

*Propiedades nucleares de los fenómenos mentales según Searle: intencionalidad, subjetividad, semánticidad**

Asunción ÁLVAREZ RODRÍGUEZ

Resumen

Este ensayo pretende ser una caracterización de la concepción searleana de los fenómenos mentales. Para Searle, la naturaleza mental de los procesos cerebrales va determinada por su capacidad de hacerse conscientes. Así, se explicitan los que Searle considera que son los rasgos principales de la consciencia – subjetividad e intencionalidad –, así como su interacción y su inscripción en el ámbito más amplio de la semánticidad del pensamiento humano.

Palabras clave: Searle, fenómenos mentales, consciencia, subjetividad, intencionalidad, semánticidad, Habitación China, inteligencia artificial.

Abstract

This essay aims to be a characterization of the Searlean conception of mental phenomena. For Searle, the mental nature of brain processes is determined by their capacity to become conscious. Thus, what Searle takes to be the main features of consciousness – subjectivity and intentionality – are expounded, as well as their interaction and their inscription within the larger framework of the semanticity of human thought.

Keywords: Searle, mental phenomena, conscience, subjectivity, intentionality, semanticity, Chinese Room, artificial intelligence.

* Quiero dar las gracias a Antonio Benítez por su ayuda y comentarios en la redacción de este trabajo, así como a los evaluadores anónimos de la *Revista de Filosofía* por sus sugerencias.

Hablar de la teoría de la mente de John Searle necesariamente requiere abordar la cuestión de la consciencia. En efecto, para Searle el carácter mental de los fenómenos va determinado por su carácter consciente, o al menos potencialmente consciente. En la teoría searleana, una caracterización de la consciencia equivale pues *de facto* a una caracterización de los estados mentales en general.

Este trabajo pretende ser una caracterización tal. Partiendo de los rasgos que Searle considera característicos de la consciencia (y, por tanto, de los fenómenos mentales), se examinarán con particular detalle los dos rasgos que mayores implicaciones tienen para la filosofía de la mente: intencionalidad y subjetividad. Ambos rasgos, argüimos, han de considerarse dentro del contexto de la semántica general que Searle les atribuye a los procesos mentales humanos, y que constituye la base de su famoso argumento de la Habitación China contra la posibilidad de un pensamiento artificial.

1. Intencionalidad

La palabra *intencionalidad* es un término técnico tomado de la filosofía escolástica medieval (*intentio*, traducción del árabe *ma'na*), donde se usaba para referirse a cosas en la mente u operaciones mentales. La *intentio* estaba relacionada con la *inexistentia*, que no quiere decir “inexistencia” sino “existencia en”, existencia del objeto en la mente o tal como se presenta a la mente, tal como ésta lo conoce y experimenta.

El término in-existencia fue retomado en el siglo XIX por Franz Brentano para designar la propiedad que distinguiría a los fenómenos psíquicos de los fenómenos físicos. Brentano arguyó que todo estado mental – tanto perceptual (oír un sonido, tener calor) como no perceptual (recordar algo, emitir un juicio) – incluye un objeto que se presenta o aparece al sujeto. El hecho de que los actos mentales incluyan tales presentaciones constituye su intencionalidad (o, en términos de Brentano, su in-existencia intencional), y sería lo que los distinguiría de los fenómenos físicos.

Searle retoma la noción brentaniana de intencionalidad como la propiedad de algunos estados mentales de estar referidos a objetos distintos de sí mismos, pero con una salvedad. Brentano consideraba la intencionalidad como la propiedad definitoria de los fenómenos psíquicos frente a los físicos: para él, todo lo mental es intencional, y todo lo intencional es mental. Para Searle, por el contrario, todo lo intencional es (en último término) mental, pero no todo lo mental es intencional.

1.1. Los actos de habla

Los orígenes del pensamiento de Searle acerca de la intencionalidad pueden rastrearse hasta la época de su trabajo sobre los actos de habla, en la huella de su maestro J.L. Austin. En su libro *Cómo hacer cosas con palabras*, Austin denunció lo que denominó la *falacia descriptiva*, la creencia tradicional de que la única función del lenguaje es la de describir estados de cosas (y por tanto de que todos los enunciados son evaluables como verdaderos o falsos). Austin señaló que tal teoría no contempla los frecuentes casos en que el lenguaje sirve precisamente para *hacer* cosas, los casos en que la acción consiste en *decir* (e.g. en promesas, amenazas, declaraciones de matrimonio). Así, el lenguaje, además de su tradicional uso *constativo*, tiene también un uso *performativo*, no evaluable en términos de verdad o falsedad.

Pero Austin dio un paso aún más radical al no limitarse a añadir el empleo performativo, realizativo, como un uso secundario respecto del uso constativo, descriptivo, primario del lenguaje. Por el contrario, Austin planteó la idea de que todos los usos del lenguaje – incluidos los descriptivos o proposicionales – serían performativos, dado que el enunciar una proposición (verdadera o falsa) constituye tanto un “hacer algo” como el pronunciar un juramento.

Esta idea habría de tener consecuencias cruciales para la teoría de la mente de Searle. En efecto, como veremos a continuación, Searle aplica el esquema austiniano de los actos de habla a los estados mentales intencionales. Del mismo modo en que Austin considera que todo uso del lenguaje es en último término performativo, esto es, supone un “hacer cosas con palabras”, así también para Searle lo mental supone también siempre un “hacer”. Examinaremos las implicaciones de esta noción en mayor detalle más adelante. Por de pronto, veamos cómo aplicó Searle la teoría de los actos de habla de Searle al ámbito de lo mental.

La labor de Searle en el campo de los actos de habla fue principalmente taxonómica. Siguiendo las ideas de Austin, Searle distingue entre:

- *acto locucionario* (de *locutio*, el decir): el acto mismo *de* decir algo;
- *acto ilocucionario* (de *in locutio*, en el decir): el acto realizado *al* decir algo;
- *acto perlocucionario* (de *per locutio*, por el decir): el acto realizado *por* haber dicho algo.

De estos tres tipos de acto que tienen lugar en el lenguaje, es el acto ilocucionario el más relevante para los fines que nos ocupan aquí. Searle distingue dos elementos dentro el acto ilocucionario: el *acto proposicional* y la *fuerza ilocucionaria*. El acto proposicional consta a su vez de un *acto de referencia* y un *acto de predicación*. El acto proposicional nos dice de qué estamos hablando, y constituye el tradicional *contenido proposicional* o contenido del acto de habla. Sin embargo, el acto ilocucionario pleno (el acto de decir algo de hecho) no se completa hasta que no se añade la fuerza ilocucionaria, que determina la naturaleza del acto. Así, en los enunciados

- a) Juan bebe
- b) ¿Juan bebe?
- c) ¡Juan, bebe!

el acto proposicional es el mismo, dado que el referente es Juan y lo que se predica de él es la acción de beber. Pero las distintas fuerzas ilocucionarias hacen que a) sea un enunciado asertórico, b) una pregunta, y c) una orden. Los actos ilocucionarios constan entonces de un elemento psicológico (la fuerza ilocucionaria) y un elemento proposicional (formado por los actos de referencia y predicación).

Como se mencionó antes, Austin desechó la idea de que todos los enunciados sean evaluables en términos de verdad o falsedad, reemplazando la idea de *verdad* de los enunciados por la de *felicidad*. Si la verdad de los enunciados descriptivos supone su correspondencia con el estado de cosas que describen, la noción más amplia de felicidad de los enunciados supone una adecuación a la situación de su emisión. De este modo, Austin describe toda una serie de *infortunios*, fallos o violaciones de las *condiciones* necesarias para que la emisión feliz de un enunciado. En este sentido, se habla de *condiciones de felicidad*, necesarias para que un enunciado llegue a buen puerto, frente a las clásicas *condiciones de verdad* – que pasan a convertirse en la teoría de los actos de habla en las condiciones de felicidad de un tipo particular de enunciado, los enunciados descriptivos.

Así, Searle forja el concepto de *dirección de ajuste* para caracterizar las condiciones de felicidad de los actos de habla. Los actos de habla *representativos* (tales como la aserción descriptiva) tienen una dirección de ajuste *palabras-a-mundo*, en tanto que el enunciado ha de corresponder con la realidad que describe (i.e. ha de ser verdadero) para que su emisión sea feliz. En cambio, los actos de habla *directivos*, tales como las órdenes o las amenazas, tienen una dirección de ajuste *mundo-a-palabras* dado que pretenden alterar

la realidad mediante el uso del lenguaje. También los actos de habla *compromisorios* tienen una dirección de ajuste *mundo-a-palabras*, puesto que el compromiso del hablante expresado por su enunciado ha de cumplirse para que el enunciado se emita felizmente. Los actos de habla *expresivos*, al servir al propósito de expresar el estado psicológico del hablante, carecen de dirección de ajuste. Y finalmente, los actos de habla *declarativos* – los actos de habla por excelencia, tales como nombrar a alguien para un cargo, bautizar, casar – tienen una dirección de ajuste *recíproca*, i.e. *mundo-a-palabras-a-mundo*, dado que este tipo de enunciados pretenden modificar la realidad, pero también han de ajustarse a ella (e.g. no se puede casar a dos chimpancés). Searle subraya que este tipo de actos se dan necesariamente en el marco de instituciones sociales concretas.

Searle aplica los tres rasgos de los actos de habla (distinción fuerza ilocucionaria / contenido proposicional, dirección de ajuste, condiciones de felicidad) a su análisis de los estados mentales intencionales. De este modo, al articular lo que denomina “rasgos lógicos” de la intencionalidad, Searle establece una analogía entre éstos y los que caracterizan a los actos de habla:

- Del mismo modo que los actos de habla están compuestos de fuerza ilocucionaria y contenido proposicional, los estados mentales intencionales tienen tanto un “*modo psicológico*” como un *contenido intencional* (o representativo).. El modo psicológico puede ser el modo de creencia, deseo, esperanza, miedo, etc., mientras que el contenido intencional concierne a lo que se cree, desea, espera, teme, etc. Sin embargo, el contenido de un estado mental (como el contenido proposicional de un acto de habla) no debiera confundirse con el objeto de este pensamiento. El contenido forma parte del estado mental. El estado intencional no es sobre su propio contenido, sino sobre algún objeto.

- Al igual que los actos de habla, los estados mentales intencionales tienen una *dirección de ajuste* (no ya entre palabras y mundo, sino entre mente y mundo), que depende del modo psicológico. Si Susana cree que está lloviendo, la dirección de ajuste es mente-a-mundo; si quiere que su amiga la venga a visitar, es mundo-a-mente.

- Los estados mentales intencionales tienen lo que Searle denomina *condiciones de satisfacción* o éxito, análogas a las condiciones de felicidad de los actos de habla, que determinan si el estado mental se ve “cumplido” (e.g. un deseo se ve cumplido si el estado de cosas que constituye su contenido intencional se crea en la realidad).

Sin embargo, aunque tome los términos para su análisis de la intencionalidad de su anterior trabajo sobre los actos de habla, Searle deja muy claro que no considera que la intencionalidad mental (i.e. la de los estados psicológicos) se derive en modo alguno de la intencionalidad lingüística (i.e. la de los actos de habla), ni de ninguna otra forma de intencionalidad o de estado mental. Es la intencionalidad del lenguaje la que se deriva de la mental: el lenguaje no le impone intencionalidad a la mente, sino a la inversa. La intencionalidad lingüística manifiesta en los actos de habla es, por así decirlo, un subconjunto de la intencionalidad general, que va más allá del lenguaje. De este modo, la estructura del lenguaje, con sus capacidades de referencia y predicación, con su dirección de ajuste y sus condiciones de felicidad, refleja la de la mente. Le conferimos significado a entidades tales como las palabras en virtud de la intencionalidad mental, pero los estados mentales *per se* carecen de significado – i.e. no representan, no remiten a ninguna otra entidad –, y suponen el punto final del análisis, el “lecho de piedra” (*bedrock*). La intencionalidad mental, asevera Searle, es intrínseca, y ha de explicarse en términos biológicos.

1.2. Intencionalidad y consciencia

La concepción searleana de la intencionalidad cobra particular relevancia dentro del estudio de la consciencia, entendida ésta en su sentido fenomenológico-cognitivo, es decir, como el “darse cuenta” propio de los fenómenos mentales. De hecho, la vinculación que establece entre intencionalidad y consciencia puede considerarse la tesis más fuerte de Searle, a la que él denomina el Principio de Conexión, y que enuncia mediante una afirmación radical:

sólo un ser que pueda tener estados intencionales conscientes puede tener estados intencionales, y todo estado intencional inconsciente es, al menos, potencialmente consciente. [Searle 1992b, p. 141]

Es decir, que la consciencia constituye un requisito tanto de la intencionalidad como de la inconsciencia. Como el propio Searle es el primero en señalar, esta tesis tiene serias consecuencias para el estudio de la mente. Por ejemplo, cualquier análisis válido de la intencionalidad ha de explicar necesariamente la consciencia, dado que Searle considera que la primera no puede darse sin la segunda. Asimismo, Searle utiliza su argumento de que cualquier contenido intencional inconsciente debe considerarse como susceptible de

pasar a la consciencia para atacar a sus contrincantes que postulan la existencia de pensamientos y procesos inconscientes que no pueden en principio conectarse con la consciencia. En efecto, Searle considera que buena parte de los intentos por separar consciencia e intencionalidad – eliminando la consciencia de las explicaciones del funcionamiento de la mente – se basan en gran medida en la asignación de un estatuto inconsciente a algunas operaciones mentales básicas. Searle encuentra la versión extrema de este enfoque en la distinción de Ray Jackendoff entre “mente computacional” (inconsciente) y “mente fenomenológica” (consciente) [Searle 1992b, pág. 160]. Searle niega la legitimidad de esta dicotomía, ya que considera que

Por profundas razones, nuestra noción de estado mental inconsciente es parásita de nuestra noción de estado consciente. [Searle 1992b, p. 161]

De este modo, la explicación searleana de la consciencia (y de la intencionalidad) pasa, por un lado, por una definición de la “mentalidad” (i.e. de la cualidad de ser mental); y por otro lado, por una revisión de la noción de lo inconsciente y sus relaciones con lo consciente.

1.2.1. Mental y no-mental

Para Searle, como dijimos, la noción de estado mental conlleva la posibilidad de acceso a la consciencia:

No tenemos noción alguna del inconsciente excepto como aquello que es potencialmente consciente. [Searle 1992b, pág. 160]

Esta concepción aleja a Searle tanto de posiciones cognitivistas como la de Jackendoff (que entiende por inconscientes los procesos “computacionales” que no llegan a la consciencia “fenomenológica”), como de las posiciones psicoanalíticas (para las que una parte de los contenidos mentales, por estar siempre sujeta a la represión primaria, no puede hacerse consciente).

Para explicar su concepción del inconsciente, Searle se opone a “nuestra noción ingenua y preteórica”, por la cual un estado mental inconsciente equivale un estado mental consciente menos la consciencia. Esto, afirma Searle, es un contrasentido: en efecto, él localiza la diferencia entre fenómenos mentales y fenómenos no mentales precisamente en la consciencia. Lo que distingue a un estado mental (como un deseo) de un estado no mental (que sin embargo puede intervenir en la producción del estado mental, como por

ejemplo la mielinización de los axones del sistema nervioso central) es el hecho de que el deseo siempre puede hacerse consciente, mientras que no tenemos por qué ser conscientes del proceso de mielinización para que éste se produzca (y de hecho, nunca podremos ser conscientes de tal proceso). Sin embargo, aunque podemos no ser conscientes de un deseo en un momento dado, tal deseo inconsciente siempre será, según Searle, susceptible de pasar a la consciencia. Mientras que para Brentano era la intencionalidad lo que distinguía a los fenómenos psíquicos de los físicos, para Searle es la consciencia – inherente a los estados intencionales, pero no exclusiva de ellos – lo que los diferencia. De este modo, podría esbozarse una escala gradual de lo mental, en la que el *macronivel* estaría constituido por la consciencia, y el Trasfondo sería el *micronivel* donde lo mental se engazaría con lo no-mental (en un *continuum*, según Searle, dentro del ámbito de lo natural).

1.2.2. *Consciente, inconsciente, no-consciente*

Como hemos visto, Searle considera que existen fenómenos que nunca podrán llegar a hacerse conscientes (incluso aunque sirvan para sostener biológicamente el fenómeno de la consciencia, como en el caso de la mielinización de los axones). Para tales estados no mentales, Searle reserva el término “no consciente”, frente a “inconsciente”.

En efecto, para Searle hablar de estados inconscientes presupone de entrada una referencia a la consciencia: un estado mental es inconsciente si puede llegar a hacerse consciente; de no ser así, se trata de un estado no consciente, y no es mental en absoluto.

Es importante, sin embargo, recordar que, según Searle, la consciencia que diferencia a los estados mentales de los estados no-mentales es explicable en términos físicos. No hay ninguna diferencia metafísica entre mente y no-mente: sencillamente, será mental todo estado neurofisiológico que tenga la particularidad de ser consciente o que tenga la capacidad especial de producir un estado consciente. En efecto, para Searle, como hemos visto, la consciencia no es, en último extremo, más que un fenómeno fisiológico entre otros, fruto de la evolución de la especie, y como tal explicable en términos naturalistas.

1.3. Aspectualidad

La aspectualidad de un estado mental intencional consiste en el hecho de que dicho estado, a la hora de tomar un objeto, necesariamente selecciona un número limitado de rasgos – *aspectos* – del objeto en cuestión: no es posible aprehender un objeto en todos sus aspectos simultáneamente (asumiendo que todos sus aspectos pudieran ser identificados).

Aunque Searle no la considera *per se* una característica de la consciencia, la aspectualidad aparece como un rasgo estrechamente vinculado a la intencionalidad por vía de la consciencia. Toda percepción consciente, afirma Searle, y por tanto toda intencionalidad, es aspectual.

Mis experiencias conscientes, a diferencia de los objetos de las experiencias, tienen siempre una perspectiva. Son siempre experiencias conscientes desde un punto de vista [...]

Obsérvese que el carácter de la experiencia consciente consistente en tener perspectiva es una buena manera de recordarnos que toda intencionalidad tiene un aspecto. Ver un objeto desde, por ejemplo, un punto de vista es verlo bajo ciertos aspectos y no bajo otros. En este sentido, todo ver es “ver como”. Y lo que vale para ver vale para todas las formas de intencionalidad, conscientes e inconscientes. Todas las representaciones representan sus objetos, u otras condiciones de satisfacción, bajo aspectos. Todo estado intencional tiene lo que yo llamo un contorno de aspecto. [Searle 1992b, pág. 140]

La aspectualidad puede ponerse en relación con la distinción entre perspectivas de primera y de tercera persona que Searle, como veremos más adelante, emplea para su explicación de la causalidad (cf. 2.2). Searle dice que una explicación causal tiene una perspectiva de tercera persona cuando refiere exclusivamente a entidades ajenas a la mente, mientras que las explicaciones causales que tienen una perspectiva de primera persona refieren a la mente que profiere tales explicaciones. Sin embargo, la perspectiva de tercera persona resulta engañosa en la medida en que, al abordar los objetos como separados de la mente que los aprehende, no tiene en cuenta el proceso mismo de aprehensión, ni la mente que aprehende, y sin la cual no hay tal perspectiva en absoluto. Todo proceso mental intencional es, según Searle, en cierto modo un proceso de primera persona (lo cual no equivale a negar el status objetivo de los fenómenos: Searle distingue claramente entre fenómenos dependientes e independientes del observador). En este sentido, la aspectualidad de los estados mentales intencionales – es decir, su capacidad de seleccionar rasgos de un objeto – se funda sobre la perspectiva de primera persona.

2. Subjetividad

Searle emplea el término “subjetividad” para referirse a la perspectiva de primera persona que caracteriza a la intencionalidad, y que como tal constituye también una característica irreductible de la consciencia y de los fenómenos mentales. Para Searle, la subjetividad es un rasgo de los fenómenos mentales del que es imposible prescindir en una descripción adecuada de la realidad: según él, es un *hecho* de la naturaleza que la consciencia es subjetiva.

2.1. Ontológico vs. epistémico, subjetivo vs. objetivo

Esta certeza sobre la existencia de la subjetividad lleva a Searle a considerarla una categoría ontológica (i.e. relativa a la realidad misma), no un modo epistémico (i.e. relativo al conocimiento de esa realidad).

Consideremos, por ejemplo, el enunciado “Tengo dolor de espalda”. Tal enunciado es completamente objetivo en el sentido de que lo convierte en verdadero la existencia de un hecho real y no depende de las actitudes u opiniones de los observadores. [Searle 1992b, pág. 106]

Es decir, que, paradójicamente, la subjetividad es un hecho *objetivo* de la naturaleza. Por el contrario, algunas entidades de la naturaleza, como el dolor, tienen un modo *subjetivo* de existencia:

el fenómeno mismo, el dolor real mismo, tiene un modo subjetivo de existencia, y es en ese sentido en el que digo que la consciencia es subjetiva. [Searle 1992b, pág. 106]

En *Mente, lenguaje y sociedad*, Searle plantea el problema de la objetividad y la subjetividad de un modo sucinto y bastante elegante, bajo la forma de un silogismo:

Es frecuente afirmar que la subjetividad nos impide dar una explicación científica de la consciencia, que la subjetividad sitúa a la consciencia fuera del alcance de la investigación científica. Pero el argumento se basa en un silogismo erróneo. Si evidenciamos la falacia de este silogismo, creo que podemos llegar a entender mejor la subjetividad. Es así como se desarrolla el argumento:

1. La ciencia es por definición objetiva (en contraposición a subjetiva).
2. La consciencia es por definición subjetiva (en contraposición a objetiva).
3. Por tanto, no puede haber ciencia de la consciencia. [Searle 2001, pp. 48-49]

La falacia, afirma Searle, estriba en la ambigüedad de los términos “subjetivo” y “objetivo”. Searle distingue dos sentidos en que se pueden entender dichos términos: el sentido epistémico y el sentido ontológico. Mientras que el sentido epistémico sólo es aplicable a proposiciones (a las que se puede asignar valores de verdad), el sentido ontológico es aplicable a hechos brutos.

Una proposición es epistémicamente subjetiva si su verdad depende esencialmente de las actitudes y sentimientos de los observadores. [Searle 2001, pág. 49]

De este modo, la proposición “Rembrandt nació en 1609” es *epistémicamente objetiva* porque es posible saber si Rembrandt nació en 1609 con independencia de los sentimientos y actitudes que tal afirmación suscite en nosotros. En cambio, la proposición “Rembrandt es mejor pintor que Rubens” es *epistémicamente subjetiva*, dado que su verdad o falsedad depende exclusivamente de las actitudes, sentimientos, o preferencias de los observadores.

Pero “subjetivo” y “objetivo” también tienen un sentido ontológico:

Mientras que el sentido epistémico se aplica a las proposiciones, el sentido ontológico se refiere al status del modo de existencia de los tipos de entidades del mundo. Las montañas y los glaciares tienen un modo ontológico de existencia porque su modo de existencia no depende de que un sujeto los experimente. Pero los dolores, cosquillas y picores, así como los pensamientos y sentimientos, tienen un modo subjetivo de existencia porque sólo pueden existir en la medida en que los experimente algún ser humano o animal. [Searle 2001, pág. 49]

La falacia del argumento contra la posibilidad de una ciencia de la consciencia consiste, según Searle, en confundir el sentido epistémico con el sentido ontológico, en

suponer que como los estados de consciencia tienen un modo de existencia ontológicamente subjetivo no pueden ser el objeto de estudio de una ciencia epistémicamente objetiva. [Searle 2001, pág. 49]

Negar que el dolor de mi pie constituye un hecho susceptible de estudio científico es incurrir en el error de confundir el *hecho* de que me duele el pie con la *proposición* “Me duele el pie”. El que me duela el pie es un hecho ontológicamente subjetivo, dado que mi dolor sólo existe en la medida en que yo lo experimento (frente a un hecho ontológicamente objetivo como el hecho de que existe el Everest, puesto que el Everest existe con independen-

cia de que alguien lo perciba) . En cambio, “Me duele el pie” es una proposición epistémicamente objetiva porque su verdad o falsedad no depende de mis actitudes o sentimientos, sino exclusivamente de si me duele el pie o no (frente a una proposición epistémicamente subjetiva como “Me gusta Rembrandt”, cuya verdad o falsedad sólo depende de mis gustos).

La ciencia es una práctica discursiva que consiste en la asignación de valores de verdad o falsedad a enunciados empíricamente verificables. La ciencia, en tanto que necesariamente maneja proposiciones, trata con enunciados epistémicos, no con hechos ontológicos. Por tanto, cualquier enunciado epistémicamente objetivo (i.e. independiente de las opiniones o sentimientos del observador) será susceptible de estudio científico – y ello incluye a los enunciados acerca de la consciencia.

Searle concede que este carácter epistémicamente objetivo de la subjetividad resulta, en principio, difícil de aceptar para una mentalidad materialista, ya que parece dejar la puerta abierta al temido dualismo. Sin embargo, para Searle cualquier descripción de la realidad *objetiva* que no incluya la subjetividad será una descripción incompleta.

Searle señala la importancia que tiene el concepto de *observación* para la definición de lo objetivo (algo es objetivo si es “accesible de la misma manera a todos los observadores competentes”): encuentra en el proceso de observación que define la objetividad la misma subjetividad que aquélla pretende desterrar. Es decir, si lo objetivo se entiende como lo universalmente observable, la subjetividad queda excluida del cuadro en tanto que se halla en el mismo ojo que observa:

Encontramos difícil explicar satisfactoriamente la subjetividad, no sólo por haber sido educados en una ideología que dice que, en último término, la realidad ha de ser completamente objetiva, sino porque nuestra idea de una realidad objetivamente observable presupone la noción de observación que es, en sí misma, ineliminablemente subjetiva, y que no puede convertirse en el objeto de la observación, como sí pueden serlo los objetos y estados de cosas objetivamente existentes. En pocas palabras, no hay manera en que podamos figurar pictóricamente la subjetividad como parte de nuestra cosmovisión porque, por así decirlo, la subjetividad en cuestión es la actividad de figuración pictórica. [Searle 1992b, pág. 109]

La solución, según Searle, no estriba en ampliar el campo de la cosmovisión para incluir al ojo que ve, sino en abandonar por completo la figuración pictórica o visual como modo de explicar la realidad:

La solución no es la de intentar desarrollar un modo especial de representación pictórica, una suerte de superintrospección, sino más bien la de abandonar en este punto las figuraciones pictóricas y limitarnos a reconocer los hechos. Los hechos son que los procesos biológicos producen fenómenos mentales conscientes, y que éstos son irreductiblemente subjetivos. [Searle 1992b, págs. 109-110]

2.2. Primera persona y tercera persona: la causación intencional

Searle, como ya se ha mencionado, considera que la intencionalidad mental es intrínseca y de carácter biológico, y que de ella se derivan todos los demás tipos de intencionalidad, como la lingüística. Tal distinción entre intencionalidad *intrínseca* y *derivada*, afirma Searle, es subsumible bajo una distinción más amplia, la que se da entre fenómenos *independientes del observador* y fenómenos *dependientes del observador*:

La distinción entre intencionalidad intrínseca y derivada es un caso especial de una distinción mucho más fundamental entre aquellas características del mundo que son independientes del observador, como la fuerza, la masa o la atracción gravitatoria, y aquellas características que son dependientes del observador, como la de ser un cuchillo, una silla o una frase en inglés. La intencionalidad intrínseca es independiente del observador: tengo un estado de hambre con independencia de lo que piense cualquier observador. La intencionalidad derivada es dependiente del observador: por ejemplo, una frase francesa sólo tiene el significado que tiene en relación con los observadores, usuarios, etc. [Searle 2001, pág. 88]

Esto le hace dar una explicación naturalista de la intencionalidad (y por tanto de la consciencia y de los fenómenos mentales), que la define como un fenómeno biológico:

La tesis básica que defiende, tesis que difiere de lo que en el momento actual se considera estándar y ortodoxo en la ciencia cognitiva, puede formularse como sigue. Supongamos que tuviéramos una ciencia física, química y biológica completa. En ese caso, acabaría por establecerse que determinados rasgos del mundo real son reales, independientes del observador, o intrínsecos. La gravitación y el electromagnetismo, por ejemplo, serían dos ejemplos de este tipo de rasgos en el campo de la física. En la biología incluirían, por ejemplo, la mitosis, la meiosis, y la fotosíntesis. Lo que yo sostengo es que también incluirían la consciencia y la intencionalidad. La consciencia y la intencionalidad, aunque sean rasgos de la mente, son independientes del observador en el sentido en que si estoy cons-

ciente o tengo un estado intencional como la sed la existencia de tales rasgos no depende de lo que piense nadie externo a mí. [Searle 2001, pág. 89]

La explicación naturalista que Searle da de la intencionalidad le permite ponerla en relación con la que habitualmente se considera como la principal relación natural entre eventos del mundo: la *causalidad*. Uno de los propósitos de Searle, de hecho, es llevar a cabo una intencionalización de la causalidad, dentro de su proyecto más amplio de naturalización de la intencionalidad.

Esta intencionalización de la causalidad es necesaria, según Searle, para desbaratar la idea, común en la filosofía de la mente y las ciencias cognitivas, de que la intencionalidad ha de poder reducirse a algún otro tipo de relación entre objetos (dado que es imposible que una entidad física sencillamente *se refiera a otra*). Desde el punto de vista de Searle, este planteamiento no es más que una nueva versión del viejo problema mente-cuerpo: la pregunta acerca de la naturaleza de la intencionalidad se formula en unos términos tales que no queda más remedio que elegir entre la opción dualista (por la cual lo mental es irreducible a lo físico) y la materialista (por la cual lo mental es reducible a lo físico).

Además de la oposición entre fenómenos dependientes e independientes del observador, Searle establece una distinción ulterior entre explicaciones de *primera persona* y explicaciones de *tercera persona*. Las tradicionales explicaciones humeanas de la causación, que dan cuenta de los procesos causales de modo objetivo y separado, serían explicaciones de tercera persona. Esta clase de explicación suele adoptar, con fines predictivos, la forma de una ley estadística, e.g. “un estímulo de tipo A provoca regularmente un comportamiento de tipo B”.

Frente a las explicaciones de tercera persona, Searle propone las explicaciones de primera persona, que no recurren para la predicción a ninguna ley implícita, sino directamente a la experiencia. Searle da el siguiente ejemplo de explicación de primera persona:

supongamos que tengo sed y bebo agua. Si alguien me pregunta por qué bebí agua, conozco la respuesta sin ninguna observación adicional. Tenía sed. Además, en este tipo de casos parece que conozco la verdad del contrafáctico sin ninguna observación adicional y sin apelar a leyes generales. Sé que, si no hubiese estado sediento en ese momento, no habría bebido agua. [Searle 1992a, pág. 128, traducción modificada]

En este caso, la sed funciona como causa, pero también como *identificador* del efecto que provoca (i.e. beber agua). La sed provoca que se produzca la satisfacción de la sed. De este modo, Searle entiende la relación de causa-efecto en la intencionalidad como el proceso de hacer que sucedan cosas.

El enfoque de primera persona de la causalidad intencional, según Searle, permite dar cuenta de la interacción entre mente y cuerpo. Los seres humanos estamos biológicamente contruidos para aprehender directamente efectos que provocamos mediante nuestras acciones, y efectos provocados en nosotros por objetos que percibimos. Este enfoque de la causalidad, obviamente, no encaja con la causalidad humeana: Searle rechaza explícitamente la noción de que las causas y efectos se produzcan en virtud de regularidades expresables como leyes, por lo menos en lo que concierne a la intencionalidad. La percepción directa de causa y efecto (entendido éste como el modo en que se hace que las cosas sucedan) obvia la necesidad de descubrir regularidades. (Por supuesto, Searle admite que las explicaciones en forma de leyes regulares son aplicables en todos los demás campos, donde no interviene la intencionalidad propia de la mente humana).

Al postular un tipo particular de causación para los fenómenos intencionales, Searle entra de lleno en el terreno de la filosofía de la ciencia. El enfoque clásico de la relación causa-efecto presupone una hipótesis acerca de la regularidad que permite justificar el proceso de inducción. En otras palabras, hay que tener una creencia (intencional) acerca de la regularidad en el Universo para justificar plenamente el funcionamiento de la ciencia. Pero según Searle, las relaciones causales se establecen mediante procedimientos de prueba y error – lo cual evita la necesidad de tener una creencia acerca de la regularidad de la naturaleza. Pues la presunción acerca de la regularidad, en su opinión, no constituye una creencia, sino que por el contrario forma parte del Trasfondo preintencional (cf. 4.1). Ello le permite retener el carácter no-intencional de la ciencia.

El tipo de causación es así un rasgo que permite distinguir la intencionalidad (intrínseca o derivada) de otros tipos de actividad mental y otros fenómenos de la naturaleza. Mientras que los fenómenos no-intencionales se pueden caracterizar por la causación humeana (que toma forma de ley), la causación propia de los fenómenos intencionales

difiere espectacularmente de la causación propia de la bola de billar o humeana: la causa y el efecto actúan como lo hacen bien porque la causa es una representación del efecto, bien porque el efecto es una representación de la causa. [Searle 2001, pág. 97]

2.3. Tres errores clásicos

Searle estima que su caracterización de lo mental como intencional, y de lo intencional como consciente, subjetivo y aspectual, contribuye a refutar tres grandes errores que, según él, la filosofía de la mente toma generalmente como tesis sobre la consciencia. Estas tres tesis falsas son las siguientes:

1. Todos los estados conscientes son *autoconscientes*.
2. La consciencia se conoce a sí misma mediante la facultad de *introspección*.
3. El conocimiento de nuestros propios estados es *incorregible*: no podemos estar equivocados acerca de ellos.

2.3.1. Autoconsciencia

Searle considera que media un abismo entre “la noción ordinaria y no problemática de autoconsciencia” y el término técnico que suele emplearse en la filosofía de la mente. Searle acepta el sentido vulgar y corriente del término: resulta obvio que hay estados de consciencia en los que uno es consciente de uno mismo. En este sentido corriente, la autoconsciencia es para Searle

una forma extraordinariamente sofisticada de sensibilidad y probablemente es poseída sólo por los seres humanos y algunas otras, muy pocas, especies. [Searle 1992b, pág. 152]

En este mismo sentido, además, Searle considera falso que todo caso de consciencia sea un caso de autoconsciencia: es perfectamente posible “olvidarse de sí” y seguir siendo consciente.

En su sentido técnico, tal como lo define Searle, la autoconsciencia se refiere a estados en los que uno es consciente de *ser consciente*, en una especie de aprehensión metaconsciente. A este respecto, Searle concede que es posible transferir nuestra atención del *objeto* de una experiencia consciente a la *experiencia* en sí, de modo semejante a como se desplaza la atención. Sin embargo, niega que *todo* estado consciente conlleve necesariamente tal focalización de la atención. Aunque es posible dirigir la atención sobre el hecho mismo de ser consciente (Searle pone como ejemplo la pintura impresionista, que llama la atención del espectador sobre la experiencia misma de la percepción), tal concentración específica, al igual que en el sentido no especializado del término, no es un requisito de la consciencia.

2.3.2. Introspección

Como en el caso de la autoconsciencia, Searle distingue entre el sentido habitual de “introspección” y su sentido técnico filosófico.

En su sentido corriente, la introspección aparece como una metáfora (de hecho, el término viene del latín *intra* “dentro” + *spectio* “visión”, i.e. “mirar dentro”). En este sentido, las personas “miran dentro” de sus estados mentales (de sus sentimientos, creencias, deseos...) para examinarlos en mayor detalle, cuestionar su validez, su pertinencia, etc. Se trata de una metáfora inofensiva, pero que, según Searle, lleva a peligrosas conclusiones cuando se toma literalmente:

El problema, según creo, no tiene que ver con el uso ordinario de la noción de introspección, sino con nuestro impulso como filósofos en tomar la metáfora literalmente. La metáfora sugiere que tenemos una capacidad de examinar nuestros propios estados conscientes, una capacidad modelada sobre la visión. Pero ese modelo o analogía es, seguramente, erróneo. En el caso de la visión tenemos una distinción clara entre el objeto visto y la experiencia visual que el preceptor tiene cuando percibe el objeto. Pero no podemos hacer esta distinción para el caso de los propios estados mentales conscientes. [Searle 1992b, pág. 153]

Searle admite el sentido común de la introspección, que califica de “crucial para el autoconocimiento”. Pero si el sentido especializado del término comporta una capacidad análoga a la visión para percibir nuestros propios estados mentales, Searle rechaza su validez:

Pero si por “introspección” queremos decir que tenemos una capacidad análoga a la visión sólo que con menos colores, que tenemos que mirar dentro, entonces me parece que no hay tal capacidad. No la puede haber porque el modelo de mirar dentro exige una distinción entre el objeto mirado y el mirarlo, y no podemos hacer esta distinción en el caso de los estados conscientes. [Searle 1992b, pág. 153]

Dado que el modelo introspectivo requiere una distinción entre objeto de introspección y sujeto espectador, y dado que tal distinción no existe en el caso de los estados mentales (dado que el mismo individuo es a la vez objeto y sujeto), no se puede hablar de introspección ahí.

2.3.3. Incorregibilidad

La tesis de la incorregibilidad de nuestro conocimiento de nuestros propios estados mentales se basa en el argumento de la *autoridad de primera persona*. Como ya vimos respecto de la causalidad (cf. 2.2), se puede distinguir entre una perspectiva de tercera persona, que trata exclusivamente con los objetos en cuestión, y una perspectiva de primera persona, que incluye al sujeto desde el que se constituye tal perspectiva. Según este argumento, uno no puede estar equivocado acerca de sus propios estados mentales dado que los conoce “de primera mano”, por así decir. Pero Searle cuestiona esto:

¿por qué podría pensar alguien que era imposible que uno estuviese equivocado sobre los contenidos de sus propios estados mentales? ¿Por qué, para empezar, habría de suponer alguien que eran “incorregibles”? La respuesta tiene que vez quizás con la confusión entre la ontología subjetiva de lo mental y la certeza epistémica. [Searle 1992b, pág. 154]

Basándose en las oposiciones ontológico / epistémico y subjetivo / objetivo que ya se trataron en 2.1, Searle afirma que los estados mentales son ontológicamente subjetivos. Sin embargo, ello no impide que uno pueda confundirse respecto de sus propios estados mentales.

Todos sabemos a partir de nuestras propias experiencias que sucede a menudo que otra persona está en mejor posición de la que estamos nosotros para determinar si, por ejemplo, estamos realmente celosos, enfadados o con sentimientos de generosidad. Es verdad que el modo en que me relaciono con mis estados mentales y, por lo tanto, el modo en que me relaciono con mis informes de mis estados mentales, es diferente del modo en que otras personas se relacionan con mis estados mentales. Y esto afecta al status de sus informes sobre mis estados mentales. Sin embargo, sus informes pueden ser más exactos que los míos. [Searle 1992b, pág. 154]

Entonces, ¿en qué queda la autoridad de primera persona? Searle, como vimos, le atribuye a los estados mentales intencionales una *perspectiva* de primera persona, frente a la de tercera persona. Sin embargo, autoridad no es perspectiva.

Searle examina el “valiente” intento de Wittgenstein, en sus *Investigaciones filosóficas*, por eliminar absolutamente la idea de las profesiones de primera persona como *informes* o *descripciones* – lo cual acabaría igualmente con la idea de la autoridad de primera persona.

Si el que yo diga “Tengo dolor” pudiera tratarse como una especie de grito ritualizado, una forma convencionalizada de conducta de dolor, entonces no se plantearía cuestión alguna sobre mi autoridad. [Searle 1992b, pág. 155]

Sin embargo, Searle considera que tal reducción conductista del lenguaje no es aceptable – no se puede negar el hecho de que en muchos casos mediante un enunciado se intenta dar una descripción cuidadosa de los propios estados mentales. No siempre se trata sencillamente de dar expresión a un estado mental, sino también de describirlo. Por tanto, si la noción de autoridad de primera persona no puede ser eliminada, ¿en qué consiste exactamente?

Creo que el modo de captar lo que hay de especial en los informes de primera persona es preguntar por qué no pensamos que tenemos la misma autoridad especial sobre los objetos y estados de cosas del mundo que son distintos de nuestros estados mentales. La razón es que en nuestros informes sobre el mundo en general existe una distinción entre cómo las cosas nos parecen y cómo son realmente. [Searle 1992b, pág. 155]

Según esta concepción, mientras que nuestra falible percepción puede llevarnos a confundirnos respecto de la verdadera naturaleza de la realidad “en sí misma”, no es posible que nos confundamos respecto de nuestra percepción:

no hay distinción que pueda hacerse entre apariencia y realidad para cómo las cosas me parecen a mí. [Searle 1992b, pág. 155]

Searle plantea dos objeciones a esta posición de indudable raigambre cartesiana.

En primer lugar, plantea la cuestión de cuál podría ser la *forma* del error acerca de los propios estados mentales, si éste no se basa en la distinción apariencia / realidad que caracteriza a los errores acerca de la realidad “externa”. Searle encuentra tres modos alternativos de errar a este respecto: el *autoengaño*, la *mala interpretación* y la *falta de atención*.

En segundo lugar, Searle cuestiona la tesis de que la distinción apariencia / realidad no sea aplicable a los propios estados mentales. Al fin y al cabo, éstos también forman parte de la realidad. Para Searle, eximir a los estados mentales del sometimiento a la distinción apariencia / realidad a la que responden el resto de fenómenos naturales es una exigencia del paradigma cartesiano. Según este paradigma, la consciencia había de constituir el funda-

mento último del conocimiento y de la ciencia: por tanto, el acceso a la consciencia había de ser infalible y libre de errores. El paradigma cartesiano, aventura Searle, es también responsable de los otros dos errores clásicos mencionados antes:

para que la consciencia nos dé una cierta fundamentación del conocimiento, tenemos que tener primero cierto conocimiento de los estados conscientes; de ahí la doctrina de la incorregibilidad. Para conocer la consciencia con certeza, tenemos que conocerla por medio de alguna facultad especial que nos dé acceso directo a ella: de ahí la doctrina de la introspección. Y – aunque tengo menos confianza en esto como diagnosis histórica – si el yo [consciente] ha de ser la fuente de todo conocimiento y significado, entonces es natural pensar que hay una conexión necesaria entre consciencia y autoconsciencia: de ahí la doctrina de la autoconsciencia. [Searle 1992b, págs. 157-158]

3. Semánticidad

La intencionalidad que Searle percibe en la consciencia y, por extensión, en los fenómenos mentales en general, puede enmarcarse en el contexto más amplio de la semánticidad. Según el Principio de Conexión de Searle, todos los estados intencionales son estados conscientes / mentales – pero ello no significa que todos los estados conscientes / mentales sean estados intencionales. Un ejemplo claro es el dolor: se trata de un estado mental consciente que carece de objeto, y por tanto no es intencional. Donde Brentano percibe la intencionalidad como una particularidad propia de los fenómenos psíquicos frente a los físicos, Searle postula la noción más amplia de semánticidad para caracterizar a lo mental. Para Brentano, los actos mentales se caracterizan por su intencionalidad, es decir, por el hecho de representarse algo; Searle rechaza tal definición de la “mentalidad”, definiendo los actos mentales por su semánticidad, que englobaría tanto los estados mentales representativos e intencionales (e.g. una percepción) como los no-representativos y no-intencionales (e.g. el dolor).

¿En qué consiste exactamente este carácter semántico general atribuido a la mente humana? *Semántico* ha de entenderse aquí en contraposición a *sintáctico*. Siguiendo la definición clásica de Charles Morris, mientras que la sintaxis se ocupa de las relaciones entre símbolos, la semántica trata con las relaciones entre símbolos y significados. La semánticidad del pensamiento conlleva por tanto una dimensión de significación y de comprensión más allá del mero juego formal entre símbolos, una significación que siempre va

mediada por la subjetividad y en ocasiones también por la intencionalidad del fenómeno mental. (En el caso mencionado del dolor, su semánticidad iría dada exclusivamente por el carácter subjetivo de la experiencia, dado que se trata de un estado mental no-intencional).

Pero ¿qué entiende Searle por *significado*? Para comprenderlo, hemos de pasar a examinar sus nociones de Red y Trasfondo.

3.1. Red y Trasfondo

Para Searle, como vimos, los estados mentales, tanto intencionales como no-intencionales, se sustentan en último extremo en un soporte puramente biológico que constituye el término último del análisis conceptual. Searle explicita tal engarce de lo mental con lo no-mental mediante las nociones relacionadas de la *Red* (*Network*) y el *Trasfondo* (*Background*).

Un estado intencional sólo determina sus condiciones de satisfacción – y solamente así es el estado que es – dada su posición en una Red de otros estados intencionales y respecto de un Trasfondo de prácticas y supuestos preintencionales que ni son ellos mismos estados intencionales ni son parte de las condiciones de satisfacción de los estados intencionales. [Searle 1992a, pág. 34]

La idea de una Red de estados intencionales que se definen por su posición respecto de los demás estados intencionales parece concordar bastante bien con la concepción estructuralista clásica de un *sistema* cuyos elementos se definen estrictamente en términos de sus mutuas diferencias. Sin embargo, Searle se resiste a prescindir de la referencia al mundo objetivo, a la realidad “dura” más allá de la representación intencional, y a la que ésta refiere.

[...] los estados intencionales son en general parte de una Red de estados intencionales y sólo tienen sus condiciones de satisfacción en relación a su posición en la Red. Distintas versiones de esta idea, generalmente llamada “holismo”, son bastante comunes en la filosofía contemporánea [...] Pero también estoy haciendo una segunda afirmación mucho más controvertida: además de la Red de representaciones, hay también un Trasfondo de capacidades mentales no-representacionales; y, en general, las representaciones sólo funcionan, sólo tienen las condiciones de satisfacción que tienen, respecto de este Trasfondo no representacional. [Searle 1992a, pág. 35]

¿Cuáles son estas capacidades mentales preintencionales y no-represen-

tacionales que constituyen el Trasfondo? Searle distingue dos niveles dentro del Trasfondo (Searle 1992a, págs. 152-153): el “Trasfondo local”, que incluiría “las costumbres culturales locales” (tales como “abrir puertas, beber cerveza de las botellas, y la postura preintencional que tenemos frente a cosas como coches, frigoríficos, dinero y cócteles”), y el “Trasfondo profundo”, que incluiría las capacidades comunes a todos los seres humanos en virtud de su naturaleza biológica compartida (tales como “andar, correr, captar, percibir, reconocer, y la postura preintencional que toma en cuenta la solidez de las cosas, y la existencia independiente de los objetos y de otra gente”).

Frente al “saber-que” propio de la Red de contenidos intencionales, el Trasfondo influye sobre el comportamiento en términos de “saber-cómo” (*know-how*).

Para que yo pueda tener ahora los estados intencionales que tengo, debo tener ciertos tipos de saber-cómo: debo saber cómo son las cosas y debo saber cómo hacer cosas, pero los tipos de saber cómo en cuestión no son, en estos caso, formas de “saber que”. [Searle 1992a, pág. 152]

Se puede hablar acerca de la solidez de una mesa y tratar eso como una creencia (la creencia de saber que es una mesa); pero según Searle un enfoque mejor sería considerar la solidez en la medida en que nos influye en términos de cómo tratamos con la mesa en nuestra vida diaria (i.e. sabemos que podemos colocar cosas encima, que podemos subirnos a ella, etc.) De este modo, no tenemos una creencia explícita de que la mesa sea sólida: es sólida en tanto que sabemos cómo tratar con ella como sólida.

En este sentido, resulta interesante señalar que Searle parece reservar el término “estado” para los elementos representacionales de saber-que propios de la Red intencional, mientras que se refiere a los componentes no representacionales de saber-cómo propios del Trasfondo preintencional como “capacidades”, “posturas”, o “costumbres”. Asimismo, y en la medida en que la Red de estados mentales intencionales parece constituir un segundo grado respecto del Trasfondo preintencional en la medida en que se apoya en éste, tal vez podría esbozarse el contraste entre ambas instancias mediante el siguiente esquema:

RED	TRASFONDO
Estados	Capacidades / Posturas / Costumbres
Representacional	No-representacional

Intencional	Preintencional
Saber-que	Saber-cómo
Cultural	Biológico

El Trasfondo parece tener así un carácter básicamente biológico (dado que el Trasfondo “profundo” depende de características de la especie), que puede adoptar diversas formas en su nivel “local” según la interacción del organismo con los diferentes entornos culturales (e.g. abrir una puerta es una acción física que depende de la configuración de brazo, codo, y muñeca, pero que también depende del hecho de encontrarse en un lugar en el que se utilizan puertas y no cortinas de jaima).

Asimismo, el Trasfondo constituye también el fundamento del carácter irreductiblemente subjetivo de primera persona de los estados mentales. Searle postula un tipo particular de causalidad para los fenómenos intencionales, la causalidad de primera persona, porque considera que la constitución biológica de los seres humanos nos lleva a percibir efectos provocados directamente por nuestras acciones o provocados en nosotros por objetos percibidos. Tal constitución biológica es el Trasfondo.

3.2. La Habitación China

Una de las aplicaciones más conocidas de la teoría de la mente de Searle pertenece al ámbito de la Inteligencia Artificial (IA). En su artículo de 1980 “Minds, brains and programs” (“Mentes, cerebros y programas”), Searle se opone a las pretensiones de lo que él denomina la IA “fuerte” de que los ordenadores piensan o al menos pueden (algún día) pensar. Para ello, presenta su célebre *argumento de la Habitación China*. El argumento se basa, según Searle, en dos hechos: por un lado, hay una relación causal entre cerebro y mente: el cerebro causa físicamente los procesos mentales. Por otro lado, una sintaxis no proporciona una semántica, i.e. la intencionalidad no es explicable en términos exclusivamente formales.

Para empezar, Searle distingue entre IA “fuerte” y “débil” según la postura respecto de las simulaciones por parte de computadores de conductas cognitivas humanas. Mientras que la IA débil se contenta con tal status de modelo de dichas simulaciones, que nos permiten “formular y probar hipótesis de un modo más riguroso y preciso”, la IA fuerte pretende que tales simulaciones no lo son en absoluto, sino que “el ordenador programado tiene esta-

dos cognitivos”. Searle considera inaceptable la postura de la IA fuerte, y para ello elabora su argumento de la Habitación China.

El argumento de la Habitación China parte de un programa de Roger Schank que pretendía simular la capacidad humana de comprender historias, siendo capaz de responder a preguntas sencillas acerca de historias sencillas. La IA fuerte, afirma Searle, no ve ninguna diferencia entre esta simulación computacional y el comportamiento humano.

Los partidarios de la IA fuerte afirman que en esta secuencia de preguntas y respuestas la máquina no sólo está simulando una capacidad humana sino también (1) que se puede decir literalmente que la máquina comprende la historia y da las respuestas a preguntas, y (2) que lo que hacen la máquina y su programa explica la capacidad humana de comprender la historia y responder a preguntas acerca de ella.

Para refutar esta tesis, Searle propone un experimento mental, pidiéndonos que imaginemos la siguiente situación. A una persona encerrada en una habitación se le da un fajo de papeles en caracteres chinos. La persona no habla, escribe, lee o comprende el chino en absoluto. Además, se le entrega un segundo fajo de papeles con caracteres chinos más un conjunto de reglas en inglés – un idioma que la persona comprende –, para poner el primer conjunto de caracteres en correlación con el segundo conjunto de caracteres. Finalmente, a la persona se le da un tercer fajo de hojas con caracteres chinos más las instrucciones en inglés acerca de cómo poner en relación los caracteres del tercer fajo con los caracteres del primer y segundo fajo, y cómo devolver ciertos caracteres como respuesta a ciertos caracteres del primer fajo. Sin que la persona lo sepa, los encargados de proporcionarle las instrucciones en inglés y los fajos de caracteres chinos denominan al primer fajo “guión”, al segundo “historia”, y al tercero “preguntas”. Además, a los símbolos devueltos al tercer fajo las denominan “respuestas a las preguntas”, y a las instrucciones en inglés las denominan “programa”.

Searle nos pide también que imaginemos que a esta persona se le da, en inglés, una serie de historias y una serie de preguntas acerca de las historias a las que la persona da respuestas (también en inglés). Obviamente, las respuestas que dé en inglés a las preguntas en inglés acerca de las historias en inglés también resultarán lingüísticamente indistinguibles de las de cualquier otro hablante competente. La cuestión es, ¿son los dos conjuntos de respuestas – las respuestas en chino y las respuestas en inglés – equiparables? La respuesta a esta pregunta determina la aceptación o no de las dos tesis mencionadas antes.

En lo que respecta a la tesis (1) de la IA fuerte – la tesis de que la máquina de Schank comprende literalmente el significado de las historias – Searle afirma que en su ejemplo resulta obvio que la persona dentro de la Habitación China no entiende nada de los símbolos chinos que manipula. Aunque sus inputs y outputs pueden ser idénticos a los de un hablante nativo de chino, no los comprende.

Por otro lado, respecto a la tesis (2) – por la cual el funcionamiento de la máquina y el programa de Schank constituye una explicación del comportamiento humano a la hora de comprender historias –: desde el punto de vista externo de alguien que lea ambos conjuntos de respuestas, tanto las respuestas en chino como las respuestas en inglés son igualmente válidas. Éste es según Searle el punto de vista de la IA fuerte – cuando la persona entiende una historia en inglés, hace exactamente lo mismo – o quizás más de lo mismo – que lo que hacía al manipular los símbolos chinos. Así, la manipulación de símbolos según una sintaxis por parte de un ordenador no difiere sustancialmente de los procesos cognitivos humanos.

Para Searle, el funcionamiento del ordenador no contribuye a la explicación de la mente humana, dado que la analogía entre ambos no incluye la *comprensión* de los contenidos mentales – los ordenadores no comprenden los símbolos que manipulan.

Entonces, ¿qué es lo que tengo en el caso de las oraciones en inglés que no tenga en el caso de las oraciones en chino? La respuesta obvia es que sé lo que aquellas significan, mientras que no tengo ni la menor idea lo que quieren decir éstas. Pero ¿en qué consiste esto, y por qué no podemos dárselo a una máquina, sea lo que sea?

Searle no ve ninguna diferencia importante, a la hora de dar cuenta del funcionamiento cognitivo, entre el hardware de un ordenador y el “hardware” orgánico de un cerebro. Lo que rechaza es que dicho funcionamiento pueda explicarse exclusivamente en términos de procesos computacionales sobre elementos formalmente definidos.

No es porque yo sea la instanciación de un programa de ordenador que soy capaz de comprender el inglés y de tener otras formas de intencionalidad [...], sino, por lo que sabemos, es porque soy una especie dada de organismo con una cierta estructura biológica (ie química y física), y esta estructura, bajo ciertas condiciones, es causalmente capaz de producir percepción, acción, entendimiento, aprendizaje, y otros fenómenos Intencionales. Y parte de este argumento es que sólo algo que tuviese esos poderes causales podría tener esa intencionalidad. Tal

vez otros procesos físicos y químicos podrían producir exactamente estos efectos: tal vez, por ejemplo, los marcianos también tienen intencionalidad, pero sus cerebros están hechos de una materia diferente. Ésa es una cuestión empírica, semejante a la cuestión de si la fotosíntesis puede producirse con una química diferente de la de la clorofila.

Es decir que, como ya vimos, la intencionalidad tiene un soporte biológico, no formal:

Pero el punto principal de este argumento es que ningún modelo puramente formal será jamás suficiente por sí mismo para explicar la intencionalidad, dado que las propiedades formales no son en sí mismas constitutivas de la intencionalidad, y no tienen en sí mismas poderes causales, salvo el poder, al instanciarse, de producir la siguiente fase del formalismo cuando la máquina está funcionando.

Es la semanticidad general de la mente lo que genera la intencionalidad del pensamiento y del lenguaje, y tal semanticidad es una propiedad de la configuración biológica del cerebro humano. Tal como lo expresa el propio Searle, esto conlleva las siguientes consecuencias filosóficas:

- Una máquina *puede* pensar. De hecho, los seres humanos venimos a ser máquinas pensantes.
- Una máquina fabricada por el hombre podría pensar si estuviera dotada de un sistema nervioso lo suficientemente parecido al nuestro como para poder duplicar las causas y efectos físico-químicos que producen la intencionalidad y la consciencia.
- Una máquina que fuera la instanciación de un programa de ordenador podría pensar. De hecho, los seres humanos venimos a ser tales instanciaciones de programas.
- Sin embargo, una máquina tal no podría pensar únicamente en virtud de ser la instanciación de un programa formal. Esto se debe a que las manipulaciones de símbolos formales carecen de intencionalidad, dado que los símbolos tienen sintaxis pero carecen de semántica. Y la intencionalidad – la semántica – es de naturaleza biológica, no formal.

4. Conclusiones

La raíz de la significación, para Searle, se encuentra en la configuración biológica del cuerpo humano: la semanticidad del pensamiento es una fun-

ción evolutiva, de carácter adaptativo, desarrollada por el cerebro orgánico que lo produce. Los programas informáticos carecen de intencionalidad porque consisten exclusivamente en una manipulación sintáctica de símbolos formales, que no constituye en sí misma una semántica.

Searle niega la plausibilidad de una inteligencia artificial fuerte partiendo de una concepción denotacional de la semántica. Según esta concepción, los símbolos formales de la sintaxis reciben su significado mediante funciones que los ponen en relación con otros objetos del mundo. En el caso de Searle, estos objetos que constituirían los significados de los estados mentales intencionales serían estados mentales no-intencionales, de carácter básico, y fundados en último extremo en configuraciones neuronales. Los símbolos manipulados por una máquina carecen de significación por no remitir a un Trasfondo preintencional de carácter biológico, mientras que los símbolos manipulados por un ser humano son semánticos por remitir a dicho Trasfondo. De este modo, Searle hace que todas las características de lo mental – consciencia, intencionalidad, subjetividad, aspectualidad –, dependan en último extremo de este Trasfondo biológico, preintencional y no-representacional.

Pero acaso otra concepción de la semántica sea posible. De hecho, Searle parece admitir, además del significado como denotación proposicional, el significado-como-uso wittgensteiniano (algo que, por otra parte, difícilmente podría rechazar un teórico de los actos de habla). En efecto, el Trasfondo que sustenta la Red de representaciones intencionales equivale de hecho a un conjunto de prácticas, un “saber-cómo” anterior a la representación y a la intencionalidad y que las funda.

Para Searle, como vimos, la afirmación de Austin de que *todos* los usos lingüísticos, incluidos los proposicionales, pueden considerarse actos de habla – que todo “decir” es un tipo de “hacer” – tiene fuertes implicaciones para la definición de la intencionalidad mental. Puesto que Searle considera que lo verbal no es sino una manifestación de lo mental, los fenómenos mentales intencionales de los que derivan los actos de habla también tienen una naturaleza eminentemente pragmática.

En este sentido, y como señala Howard Gardner (Gardner 1987), la teoría de Searle parece obviar en último término la necesidad de postular un nivel de representación interna de carácter simbólico que sirva de soporte a los procesos de pensamiento, mediando entre la percepción de la realidad externa y la consciencia. Searle pareciera adscribirse así más a la “segunda generación” cognitivista, que encuentra inspiración en las ideas del Wittgenstein de las *Investigaciones filosóficas* y cuestiona la idea misma de representación men-

tal, frente a la “primera generación” cognitivista, basada en la analogía de la mente con el ordenador y en la noción de la mente como receptáculo de representaciones internas.

Sin embargo, Searle no lleva la idea de la semánticidad como praxis a sus conclusiones lógicas, negándose a ubicar el fundamento del significado en ningún lugar distinto del mundo físico. Donde Wittgenstein habla del uso lingüístico como basado en una serie de “juegos de lenguaje” intersubjetivos, Searle refiere las prácticas preintencionales, en último término, a una realidad física no lingüística: específicamente, a los procesos cerebrales que causan los procesos mentales básicos que conforman el Trasfondo.

La teoría searleana de la mente parece caer así en una ambigüedad que quedaría cristalizada en la oposición entre Red y Trasfondo. Por un lado, la noción de la Red social recoge el carácter performativo de los estados mentales, así como su aspectualidad, su subjetividad y su causación de primera persona. Pero por otro lado, la noción del Trasfondo biológico supone un anclaje último en las estructuras cerebrales que contradice el carácter externo y social de la semánticidad. Pese a esta ambigüedad, encontramos que su concepción de la semánticidad como definitiva de los fenómenos mentales humanos, y su relación con los dos rasgos principales de tales fenómenos, la intencionalidad y la subjetividad, constituye la aportación más interesante de Searle a la filosofía de la mente.

Según esta aportación, la semánticidad constituiría una propiedad general y específica de la mente humana, de la cual intencionalidad, aspectualidad y subjetividad no serían sino manifestaciones particulares. La semánticidad supone una extensión del principio austiniano por el que toda enunciación lingüística es realmente un *acto* que toma su sentido de su contexto de preferencia y de sus condiciones de felicidad. De igual manera, los fenómenos mentales aparecen como dotados de un sentido que sólo puede aprehenderse examinando holísticamente el entramado de otros fenómenos mentales, tanto propios como ajenos.

Referencias bibliográficas

- AUSTIN, John L. *Cómo hacer cosas con palabras*. Barcelona: Paidós.
GARDNER, Howard (1987) *La nueva ciencia de la mente. Historia de la revolución cognitiva*. Barcelona: Paidós.
SEARLE, John (2001) *Mente, lenguaje y sociedad. La filosofía en el mundo real*. Madrid: Alianza.

- SEARLE, John (2000) *Razones para actuar. Una teoría del libre albedrío*. Oviedo: Nobel.
- SEARLE, John (1992a) *Intencionalidad en la filosofía de la mente*. Madrid: Tecnos.
- SEARLE, John (1992b) *El redescubrimiento de la mente*. Barcelona: Crítica.
- SEARLE, John (1986) *Actos de habla*. Madrid: Cátedra.
- SEARLE, John (1980) "Minds, brains and programs", in *Behavioral and Brain Sciences*, 3, 417-24.