A. Boden (ed.), *The Philosophy of Artificial Intelligence*, Oxford University Press, 1990, pp. 67-88. Originalmente publicado en «The Behavioral and Brain Sciences», III (1980): 417-424, Cambridge University Press.

²² J. SEARLE, *Mentes, cerebros y ciencia*, Cátedra, Madrid 1994, p. 50. El título original es *Minds, Brains and Science*. Este volumen recoge *The 1984 Reith Lectures*, hechas para la BBC y publicadas en BBC Publications, Harvard Univ. Press, Harvard 1984.

²³ J. SEARLE, *Mentes, cerebros y ciencia*, pp. 33-34.

tre la explicación de sentido común que damos a las cosas que ocurren en la mente y el nivel de explicación neurofisiológico. Esta laguna realmente existe, porque no tenemos explicaciones de este orden con respecto a la conducta humana o intenciones de la vida diaria, y nos parece que tendríamos que tenerlas, pues una psicología del sentido común -o de la abuela, como también la llaman- no puede tener una sustentación científica. Muchas teorías, desde el conductismo, han tratado de salvar este hiato, la más importante en estos momentos es la ciencia cognitiva. Una explicación del afán por superar esta desvinculación es la concepción objetivista que tenemos de lo mental, por la que queremos reducir o entender lo mental bajo los paradigmas de las ciencias verificables. S. piensa que este hiato realmente no se da, porque lo mental no se explica correctamente en términos neurofisiológicos, aunque efectivamente tenga una conexión con procesos neurofisiológicos. Intentaré explicar esta afirmación desde diferentes argumentaciones que expone S. a través de sus obras.

4. Intencionalidad intrínseca y relativa al observador

Un instrumento útil para comprender esta idea es la distinción, que ya se ha mencionado en el n. 2.1, entre la *intencionalidad intrínseca* de los estados mentales y lo que llama *adscripciones de intencionalidad relativas al observador*. En este último caso es relevante la adscripción que él considera meramente metafórica, según la cual podemos decir, por ejemplo, que el termostato tiene intencionalidad o creencias, porque indica las condiciones de temperatura, ya que este es un caso asimilable a la comparación entre mente y computador. Para S. el computador tiene sólo una intencionalidad relativa al observador y metafórica. En los casos de intencionalidad relativa al observador, no hay propiamente estados mentales ni tampoco significado, en sentido propio, de los términos que utilizamos, sino que se trata solamente de una manera de hablar acerca de la intencionalidad de los observadores. Ya se ha señalado que no hay dos tipos de estados mentales intencionales, sino sólo uno, el que tiene intencionalidad intrínseca, pues las adscripciones de intencionalidad relativas a un observador son siempre dependientes de la intencionalidad intrínseca del sujeto. Se puede plantear la objeción -siguiendo la teoría de McCarthy²⁴, que sostiene que los termostatos tienen intencionalidad-, que pudiéramos adscribir intencionalidad a los objetos, porque podemos considerar el cerebro como un objeto, como una cosa, solamente que con procesos más complejos que los de otros aparatos artificiales. S. respondería que ni aun en ese caso el cerebro tendría una intencionalidad intrínseca, porque los procesos artificiales son del todo diferentes de los procesos biológi-

²⁴Cfr. J. MCCARTHY, *Adscribing Mental Qualities to Machines*, in Ringle, M. (ed.), *Philosophical Perspectives in Artificial Intelligence*, Humanities Press, Atlantic Highlands (NJ) 1979, pp. 161-195.

cos. El cerebro se explica por la línea de la vida y eso nunca puede ser igualado a los procesos formales de las cosas artificiales. Ese es el error en que también cae el funcionalismo, que trata la intencionalidad desde un punto de vista de adscripción del observador y no como una realidad que surge de la vida del sujeto de los estados mentales²⁵.

Si igualamos el cerebro a una máquina o artefacto artificial, nunca podremos responder a la cuestión semántica de cómo obtienen las oraciones que están en la cabeza sus significados, sino que más bien se plantea la pregunta ¿cómo funciona el cerebro al procesar esas oraciones? Y se responde típicamente que el cerebro funciona como un computador digital realizando operaciones computacionales sobre la estructura sintáctica de las oraciones que están en la cabeza. Por mucho que se intente establecer una noción más estricta de la computación, el problema se mantiene irresoluto, porque el punto realmente importante para S. es que la sintaxis es esencialmente una noción relativa al observador. Una demostración de lo dicho es la realizabilidad múltiple de los procesos computacionales, que les posibilita ser llevados a cabo en diferentes medios físicos, pues ahí se ve no sólo que son abstractos o formales, sino también que no son en absoluto intrínsecos al sistema. Queda patente que dependen de una interpretación desde fuera²⁶.

5. La habitación china

Otra explicación es la del famoso experimento mental de la habitación china, en la que S. muestra que la sintaxis no sustituye lo propiamente mental que es la semántica. Resumidamente el experimento es el siguiente. Se ha diseñado un programa que capacita a un computador para simular que entiende chino. Si al computador se le hace una pregunta en chino, confrontará la pregunta con su memoria o base de datos y producirá respuestas adecuadas a las preguntas en chino con tal corrección que iguala a un hablante nativo del chino. Para responder a la pregunta de si efectivamente el computador entiende chino, S. imagina que en lugar del computador la operación de traducción la realiza una persona a la cual se la encierra en una habitación llena de diversas cestas de símbolos chinos. La persona no sabe nada de chino, pero se le da un libro en castellano con las instrucciones para manipular esos símbolos. Las reglas recibidas sólo especifican la manipulación de símbolos de una manera puramente formal, en términos de su sintaxis, no de la semántica. El experimento se hace más complejo al suponer que, sin que la persona lo sepa, se le hacen llegar preguntas en chino y el sujeto, siguiendo las reglas que tiene, las responde perfectamente, de tal manera que pa-

²⁵Cfr. J. SEARLE, *Intrinsic Intentionality*, Reponse/Searle: *Minds, brains, and programs*, *Behavior and Brain Science*, Vol. III (1980), pp. 450-456.

²⁶Cfr. *TRM*, pp. 201 y 209.

rece un hablante nativo de chino. Sobre la base de esta situación, tal como se ha descrito, la persona que baraja los símbolos chinos no entiende absolutamente nada de chino, ni es posible que con este sistema aprenda el idioma, porque siempre está moviéndose en un nivel formal²⁷. Concluye que la sintaxis formal del programa no garantiza por sí misma la presencia de contenidos mentales.

En esta misma obra, S. hace hincapié en que es esencial para nuestra concepción de computador digital que sus operaciones puedan especificarse de manera completamente formal. Esto significa que nosotros estipulamos los pasos de la operación del computador en términos de símbolos abstractos. Los símbolos que explican el comportamiento de la máquina, en este caso del computador, no tienen ningún significado, no tienen ningún contenido semántico, no se refieren a nada: «Pero este rasgo de los programas, el que estén definidos de manera puramente formal o sintáctica, es fatal para el punto de vista de que los procesos mentales y los procesos de programas son idénticos. Y la razón puede enunciarse de manera completamente simple. Tener una mente es algo más que tener procesos formales o sintácticos. Nuestros estados mentales internos tienen, por definición, ciertos tipos de contenido. (...) Esto es, incluso si mis pensamientos se me presentan en cadenas de símbolos tiene que haber algo más que las cadenas abstractas, puesto que las cadenas por sí mismas no pueden tener significado alguno. Si mis pensamientos han de ser *sobre* algo, entonces las cadenas tienen que tener un significado que hace que sean los pensamientos sobre esas cosas. En una palabra, la mente tiene más que una sintaxis, tiene una semántica. (...) Las mentes son semánticas, en el sentido de que tienen algo más que una estructura formal: tienen un contenido»²⁸.

Es un grave error de la ciencia cognitiva suponer que los cerebros procesan información en el mismo sentido que lo hacen los computadores. Un agente externo codifica alguna información de forma que pueda ser procesada por los circuitos del computador; se proporciona una realización sintáctica de la información que el computador puede implementar en diferentes niveles de voltaje. En el proceso, toda la interpretación corre por cuenta del observador. El producto final, el *output*, es un fenómeno físico que un observador puede interpretar como símbolos con una sintaxis y una semántica. En contraste, en el cerebro ninguno de los procesos neurofisiológicos relevantes son relativos al observador, aunque puedan ser descritos como vistos desde el observador: «La realidad biológica no es que el sistema visual produzca todo un ramillete de palabras o símbolos; más bien se trata de un evento visual concreto específico y consciente -esta misma experiencia visual». Aunque podamos hacer un modelo de procesamiento de esta información que da lugar a la percepción, la percepción misma no es un sistema de procesamiento de información, ni tampoco lo es el cerebro²⁹. Este es un caso claro del contraste entre la perspectiva de la primera y tercera persona.

²⁷J. SEARLE, *Mentes, cerebros y ciencia*, pp. 37-39.

²⁸*Ibidem*, p. 37.

²⁹J. SEARLE, *TRM*, pp. 224-225.

Este experimento mental de la habitación china pone de relieve nuevamente que lo mental es específicamente semántico, porque acentúa el divorcio entre la sintaxis y la semántica, ya que en este ejemplo el nivel semántico prácticamente desaparece. No podrían darse en este lenguaje metáforas, analogías, etc., sino sólo un decir algo con un significado que exclusivamente puede captar el oyente que es quien maneja la semántica. Es decir un nivel mínimo de lenguaje. S. no hace estas observaciones, que harían aún más potente su propia argumentación.

6. La sintaxis no es intrínseca a la física y el olvido de la conciencia

En *TRM* S. plantea una nueva manera de enfocar el problema de la IA, que completa el argumento de la habitación china, en el que el hincapié se puso en demostrar que la sintaxis y la semántica son dos aspectos diferentes en lo que podríamos llamar la comprensión o los estados mentales. Y que aun cuando se pueda operar sintácticamente, no hay garantía de que realmente se comprenda un símbolo, una sentencia, etc., porque falta el elemento semántico. También se insiste en que el elemento exclusivamente formal, propio de la computación, no permite acceder a la intencionalidad propia de los estados mentales. Lo que ahora S. quiere mostrar es que el problema es más profundo, porque incluso el elemento sintáctico solo no basta para explicar un proceso computacional, como si éste fuese intrínseco y así igualarlo a un proceso o estado mental. No se puede hablar de una sintaxis intrínseca, sino que ésta es esencialmente una noción relativa a un observador. Todo proceso computacional depende de una interpretación desde fuera y como tal no puede descubrirse desde la física, sino que son procesos que se *asignan* a la física. El *quid* de este argumento es que la sintaxis no es intrínseca a la física (en cambio el de la habitación china era que la semántica no es intrínseca a la sintaxis). Con esto quiere enfatizar S. que no hay nada que haga que un computador sea intrínsecamente tal, sino que siempre su afirmación como computador será relativa a un observador que asigna una interpretación sintáctica a los rasgos puramente mecánicos o artificiales del sistema. Una vez más S. insiste en esta separación entre el nivel de lo físico, al que pertenecen los estados mentales, y el de lo artificial al que pertenece la computación. Es muy significativa la afirmación con la que cierra este argumento: «*El objetivo de la ciencia natural es describir y caracterizar rasgos que son intrínsecos al mundo natural. Por sus propias definiciones de computación y cognición, no hay modo en que la ciencia cognitiva computacional pueda ser jamás una ciencia natural, puesto que la computación no es un rasgo intrínseco del mundo. Se asigna de manera relativa a observadores*»³⁰.

Pero también introduce otra rectificación importante, que lleva a ahondar más en el asunto y no ver sólo el error de igualar la mente a un computador imple-

³⁰ *Ibidem*, p. 212. La cursiva es del autor.

mentado con un programa en el *hardware* del cerebro. Reconoce que hemos sido negligentes respecto a la centralidad de la conciencia en el estudio de la mente. Desde Descartes, pensamos que la conciencia no es un sujeto apropiado para una ciencia seria o científica de la mente y la dejamos de lado³¹. A pesar de estas observaciones, se podría objetar a S. un cierto fisicalismo en su pensamiento, porque no termina de reconocer la especificidad de lo mental o de lo consciente, sino que sobrepone estas características mentales y conscientes a las físicas, que están como a la base de un modelo de estructuras. Esto es coherente con su biologicismo naturalista, pero no deja de ser problemático al contraponerlo al esfuerzo que realiza por concebir lo mental de un modo diferente a como lo hace la ciencia cognitiva y dentro de ella la IAF.

7. La hipótesis del lenguaje del pensamiento

Tampoco es posible para S. aceptar tesis que sustenten que en la mente hay algo más que operaciones sintácticas similares a las de un computador digital, partiendo del supuesto que los estados mentales sean *al menos* estados computacionales y que los procesos mentales operen sobre la estructura formal de esos estados mentales. Esta sería la hipótesis que aceptaría oraciones que ya están en la cabeza, al estilo del lenguaje del pensamiento de Fodor. También aquí se plantea, en primer lugar, la dificultad de la ausencia de la actividad semántica: ¿cómo obtienen estas oraciones que están en la cabeza sus significados? Sólo se estaría planteando: ¿cómo funciona el cerebro al procesar esas oraciones? Y la respuesta sería que el cerebro funciona como un computador digital realizando operaciones computacionales sobre la estructura sintáctica de oraciones que están en la cabeza. En segundo lugar, surge la cuestión de la incoherencia de tal tesis, pues no hay manera en que pueda descubrirse que hay, intrínsecamente, oraciones desconocidas que están en la cabeza, puesto que algo es una oración sólo en tanto algún agente o usuario la usa como oración³².

Se ha visto que sólo principios formales incorporados a un computador no son suficientes para explicar el entender, aunque el hombre fuera capaz de seguir los principios formales sin entender nada. Tampoco hay razón alguna para suponer que tales principios son necesarios o contribuyen de alguna manera para la comprensión del hombre, pues somos capaces de entender, por ejemplo el propio idioma, sin ningún programa formal³³.

³¹J. SEARLE, *Consciousness, explanatory inversion, and cognitive science*, *Behavioral and Brain Sciences*, 1990, 13, p. 585. El tema de la conciencia no podrá ser tratado en este artículo por su extensión, pero al menos menciono la importancia que Searle actualmente asigna a este aspecto. Ciertamente hay filósofos posteriores a Descartes que han dado gran importancia a la conciencia, por ejemplo, Husserl.

³²Cfr. J. SEARLE, *TRM*, pp. 201-202 y pp. 210-211.

³³Cfr. J. SEARLE, *Minds, Brains, and Programs*, p. 71.

Al concluir el capítulo en que ha desarrollado el experimento mental de la habitación china, en *Mentes, cerebros y ciencia*, S. cierra la discusión con esta afirmación: «El resultado de esta discusión es recordarnos algo que ya sabíamos desde el principio, a saber: que los estados mentales son fenómenos biológicos. La conciencia, la intencionalidad y la causación mental son todas ellas parte de la historia de nuestra vida biológica, junto con el crecimiento, la reproducción, la secreción de la bilis y la digestión»³⁴. Me detendré en este reforzamiento de la idea de que la mayor distancia que se puede establecer entre la IA y los estados mentales es que estos últimos son esencialmente biológicos o naturales, mientras que los fenómenos computacionales nunca podrán traspasar la frontera de la vida, que incluye, aunque sea en ciernes, el tema de la conciencia, pues como veremos, para S. la conciencia es también un fenómeno biológico.

8. El caso del robot

Como se verá en el caso del robot, la biología ocupa un lugar central en el pensamiento de S. Oponer lo biológico a lo formal es el argumento básico de sus discusiones. Cuando S. dice que los procesos mentales son fenómenos biológicos reales en el mundo, como la digestión, etc., no se trata de explicarlos neurofisiológicamente, porque nadie sabe en detalle actualmente cómo funciona el cerebro humano, sino que el objetivo es mostrar cómo es posible que los estados mentales puedan ser fenómenos biológicos del cerebro. Esto lo explica nuestro autor diciendo que al igual que fenómenos como la visión, la audición, etc., los pensamientos y el resto de nuestra vida mental están causados por fenómenos que suceden en el cerebro. Los fenómenos electroquímicos tienen como resultado los fenómenos mentales, del mismo modo que la digestión es el resultado de procesos químicos que suceden en el estómago³⁵.

Pero analizando contraargumentos que le han formulado, estudia el hipotético caso del robot que tiene un computador que imita perfectamente las sinapsis del cerebro humano; la conducta completa del robot es indistinguible de la conducta humana; además su mecanismo no está basado en *inputs* y *outputs*, sino que en él está todo unificado. Seguramente, afirma el que plantea la hipótesis, en tal caso tendríamos que adscribir intencionalidad al sistema. S. está de acuerdo en que en este hipotético caso sería racional e incluso irresistible adjudicar intencionalidad al robot, si no conociéramos nada más acerca de su funcionamiento³⁶.

³⁴J. SEARLE, *Mentes, cerebros y ciencia*, p. 48.

³⁵Cfr. J. SEARLE, *Minds and Brains without Programs*, p. 220. En *Mindwaves*, p. 209-233. Este artículo, publicado en forma separada para este volumen, se elaboró basándose en las conferencias Reith del año 1984. S. repite la mayor parte de las ideas de esa conferencia, pero también explicita varios de los puntos ahí expuestos.

³⁶Cfr. J. SEARLE, *Minds, Brains, and Programs*, pp. 78-79, donde se expone y analiza este argumento.

Pero en realidad, esto sería una apreciación superficial, porque analizando bien el argumento se conoce que aboga por la IAF. Para esta postura, no es relevante la conducta para adjudicar intencionalidad a un computador. Es una condición suficiente que el programa formal conecte adecuadamente los *inputs* y *outputs*. La esencia de lo mental, para ellos, es la operación de un sistema físico simbólico. Este caso parece diferente, porque el acento se ha puesto en la conducta similar o casi indiferenciable de la humana, lo que podría hacer pensar que no se está recurriendo a la estructura formal-simbólica del computador que tiene el robot, sino más bien a algo interno capaz de producir estos estados mentales. Llegado a este punto, S. razona de una manera bastante semejante al argumento de la habitación china. Supongamos que aquello que llamamos «interno» no es más que un hombre que opera los símbolos que el robot a través de sus mecanismos receptores capta. El hombre, conforme a unas reglas, interpreta y expresa estos símbolos que permiten actuar al robot con una conducta como la nuestra. Estaríamos ante un robot que es, en realidad, un ingenioso maniquí. Pero la hipótesis de que el maniquí tiene intencionalidad, nos sería ahora increíble e innecesaria. Al único que adscribimos intencionalidad es al hombre que está operando a través del robot. Se puede concluir que la conducta de un robot, aparentemente igual o muy semejante a la humana, podría recibir una explicación en la que no fuera necesario recurrir a la adscripción de intencionalidad.

En cambio, en el caso de especies animales como las de los simios o animales domésticos como el perro, no podríamos comprender su conducta sin la adscripción de intencionalidad. Por una razón principal: estas bestias están conformadas biológicamente como nosotros y esto causa que su estructura o su conformación explique causalmente su conducta. Hay una relación entre lo que se podría llamar el material biológico causal de la conducta animal y el humano y por eso ambos tienen en común la intencionalidad.

La gran diferencia está entre una explicación simbólico-formal de la conducta y una explicación biológica. Nunca se podrán homologar estas dos explicaciones causales y la argumentación va siempre por el mismo lado: la explicación simbólico-formal requiere la intervención de una intencionalidad humana, o, al menos, biológica, y por eso siempre es posible reemplazar lo que se atribuye a una inteligencia artificial, por una inteligencia humana y, más aún, no sólo reemplazar, sino demostrar que lo que se atribuía a la inteligencia artificial es imposible sin la intencionalidad humana.

9. Las reglas y los procedimientos formales

Otro camino para refutar a la ciencia cognitiva es la distinción que S. establece entre seguir una regla y actuar de acuerdo a ciertos procedimientos formales, teniendo en cuenta que el seguir reglas es uno de los elementos explicativos de la IAT. En realidad, seguir una regla no es lo que la IAT efectivamente realiza, pues

el significado de la regla ha de jugar algún papel causal en la conducta del que actúa conforme a esa regla, y lo que en realidad en la IAT se entiende por regla es el actuar de acuerdo a procedimientos formales carentes de significado. Los computadores actúan como si siguieran reglas, pero esta es una forma metafórica de hablar que no puede convertirse en un sentido literal³⁷. Y no pueden seguir reglas, porque no tienen contenido intencional intrínseco al sistema que está funcionando para producir la conducta. El cognitivismo nos dice que el cerebro funciona como un computador comercial y que esto causa la cognición, pero en realidad, de acuerdo a esta explicación, tanto el cerebro como el computador tendrían sólo modelos y los modelos no tienen poderes causales adicionales a los que tienen los medios que los implementan. Sólo puede tener poder causal algo con existencia real y no un objeto, como el computador, que sólo tiene una intencionalidad relativa al observador. Por esta razón, el cognitivismo no podría dar ninguna explicación causal del conocimiento³⁸.

S. ahonda más en esta argumentación, preguntándose si acaso no se puede pensar en una causalidad del computador, porque después de una determinada operación, como pulsar una tecla, éste nos entrega un resultado, en la medida que la máquina está implementando un determinado programa. Pero, ¿es éste realmente un proceso causal, teniendo en cuenta que no hay ningún proceso físico? Es como si dijéramos que el termostato causa la temperatura, porque la indica o regula. Lo que hay, responde S., es que se simula una causalidad, pero en realidad no hay tal. Más aún, la simulación no nos dice realmente lo que ocurre, no es el fenómeno mismo. La simulación de la cognición por ciertos procesos computacionales no nos dice lo que es la cognición ni tampoco qué la causa, porque en cuanto dejamos de lado al hombre que implementa los programas, éstos no nos dicen nada por sí mismos. S. fundamenta este argumento comparando lo que sucede cuando dos sistemas, uno humano y otro mecánico, implementan un algoritmo: «El computador humano recorre conscientemente los pasos del algoritmo, de modo que el proceso es a la vez causal y lógico: lógico porque el algoritmo proporciona un conjunto de reglas para derivar los símbolos de *output* a partir de los símbolos de *input*, y causal porque el agente está haciendo un esfuerzo consciente para recorrer todos los pasos. En el caso del computador mecánico, el funcionamiento total del sistema incluye un homúnculo externo, y con el homúnculo el sistema es a la vez causal y lógico: lógico porque el homúnculo da una interpretación a los procesos de la máquina, y causal porque el *hardware* de la máquina causa que recorra todo el proceso. Pero estas condiciones no pueden ser cumplidas por las operaciones neurofisiológicas brutas, ciegas, no conscientes del cerebro. En el computador cerebral no hay implementación intencional consciente alguna del algoritmo como la hay en el computador humano, pero no puede haber tampoco implementación no consciente alguna como la que hay en el

³⁷Cfr. *Mentes, cerebros y ciencia*, p. 55.

³⁸Cfr. *TRM*, p. 216.

computador mecánico, puesto que esto exige un homúnculo externo que añada una interpretación computacional a los eventos físicos»³⁹.

En definitiva, lo que S. quiere demostrar es que no podemos atribuir propiamente cognición a la IA, sino sólo adscripciones de reglas que son siempre relativas a un observador, es decir, el homúnculo que asigna interpretaciones computacionales a los estados de *hardware*. «Sin el homúnculo no hay computación, sólo hay un circuito electrónico»⁴⁰. La IA no consigue dar cuenta de lo que sucede en la mente humana donde sólo hay un mecanismo físico, que es el cerebro, que tiene diversos niveles de descripción causal: físico- reales y físico-mentales.

Siguiendo esta misma línea argumental S. establece una clara distinción entre procesar información y pensar o tener procesos mentales. Los computadores y las calculadoras procesan información, pero desconocen absolutamente esta información. Sucede *como si* ocurriese algún proceso mental de información, pero en realidad, cuando ocurre un proceso mental, además de procesar la información, hay pensamiento, porque se conoce la información; ésta tiene un significado para el sujeto. El cognitivismo, según S., confunde estas dos situaciones y olvida el *como si*⁴¹. S. vislumbra un «algo más» propio de la mente considerada en tanto que diferente a los procesos cerebrales neurofisiológicos, pero no lo distingue de una actividad netamente biológica, sino solamente le otorga un rango de descripción superior que será el físico-mental. Es por esta razón que insiste tanto en la diferencia que hay entre la explicación formal que da la ciencia cognitiva y la explicación biológica. La primera se maneja en el nivel de una descripción en la que en definitiva se consigue obtener un conjunto de símbolos como *output* en respuesta a un conjunto de símbolos como *input*. Pero esta explicación es diferente a la que podemos encontrar en lo que no es un modelo, sino la realidad física misma, como por ejemplo en una experiencia visual que es un evento consciente concreto que se produce en el cerebro por procesos biológicos electroquímicos específicos⁴². Más adelante, en el mismo texto, aclara que el cerebro con sus operaciones intrínsecas no hace un procesamiento de la información, sino que el órgano biológico específico y sus procesos neurobiológicos también específicos, causan formas específicas de intencionalidad, entre las cuales está, en algunos casos, la conciencia⁴³.

10. El dualismo de la Inteligencia Artificial

En el artículo *Minds, Brains, and Programs* hace una última crítica a la IA y es que ésta necesariamente cae en el dualismo, porque afirma que donde la men-

³⁹Cfr. *TRM*, pp. 219-220.

⁴⁰*Ibidem*, p. 221.

⁴¹Cfr. *Mentes, cerebros y ciencia*, p. 57.

⁴²Cfr. *TRM*, pp. 224-225.

⁴³*Ibidem*, p. 226.

te tiene un papel, el cerebro no tiene importancia. Es decir, se vuelve como en el cartesianismo a separar mente y cerebro, ya que en la IAF y en el funcionalismo lo importante son los programas y éstos son independientes de su realización en máquinas; así pues en la IA el mismo programa podría ser realizado por una máquina electrónica, una sustancia mental cartesiana, o un espíritu del mundo hegeliano. Al exponer estas ideas a los que trabajan en IA, S. ha comprobado que no han comprendido cómo los fenómenos mentales deben ser dependientes de las propiedades físico-químicas de los actuales cerebros humanos. El proyecto de la IA cae en un nuevo cartesianismo que es tratar de reproducir y explicar lo mental diseñando programas, pero para esto hay que suponer que la mente no es sólo conceptual, sino empíricamente independiente del cerebro, puesto que el programa es completamente independiente de cualquier realización. Este sería un dualismo en su forma más fuerte⁴⁴. En *TRM* hace notar que se supera el dualismo si no se separa la existencia de la conciencia de una característica física, porque los estados mentales conscientes son sólo un rasgo biológico ordinario, a un nivel más alto. De tal manera que la conciencia no puede sino ser considerada como una propiedad emergente del cerebro, como la liquidez es una propiedad emergente de las moléculas de H₂O. Según esto, la conciencia es una propiedad mental y, por lo tanto, física del cerebro. El que una característica sea mental no significa que no sea física y, al revés, el que sea física no es obstáculo para que sea mental; y éste es el error dualista en el que incurre la IAF al desvincular lo mental de lo físico, pues lo mental se iguala a lo formalmente computarizado⁴⁵.

11. Algunas observaciones críticas

En la mayor parte del artículo he expuesto el pensamiento de S. respecto a los temas seleccionados, de una manera más bien descriptiva, intentando ser fiel al autor. Ahora expondré algunas críticas que se refieren principalmente a la carencia de una explicación adecuada de lo que S. llama rasgos biológicos superiores.

11.1. El biologicismo de Searle

S. claramente ha defendido que la conciencia y a la intencionalidad son rasgos biológicos superiores. Aunque, como advertí, no introduciré explícitamente el tema de la conciencia por su extensión, los textos a los que voy a aludir la incluyen; los consideraré bajo este aspecto puntual que es el que interesa.

Conviene partir desde más atrás y ver, resumidamente, en qué consiste el biologicismo de S. Ya en la introducción a *TRM* afirma: «Lo que he argumentado

⁴⁴Cfr. *art. cit.* p. 86.

⁴⁵Cfr. *TRM*, pp. 14-15.

respecto de ambas posturas (Searle, *Minds, Brains, and Science: The 1984 Reith Lectures*, Harvard University Press., Cambridge, Ma.), y repito aquí es que uno puede aceptar los hechos obvios de la física -que el mundo consta enteramente de partículas físicas en campos de fuerza- sin negar que entre los rasgos físicos del mundo hay fenómenos biológicos tales como estados de conciencia cualitativamente internos e intencionalidad intrínseca»⁴⁶.

Varias son las cuestiones que surgen a partir de este texto. En primer lugar, cómo conecta S. los rasgos físicos del mundo, que son los que lo componen enteramente, con estos fenómenos biológicos que claramente no pueden ser estudiados, ni considerados desde la misma perspectiva que aquellas realidades compuestas de partículas físicas. Es evidente que la intencionalidad y la conciencia no pueden ser vistas desde la perspectiva que se señaló, porque son realidades que ni siquiera se advierten en la experiencia desde una perspectiva exclusivamente física, tal como se pueden llegar a conocer los rasgos físicos del mundo que menciona. Antes, o junto a esta cuestión, está el tema de si efectivamente podemos decir que el mundo está totalmente compuesto por partículas físicas, pues es necesario aclarar, ¿de qué mundo hablamos? Se trata del mundo entendido exclusivamente como una realidad física, entendiendo por ella material. Pero ¿qué sucede si ampliamos el concepto de mundo a «realidad», a «todo aquello que existe»? Me parece que si no se hace esta aclaración, se caería en un reduccionismo materialista, pues hay realidades, específicamente las mentales, de las cuales no podemos dar cuenta desde una perspectiva exclusivamente física. Este punto lo trataré más en el n. 3, en el que me refiero al enfoque objetivista y al de la primera persona.

Una segunda cuestión es que S. siempre ha insistido en el carácter intrínseco e interno de la conciencia y la intencionalidad, tal como se ha expuesto a lo largo del artículo. Sin embargo, resulta difícil ver cómo compaginar estas características con su énfasis en unificar conciencia e intencionalidad con un fenómeno físico más en el mundo natural⁴⁷. Se pueden dar explicaciones a este punto, para comprender mejor desde dónde S. hace estas afirmaciones. Una de ellas es que cuando se refiere a las diferencias entre rasgos intrínsecos del mundo y rasgos relativos al observador, afirma que los rasgos intrínsecos son aquellos que se pueden incluir en la ciencia natural, como por ejemplo la masa de un cuerpo cualquiera, y los relativos al observador son aquellos de los que no cabe ciencia natural, como es el caso de que ese objeto cualquiera sea una bañera.

Es decir, la atribución social que hacemos de las cosas, particularmente de las artificiales, no son rasgos intrínsecos o internos de las cosas y, por lo tanto, tampoco podemos llamarlos estrictamente físicos. En este sentido, la conciencia y la intencionalidad son características físicas, porque, desde el punto de vista de Se-

⁴⁶J. SEARLE, *TRM*, XII. Las posturas con las que discute son el materialismo y el dualismo de propiedades.

⁴⁷Cfr. *ibidem*, p. 90.

arle, se pueden explicar como fenómenos del mundo natural. Pero la cuestión aquí es qué realidades estamos contraponiendo. Efectivamente las características de la conciencia son diferentes a las de la bañera, pero no sólo porque la última es tal por una práctica social y por un artificio del hombre y, por lo tanto, sus rasgos son extrínsecos a la cosa misma, pues también la bañera con el paso del tiempo o en alguna ocasión podría usarse, por ejemplo, como jardinera. Resulta evidente en esta contraposición que uno es un rasgo intrínseco, el otro extrínseco. Pero la cuestión no lo es tanto si se contraponen dos rasgos intrínsecos del hombre o de los seres conscientes, como son los rasgos biológicos y los mentales. En este ámbito, S. no establece contraposición, porque los rasgos mentales, desde su punto de vista, son rasgos biológicos sólo que de un nivel superior.

11.2. La emergencia de los estados mentales a partir de propiedades biológicas

¿Podemos considerar, por otra parte, igualmente intrínsecos los estados mentales y los rasgos o características biológicas?

S. es claro, al finalizar su libro *TRM*, en afirmar cuál es su visión general de la mente y de su relación con el mundo natural: «Veo el cerebro humano como un órgano como cualquier otro, como un sistema biológico. Su rasgo especial, por lo que respecta a la mente, el rasgo en el que difiere notablemente de otros órganos biológicos, es su capacidad de producir y mantener toda la enorme variedad de nuestra vida consciente»⁴⁸. Con anterioridad en esta misma obra, como también en otros artículos, ha explicado cómo es posible que un órgano biológico produzca y mantenga la vida de la conciencia. Hay que tener en cuenta que al buscar una explicación, S. está reconociendo, de una manera explícita, una diferencia entre la actividad biológica de otros órganos que no requieren más que una explicación física, y en este sentido material, y la actividad del cerebro que produce conciencia y que exige una explicación de orden filosófico, porque de algún modo ve esta diferencia de niveles.

La explicación está más detallada en el artículo *Minds and Brains Without Programs*. Por eso seguiré esta, más que la que expone en *TRM*. S. propone un nuevo modelo de causación, habitual en la física, y que se puede aplicar a esta situación tan peculiar del cerebro y los estados mentales. Se trata de la distinción entre micro y macro-propiedades de los sistemas. Estos están compuestos de micro-partículas que tienen características en el nivel de las moléculas y de los átomos, así como en el de las partículas subatómicas. Pero cada sistema tiene también ciertas propiedades como la solidez en el caso de la mesa, o la fluidez en el caso del arroyo, que son macro-propiedades o propiedades de superficie (*surface properties*) de los sistemas físicos. Algunas macro-propiedades, no todas, pueden ser causalmente explicadas en base al comportamiento de los ele-

⁴⁸J. SEARLE, *TRM*, p. 227.

mentos en el micro-nivel. Por ejemplo, la solidez de la mesa frente a mí es explicada causalmente por la estructura reticular de la que se compone la mesa. De manera similar, la fluidez del agua es explicada causalmente por el comportamiento de los movimientos de las moléculas de H₂O. Pero no todas las macropropiedades tienen una explicación causal en términos de micro-comportamiento. Por ejemplo, la velocidad del arroyo no se explica en base al movimiento de las moléculas, sino más bien por el ángulo de la pendiente, la atracción de la gravedad y la fricción provista por el lecho. Sin embargo, en el caso de las macrocaracterísticas que son explicadas causalmente en base al comportamiento de los elementos en el micro-nivel, le parece a S. encontrar un modelo perfectamente normal para explicar las intrigantes relaciones entre la mente y el cerebro: «Mi modo preferido de enunciar este punto es decir que las características de superficie F *son causadas por* el comportamiento de los micro-elementos M, y que al mismo tiempo está *realizado en* el sistema de los micro-elementos. Las relaciones entre F y M son causales, pero al mismo tiempo F es simplemente una característica de nivel más elevado del sistema que consiste en los elementos M»⁴⁹.

Una objeción que se puede plantear a esta explicación es que F es idéntico a las características de M. Por ejemplo, científicamente se puede definir la solidez como cualquier micro-estructura que cause estos toscos rasgos observables, como el que yo pueda poner libros sobre la mesa. Se puede decir que la solidez es sólo la estructura reticular del sistema de moléculas y que así definida, la solidez causa, por ejemplo, la resistencia al tacto y a la presión; o se puede decir que la solidez consiste en cosas tales como la rigidez y la resistencia al tacto y a la presión, y que está causada por el comportamiento de los elementos en el micro-nivel. Este paso de la causación a la identidad definicional es muy común en la historia de la ciencia y a S. o no le causa problemas, no considera que sea una objeción al análisis que plantea.

Si se aplican estas explicaciones al estudio de la mente, no habría para S. dificultad en dar cuenta de las relaciones metafísicas de la mente con el cerebro en términos de una teoría causal del funcionamiento del cerebro como productor de estados mentales: «Así como la fluidez del agua está causada por el comportamiento en el micro-nivel y es al mismo tiempo una característica realizada en el sistema de micro-elementos, exactamente en el mismo sentido de ‘causado por’ y ‘realizado en’ los fenómenos mentales son causados por procesos que suceden en el cerebro en el nivel neuronal o modular, pero están realizados en el mismo sistema que consiste en neuronas organizadas en módulos. Y así como necesitamos la distinción micro-macro para cualquier sistema físico, por las mismas razones necesitamos la distinción micro-macro para el cerebro». Después de explicar que podemos aplicar al cerebro el mismo esquema causal que nos es familiar respecto del resto de la naturaleza, afirma: «Las características intrínsecamente

⁴⁹J. SEARLE, *TRM*, p. 226.

mentales del universo son las características *físicas* de alto nivel de los cerebros»⁵⁰.

Varias cuestiones se pueden plantear al respecto. La primera y quizás crucial dentro del esquema de análisis en el que se mueve S. es tratar de responder a la pregunta de qué es una propiedad física de los cuerpos, pues las propiedades físicas no son estrictamente intrínsecas, si las consideramos desde una perspectiva de distinción entre la primera y la tercera persona. La física, considerada como una perspectiva de análisis de hechos del mundo, atribuye a los cuerpos ciertas características o propiedades a las cuales da una denominación, por ejemplo, masa y peso. Pero se las atribuye por los efectos que son observables en otros cuerpos. La noción de causación que maneja la física es ésta, y así hablamos, por ejemplo, del calor, del color, etc. Es decir, el estado de la materia se define por relación a la acción que nosotros podemos ejercer sobre un cuerpo o que ese cuerpo ejerce con respecto a nuestros sentidos. Como dice S., reconocemos la solidez de la mesa, porque sobre ella puedo realizar una acción que no la deforma, por la cual no puedo penetrar en ella, etc. Se trata de nociones relativas al observador o al agente, pero ¿hasta qué punto podemos hablar de nociones que expresen una propiedad intrínseca?⁵¹ Porque según los ejemplos anteriores, la mesa es sólida con respecto a alguien. Es verdad que esta solidez es la resultante de la asociación y fuerza de la asociación de las moléculas que, siguiendo el ejemplo, componen la mesa, pero esto no podría ser conocido, sino por los efectos que causan en los seres sentientes, por lo cual las propiedades físicas son intrínsecas sólo en cierto sentido: en cuanto inferimos algo del ser propio de ese objeto. La física moderna y experimental conoce este aspecto intrínseco de un modo mediato, porque inmediatamente el aspecto intrínseco sólo es conocido de una manera sensible. Dicho de otro modo: si la física corta todo nexo con el mundo de la experiencia tal como éste se *da* en ella, se queda sin *explanandum*⁵².

De ahí que sea difícil compaginar en S. lo que se podría llamar una visión internalista de la ciencia natural, con lo que efectivamente esta ciencia puede dar de sí por sus mismos métodos y características propias. La crítica a la ciencia cognitiva de pretender hacer de la computación una ciencia natural, también se puede aplicar a la biología, pues ésta por sus propios métodos se realizará siem-

⁵⁰ *Minds and Brain Without Programs*, pp. 223-224.

⁵¹ *Ibidem*, pp. 224-225.

⁵² «En general, nos representamos las características de las entidades naturales como si correspondiesen de algún modo a nuestras características: las *tendencias* de los animales, la *actividad* de las substancias químicas, el *trabajo* realizado por una fuerza, la *duración* de los procesos físicos, son conceptualizaciones en las cuales tomamos como punto de referencia las experiencias de nuestra propia existencia. Sin duda, al construir los conceptos científicos se intenta evitar las connotaciones antropomórficas, pero cuando se construye una imagen de la naturaleza, resulta forzoso recurrir a representaciones que se refieren a nuestra propia experiencia» (M. ARTIGAS, *La inteligibilidad de la naturaleza*, Eunsa, Pamplona 1992, p. 452).

pre desde la perspectiva de la tercera persona. Este es el problema que se plantea en el intento de Searle de fisicalizar lo mental.

11.3. La perspectiva de la primera y tercera persona

Por esto, se puede plantear la pregunta de si realmente los estados mentales pueden considerarse como propiedades físicas. La cuestión aquí es que si los consideramos como propiedades físicas tenemos que reconocer en ellos las características de las propiedades físicas, que son en su mayoría inferencias de causalidades sensibles. ¿Qué decir del fenómeno del conocimiento, de la volición, etc.? Además, en los estados mentales nos estamos moviendo exclusivamente en el orden de la primera persona, y ya hemos visto que las propiedades físicas exigen una consideración objetiva, desde la perspectiva de la tercera persona. Incluso la experiencia sensible es del orden de la primera persona, y para convertirse en lo que llamamos un conocimiento de una «propiedad física» tenemos que hacer una inferencia que la pone en el orden objetivo. Los fenómenos conscientes son reales, pero no reúnen todas las características o propiedades de lo que llamamos realidades físicas, y desde luego, se sitúan en el orden de la primera persona⁵³. ¿Cómo entonces afirmar que son producto del cerebro como propiedad física? S. podría responder diciendo que se ha reconocido que son propiedades de un nivel superior. Pero esa superioridad ¿en qué estriba? Son superiores, podría responder, porque son del todo diferentes. Pero ¿por qué son diferentes? se le puede replicar, o ¿no habría que precisar en qué radica la diferencia?

Por otro lado, ¿cómo argumentar que los estados de la mente son diferentes a las propiedades de la materia? Se puede argumentar que una propiedad física no es diferente a la materia, o no es algo nuevo con respecto a ella. Las propiedades físicas no son más que una mayor o menor deformabilidad de los componentes y de las relaciones o lazos que se establecen entre ellos. En ese sentido, S. también acepta que F sea igual que M cuando se le objeta la similitud, pero argumenta que aun siendo iguales, se puede distinguir entre las micro y macro-propiedades y establecer una relación entre *causado por* y *realizado en* el cerebro. Pero el punto de discusión sigue vigente, porque se puede objetar que los estados mentales no sólo son propiedades de mayor nivel, sino de diferente categoría o que tienen unas características que las sitúan en un estado de realidad de otro orden que el de las propiedades físicas. Hablando en el mismo lenguaje de S., se diría que lo más radicalmente diferente de los estados mentales es su relación al sujeto. En este sentido su fuerte carácter de relacionalidad los distancia claramente de los fenómenos físicos. Los fenómenos mentales no son reductibles a la observación. Y cuando reduzco a la observación fenómenos sensibles, también subjetivos en

⁵³En este, como en otros temas del artículo, agradezco las sugerencias del Prof. Dr. Alejandro Vigo P. de la Universidad de los Andes.

el sentido de fenómenos de primera persona, es necesario que efectúe una mediación objetiva, que de algún modo sitúe en el nivel de la tercera persona el modo como experimento la estructura molecular de las cosas. Si ya en este nivel del conocimiento sensible se encuentran diferencias, surgirán mayores cuando se trate de analizar fenómenos más propiamente relacionados con la conciencia, como son el conocimiento, los deseos, etc., porque con respecto a ellos no tengo un acceso en absoluto semejante al que tengo con respecto a los fenómenos físicos. ¿Cómo explicar esta diferencia con la sola consideración del fenómeno consciente como un fenómeno físico de nivel superior?

Además, el propio S. en su crítica a la ciencia cognitiva afirma que lo que explica actos como la cognición intelectual o sensible no es la adscripción externa, sino la explicación de la ocurrencia de fenómenos concretos como son la comprensión consciente de una oración o la experiencia visual de una escena (cfr. n. 5). Otra dificultad que surge en la explicación de S. es que naturalmente tenemos la intuición de que cuando conocemos lo hacemos como un todo; no es una parte de nuestra persona la que conoce, no es nuestro cerebro el que conoce. ¿Cómo hacer compatible esta experiencia humana con la afirmación de que nuestra actividad mental es el producto de una parte de nuestra persona como es el cerebro?⁵⁴

Es difícil ver cómo S. respondería a estos cuestionamientos, cuando es tan rotunda su afirmación de que la conciencia y la intencionalidad pertenecen al mundo biológico y natural y que no hemos alcanzado un conocimiento correcto de la conciencia justamente por la exclusión de la mente y de la conciencia de este mundo biológico natural: «En pocas palabras, la conciencia es un rasgo biológico de los cerebros humanos y de ciertos animales. Está causada por procesos neurobiológicos y es una parte del orden biológico natural como cualquier otro rasgo biológico, como lo son la fotosíntesis, la digestión o la mitosis»⁵⁵. También ilustra su pensamiento esta otra afirmación: «la conciencia encuentra su lugar naturalmente como un rasgo fenotípico de ciertos tipos de organismos con sistemas nerviosos altamente desarrollados»⁵⁶. Excluir la conciencia del mundo natural, según S. fue un gran error; fue un recurso heurístico útil en el siglo XVII porque permitió a los filósofos concentrarse en fenómenos que fueran medibles, objetivos, desprovistos de significado, es decir, libres de intencionalidad. Pero esta exclusión estaba basada en la falsedad de pensar que la conciencia no es parte del mundo natural. Es por esto que, según S., nos hemos encontrado con la enorme dificultad de estudiar la conciencia con las herramientas científicas a nuestro alcance.

⁵⁴Estas reflexiones y cuestiones surgieron de una conversación con el Prof. de la Universidad de los Andes, Alejandro Serani M., neurólogo y Dr. en Filosofía. Cfr. A. SERANI MERLO, *El viviente humano. Estudios Biofilosóficos y Antropológicos*, Eunsa, Pamplona 2000. Para este tema ver especialmente los cap. I y II.

⁵⁵J. SEARLE, *TRM*, p. 90.

⁵⁶*Ibidem*.

La gran pregunta que surge de estas afirmaciones es: ¿De qué manera podríamos estudiar la conciencia y la intencionalidad, fenómenos de primera persona, con herramientas científicas diseñadas para estudios objetivos, hechos desde la tercera persona? Hay puntos en los que S. no guarda una perfecta coherencia entre la intención de hacer un análisis desde la primera persona y el resultado efectivo, porque sin advertirlo cae en la perspectiva de la tercera persona que distorsiona desde su misma raíz el correcto enfoque y la posible solución al problema.

A pesar de estas objeciones, el esfuerzo de S. por rescatar lo específicamente mental es muy interesante y digno de ser contrastado con posturas divergentes. Su argumentación contra la IAF es sólida y valida una clara distinción entre la inteligencia humana y la artificial. Se entiende la preeminencia tanto de la semántica con respecto a la sintaxis, como también la de la conciencia e intencionalidad con respecto a modelos puramente formales. En este sentido, su internalismo es una respuesta rotunda a una consideración exclusivamente mecanicista o reduccionista de los estados mentales. S. recoge también, aunque sin lograr explicarlo de una manera adecuada, la unidad del viviente y particularmente de la persona humana, que no tiene como algo extrínseco su dimensión física, sino que constituye en ella un aspecto básico. La dificultad que a mi parecer no logra superar S. es el dar cuenta de la especificidad del ser humano, porque parte de premisas, como es la de su biologicismo naturalista, que cierran la posibilidad al estatuto ontológico propio de la persona, como son las capacidades racionales y volitivas que no son posibles de explicar satisfactoriamente desde la exclusiva biología.

* * *

Abstract: *Per comprendere adeguatamente l'internalismo sostenuto da Searle occorre riflettere su ciò che lui chiama intenzionalità estrinseca e confrontare quest'ultima con la scienza cognitiva. Nella sua teoria intenzionalista la "rappresentazione" e la "autoreferenzialità" sono nozioni-chiave. Questi concetti sono esposti soprattutto in Intentionality, anche se vengono già usati in Speech Acts. Il presente articolo è suddiviso in due parti: dapprima si espongono le caratteristiche dell'internalismo in Searle e poi viene raffrontata la sua posizione con la scienza cognitiva, particolarmente in riferimento a The Rediscovery of Mind. Infine, viene messo in risalto che nelle opere di Searle il suo biologicismo naturalista è incompatibile sia con la comprensione degli stati mentali sia con quella che è chiamata la prospettiva della prima persona.*