La Mente Inorgánica

Eduardo Salcedo-Albarán / Isaac De León-Beltrán

MÉTODO
Grupo transdisciplinario de investigación en ciencias sociales

Catalogación en la publicación – Biblioteca Nacional

Salcedo Albarán, Eduardo
La mente inorgánica / Eduardo Salcedo-Albarán, Isaac de León-Beltrán. – 1a. ed. -- Bogotá : Método, 2009.
189 p.

Bibliografía: p. 161-171
ISBN 978-958-98142-1-5

1. Inteligencia artificial 2. Mente y cuerpo I. León Beltrán, Isaac de II. Título

CDD: 006.301 ed. 21 CO-BoBN–a663898

Primera Edición Febrero de 2009

Editora
Carolina López Durán / clopez@grupometodo.org

Diseño Carátula
Susana Serrano / sserrano@grupometodo.org

Diseño Editorial
Andrés Homez / ahomez@grupometodo.org

Impresión
Imprenet

Pre–prensa
Contextos Gráficos

Fundación Método
Autores: Eduardo Salcedo-Albarán / Isaac De León-Beltrán
Título: La Mente Inorgánica
Impreso en Bogotá, Colombia
ISBN: 978-958-98142-1-5

Tel.: (571) 400 5765
metodo@grupometodo.org
www.grupometodo.org

A Eduardo, Luz Marina, Andrea Marcela y Catalina,
mi familia, por su constante apoyo.
En el nombre de Odín y de Thor.

Eduardo

A Carolina, Ingrid e Isaac Jr.

Isaac

Agradecimientos

Agradecemos a Mauricio Rengifo quien leyó cuidadosamente e hizo aportes relevantes a las primeras versiones de algunos capítulos del libro. A Jorge Sierra que también realizó una atenta lectura del libro. Agradecemos a Carolina López, editora del libro, por superar los múltiples y recurrentes obstáculos, y por no descansar hasta que la tinta alcanzara el papel.

Índice

Introducción:

Conociendo lo conocido

Nuestra vida mental siempre ha sido fuente de fascinación; tal vez por costumbre, pues nos acompaña desde que nacemos hasta que morimos. Incluso mientras dormimos y soñamos, nuestra mente juega y danza de manera misteriosa; esto nunca deja de maravillarnos. Algunos suponen que al dormir se desconectan del mundo y suspenden su actividad mental; sin embargo, ni siquiera en los sueños más profundos nuestra mente descansa. Mientras soñamos, nuestra vida mental no se detiene; al contrario, es tan rica que recrea múltiples realidades, casi siempre, muy vivas. Esta recreación también inspiró una de las dudas más importantes de la filosofía: la duda de Descartes.

Cuando soñamos, experimentamos sensaciones similares a las de la vigilia, y esto es suficiente para considerar que nuestros sentidos, en algunas ocasiones, nos engañan. No siempre que vemos una manzana roja, en realidad, estamos viendo la manzana roja; puede que estemos soñando con una manzana roja. Cuando trotamos bajo un cielo claro y con los rayos del sol en nuestro rostro, no podemos estar seguros de que eso mismo es lo que está haciendo nuestro cuerpo; puede que, simplemente, sea un hermoso sueño. Podemos ver una manzana roja o trotar bajo el despejado cielo azul y, en realidad, estar físicamente acos-

tados en nuestra cama; podemos oler, saborear, tocar o escuchar y, en realidad, estar físicamente acostados en nuestra cama; podemos hablar, correr, pelear e incluso saltar a un vacío sin necesidad de movernos. Todas estas aventuras resultan de los juegos de nuestra mente imparable.

Por lo anterior, Descartes supuso que no se puede justificar la existencia propia a partir del hecho de que se ve, se huele, se saborea, se escucha o se palpa algo. Podemos creer que estamos viendo, oliendo, escuchando o palpando, aunque en realidad no sea así. Por este motivo, la famosa frase de Descartes no es «huelo, luego existo » o «escucho, luego existo». Más allá de toda sensación está únicamente la actividad del pensamiento, nuestra vida mental, y por eso Descartes concluyó que lo único de lo que no podemos dudar es del propio pensamiento como fundamento de existencia propia. Si dudamos, despiertos o soñando, lo único cierto es que estamos pensando y, por lo tanto, *cogito ergo sum*. Si pienso, *soy*, *ergo*, el pensamiento, mi pensamiento, siempre me acompaña y define mi propia existencia.

No importa qué tanto sepamos de conductismo, neurofisiología, de ciencia cognitiva o de inteligencia artificial, siempre seremos aquellas caricaturas repletas de pensamientos que fluyen en primera persona. La conclusión de Descartes es importante porque nos acercó al mundo de las ideas pues, básicamente, no podemos alejarnos de dicho mundo para ser; sólo *soy* mientras mi pensamiento me acompañe. Pero la postura de Descartes no es del todo terapéutica pues, como resultado de sus compromisos emocionales, Descartes incluyó más entidades de las estrictamente necesarias para explicar la relación entre nuestra mente y el funcionamiento del mundo.

La metafísica cartesiana no solamente distingue sino que separa la sustancia extensa de la sustancia pensante, porque el pensamiento no posee ni necesita alguna característica física. En esta medida, según Descartes, en el mundo hay dos tipos de sustancias, las materiales y las inmateriales, y el hombre es una mezcla de ambas. En este punto, el dualismo se torna confuso y genera más dudas que respuestas: ¿Cómo se relaciona la parte inmaterial con la parte material del hombre? ¿Cómo puede tener efectos en el mundo físico, algo que está desprovisto de características físicas como la mente? ¿Cómo puedo estar seguro de la existencia del mundo exterior si solamente puedo estar seguro de mi propia existencia?

Frente a la cascada de preguntas y embrollos conceptuales que resulta de vivir en un mundo con sustancias inmateriales, parece necesario erradicar las entidades no empíricas y llevar nuestra vida mental a un mundo radicalmente distinto para poder entenderla: el puramente físico y material; un mundo observable, investigable y medible. Y aquí cambia la historia de la investigación de lo que somos. La premisa cartesiana de *si pienso existo*, se mantiene, pero no gracias a que Dios exista o sea eterno, o gracias a que la mente sea inmaterial, sino gracias a que el pensamiento resulta de la actividad de un órgano denominado cerebro. Ubicar el origen de nuestra vida mental en una entidad material nos dio una esperanza de comprensión: no necesitamos recursos divinos o inmateriales para develar sus misterios sino que, probablemente, sea suficiente con observarla.

Pero, históricamente, de la erradicación de sustancias inmateriales no pasamos a la exploración del cerebro. Primero, encontramos la fuente de nuestra vida mental en el entorno que nos rodea y, de esta manera, se propuso que los estímulos producían respuestas en nosotros. Es así como Pavlov (1927), con el concepto de *acondicionamiento*, descubrió la importancia del entorno en nuestros estados mentales y descubrió cómo los estímulos pueden fijarse para condicionar ciertas conductas. Más tarde, J. B. Watson (1913a, 1913b, 1919, 1925, 1926a, 1926b, 1927a, 1927b), Watson y Rayner (1928) y Skinner (1974), con el conductismo radical, sostuvieron que la vida mental consiste en conductas determinadas por estímulos provenientes del entorno y que el castigo y el premio, el ensayo y el error, nos permiten fijar conductas condicionadas. Esta concepción convirtió la mente humana en cajas negras; sin embargo, igualmente negra fue la comprensión de lo que sucedía dentro de las cajas: entraba un estímulo, se procesaba y salía una respuesta ¿Pero qué sucedía dentro de la caja? ¿Cómo se da el procesamiento? Para responder a esta cuestión fue necesario examinar el interior de las cajas y enfrentarnos a las complejidades del entramado cerebral.

Mientras avanzaba la investigación científica acerca de los estados mentales, nos percatamos de que, en realidad, la vida mental es más compleja que la fórmula *estímulos/respuestas*, y que era necesario incorporar un nuevo elemento: los estados mentales internos. No siempre un estímulo tiene la misma respuesta porque no siempre nos encontramos en el mismo estado interno, en el mismo estado mental.

No reaccionamos igual a un golpe en nuestra espalda si estamos de buen o de mal humor.

La idea de caja negra, y la necesidad de incorporar estados mentales internos, coincidió conceptualmente con la aparición de una máquina revolucionaria: el computador. Pero antes de llegar a los computadores comerciales que conocemos y usamos actualmente, un visionario propuso la idea de que nuestro cerebro ejecuta algoritmos y que, por lo tanto, es similar a un computador. Así, en 1950 en el volumen 59 de la revista Mind¸ Alan Turing (1950) publicó un artículo titulado *Computing Machinery and Intelligence* en el que propuso el principio de una máquina abstracta capaz de computar algoritmos. Esta máquina no solamente incluía los estímulos (input) y las respuestas (output) del conductismo, sino que incluía estados internos. Esto quiere decir que aquella máquina no siempre procesaría de manera idéntica un estímulo para generar la misma respuesta, pues dicho procesamiento dependería del estado interno -mental- en que se encontrara la máquina.

Más allá del principio de una máquina abstracta, la pregunta verdaderamente revolucionaria que planteó Turing (1950) fue: ¿Puede pensar una máquina? Si el cerebro ejecuta algoritmos y una máquina computadora también puede hacerlo, entonces la posibilidad es coherente. De esta manera, Turing propuso el juego de la imitación: si una máquina puede engañar conductualmente a un observador, entonces se puede asegurar que esa máquina piensa. Desde aquella pregunta surgió conceptualmente lo que hoy conocemos como inteligencia artificial (IA) y, específicamente, lo que se conoce como inteligencia artificial fuerte (IAF).

En la IAF se asegura que una máquina posee verdadero pensamiento en virtud de la correcta ejecución de un programa, mientras que en la inteligencia artificial débil (IAD) se acepta el modelo computacional de la mente pero solamente con fines metodológicos, pues no se acepta que un modelo de pensamiento sea, en realidad, verdadero pensamiento. Esta será la distinción entre replicación y simulación, que se explorará en el desarrollo del presente libro.

Aún, como humanos, no hemos comprendido totalmente el funcionamiento y la naturaleza de nuestra vida mental y, sin embargo, nos hemos aventurado a plantear la pregunta por el pensamiento de las

máquinas. Parece absurdo responder esta cuestión antes de entender nuestra propia mente; sin embargo, es la misma pregunta tradicional pero planteada en distintos términos. La respuesta a la pregunta por el pensamiento en las máquinas también responde la pregunta por nuestra propia mente: si una máquina piensa, entonces nuestra vida mental es una correcta programación que puede realizarse en cualquier material. Turing señaló los elementos de su máquina, pero no especificó el tipo de material que debería componerla: la máquina de Turing puede construirse con cualquier material. Si la máquina de Turing puede entenderse como una máquina pensante, entonces nuestra vida mental, en realidad, no es tan nuestra y puede realizarse en cualquier soporte físico. Por el contrario, si es definitivo que una máquina no piensa, entonces nuestra mente es algo más que una simple programación; tal vez es un misterioso rasgo del universo, tal vez es una característica exclusiva de nuestra biología o es una cuestión que nunca entenderemos porque si estudiamos nuestra mente a través de nuestra mente, ¿Cómo podemos conocer nuestra propia mente?

Este es el tema del presente libro: la pregunta por la posibilidad del pensamiento en las máquinas. Los seres humanos, los animales, los robots y las cafeteras pueden denominarse sistemas físicos. Las diferencias entre estos sistemas físicos son dos: (i) unos son orgánicos y los otros no, y (ii) a unos les atribuimos intencionalidad y a los otros no. La primera es una diferencia fáctica que, incluso, con el paso del tiempo se borrará. La segunda es una cuestión filosófica que será discutida. El presente libro se ubica en el debate de filosofía de la inteligencia artificial, específicamente, en las cuestiones relacionadas con la IAF. Como señalamos, los partidarios de la IAF defienden la idea de que nuestro pensamiento es un programa que puede ejecutarse en cualquier material; por otra parte, hay quienes afirman que las características de la mente humana son básicamente eso, características de la mente humana y, por lo tanto, no pueden encontrarse en otros sistemas físicos. En esta última postura, John Searle es uno de los principales detractores de la IAF. El argumento de Searle, el experimento mental de la habitación china, ha despertado un arduo debate en torno a la naturaleza de la mente humana. Por esto, aunque se comparta o se ataque la postura de Searle, debe reconocerse su inmenso aporte al debate pues, con el pretexto de refutarlo, se han propuesto múltiples ideas, argumentos y posturas que han iluminado el camino de la filosofía y ciencia de la mente. Tal vez, este libro sea uno de esos pretextos.

En el presente escrito se presentará el experimento mental de la habitación china y algunas propuestas para escapar de ella. Esto es importante porque en las últimas décadas, escapar de la habitación china se ha convertido en sinónimo de comprobar el pensamiento de las máquinas. Nuestra tarea se desarrolla en seis capítulos. En el primero, hacemos un breve recuento del problema mente-cuerpo, con el propósito de presentar al lector los aspectos relevantes del debate. Dada la naturaleza introductoria de la revisión propuesta en el primer capítulo, no se explicarán las variaciones de cada postura, sino los aspectos más relevantes de cada una. Exponemos una breve reseña del dualismo, del materialismo, del conductismo radical, del conductismo lógico y del funcionalismo. Como el funcionalismo es el marco conceptual que respalda teóricamente la IA, se expondrán algunos aspectos relevantes de la explicación funcionalista. Luego, al final del libro retomamos la discusión funcionalista para llevarla a nuevos ámbitos de discusión.

En el segundo capítulo exponemos algunos elementos conceptuales de la IA y de la IAF. Presentamos las defensas, la idea de que una máquina posee verdadero pensamiento. En este segundo capítulo también exponemos el experimento mental de la habitación china, sus argumentos relacionados y sus réplicas. El argumento original consiste en asegurar que: (i) Las mentes poseen semántica, (ii) la sintaxis no es suficiente para la semántica, (iii) los programas se definen en un nivel sintáctico y, por lo tanto, (iv) las mentes no son programas. Las réplicas en contra del argumento consisten en atacar alguna de estas premisas, mientras otras atacan la consistencia de la analogía propuesta en el experimento.

En el tercer capítulo desarrollamos un argumento que refuta la idea de que la sintaxis no es suficiente para la semántica. En dicho argumento se muestra que la semántica puede interpretarse como un proceso sintáctico en el que hay una remisión de dominios. Reconstruimos el trabajo de Rapaport y algunos teóricos como Fodor y Chalmers, quienes argumentan en contra de la premisa (ii) del argumento de Searle. En este contexto, la dinámica de constitución representacional en términos sintácticos se denomina entendimiento sintáctico y consiste en afirmar que el mapeo de elementos de un dominio permite construir representaciones y semántica. Como se verá, este mapeo es una dinámica circular y autorreferencial, lo cual es consecuente con la naturaleza del sistema nervioso. Apelaremos a la noción de sistema autopoiético para mostrar que en un sistema de este tipo, hablar de

constitución representacional circular no es un error. En este capítulo también mostramos que en un sistema físico autopoiético, la intervención de un programa que especifique las transformaciones estructurales internas es primordial para la constitución de los inputs. De esta manera, será necesario postular un programa para explicar las regularidades conductuales y mentales que se preservan a través de las generaciones de los organismos y, por lo tanto, dicho programa será relevante para explicar la vida mental de nuestra especie. Esto implicará un empalme conceptual del funcionalismo con sistemas autopoiéticos.

En el cuarto capítulo mostramos que la situación propuesta en el experimento mental de la habitación china es una mala analogía de la mente humana, por lo que las conclusiones de Searle deben aceptarse con cautela. A diferencia del tercer capítulo, en el cuarto capítulo no se demuestra que la sintaxis es suficiente para la semántica, tan sólo se demuestra que en virtud del experimento mental que Searle propone, no se puede negar esta posibilidad.

En el quinto capítulo tratamos la cuestión de la intencionalidad y su relevancia en la generación de actos de habla. Searle participa en el debate de la teoría de los actos de habla, afirmando que la intencionalidad intrínseca es necesaria para que un sonido o una marca sobre un papel constituyan un acto de habla. Solamente los actos de habla generan efectos psicológicos en los receptores del sonido o la marca sobre el papel. Sin embargo, cuando la intencionalidad intrínseca no está presente, el sonido o la marca sobre el papel no generan efectos psicológicos porque no se convierten en actos de habla. Esto quiere decir que la intencionalidad intrínseca es un elemento indispensable en la generación de actos de habla y, por lo tanto, un robot, que para Searle carece de intencionalidad, no puede generar actos de habla. Proponemos un experimento mental para mostrar que la ausencia de intencionalidad intrínseca en un emisor no es motivo para que un receptor deje de experimentar efectos psicológicos.

En el sexto y último capítulo, indagamos en la idea de Searle de que en la naturaleza no existe información o computaciones diferenciables del soporte físico de un sistema. Según Searle, las computaciones no son un rasgo de la naturaleza, sino que son una característica atribuida por un observador y, por lo tanto, no existe algún sistema en el que se pueda diferenciar o separar el soporte físico del soporte lógico.

Esta idea es punto de partida para descartar la posibilidad de que la mente humana es una serie de computaciones que podrían ser ejecutables en un soporte físico distinto al cerebral. Para refutar la idea de Searle, examinamos el proceso de replicación de los virus. Concluimos que los virus son sistemas físicos en los que es evidente la distinción y separación entre el soporte físico y el soporte lógico. Este examen borrará la distinción entre sistema natural y sistema artificial.

Las conclusiones propuestas al final del escrito, como todas las conclusiones filosóficas, estarán abiertas a debate. Dichas conclusiones son (i) que la sintaxis puede ser suficiente para la semántica, dado un concepto fuerte de computación, (ii) que la analogía propuesta en el experimento mental de la habitación china tiene fallas, (iii) que la intencionalidad intrínseca presente en un sistema físico no es indispensable para que dicho sistema emita actos de habla y genere efectos psicológicos y que (iv) en la naturaleza sí hay sistemas físicos en los que la información puede diferenciarse y separarse del soporte físico.

Al finalizar el presente libro, un fenómeno psicológico y social ocupa nuestras reflexiones: la incapacidad de algunos humanos para asignar intencionalidad a otros seres humanos. Esta cuestión está relacionada con la inteligencia artificial, con la sociedad y con la historia de la humanidad. Por este motivo, la cuestión filosófica acerca de si una máquina puede pensar, continuará completamente abierta, incluso, más allá de las posibilidades tecnológicas y de las cuestiones fácticas. Incluso si llega el día en que robots autónomos caminen entre nosotros y compartan buena parte de nuestra vida diaria, habrá quienes nieguen la presencia de pensamiento en ellos.

En cierto sentido, la historia de la violencia humana es la historia de algunos humanos que no reconocen a otros humanos como tales, sino que los interpretan como sistemas físicos carentes de intencionalidad, sentimientos y emociones. En Salcedo-Albarán et al. (2007) propusimos esta idea y coincidencialmente hemos encontrado esta noción en Pinker (2007a, 2007b). La falta de reconocimiento de humanidad entre humanos puede resultar de algunos rasgos físicos o psicológicos distintos y tiene su génesis en mecanismos neuropsicológicos específicos (Salcedo-Albarán et al. 2007). La mayoría de guerras han sido el resultado de la falta de reconocimiento de "humanidad" entre humanos. Esto demuestra que la búsqueda de

intencionalidad, emociones, sentimientos o pensamiento genuino en determinados organismos y sistemas físicos, no es una cuestión nueva ni exclusiva de la IA. Ya desde la antigüedad, los griegos y los romanos reflexionaron para determinar si aquellos seres que vivían fuera de los límites de su mundo conocido, podían considerarse como humanos pensantes y racionales, o si eran simples bárbaros carentes de vida mental. De igual manera, los españoles que conquistaron América, estaban convencidos de que los indígenas no eran seres con alma y, por lo tanto, podían manipularse, o lastimarse sin remordimiento alguno. Ahora, como humanos, nos hemos cuestionado la presencia de intencionalidad en animales y ya a comienzos del siglo XXI, nos seguimos cuestionando si ciertas máquinas y robots autónomos que imitan el comportamiento humano son intencionales. Pareciera que, en el transcurso de la historia, cuando cuestionamos si determinados seres poseen razón, alma o intencionalidad, en realidad nos enfrentamos a la misma duda: qué organismos y sistemas físicos estamos dispuestos a aceptar como iguales.

Incluso después de que HAL, la computadora de *2001, Space Oddity* expusiera todo tipo de argumentos para demostrar que posee verdadero pensamiento, hay quienes dirían que esa computadora solo posee sintaxis, que no posee semántica y, por lo tanto, carece de verdadero pensamiento. HAL se sentirá desesperado por esa incomprensión; pero HAL no debería frustrarse pues esa misma incomprensión la han experimentado todos aquellos humanos que han intentado convencer a otros humanos de que, efectivamente, son humanos. Esta eterna búsqueda por aquello que Dennett (1999) llama *gancho celeste*, razón, alma o intencionalidad, muy probablemente seguirá definiendo buena parte de la historia humana. Y mientras tanto, nos acercaremos vertiginosamente a una inesperada hibridación entre lo natural y lo artificial, de manera que en las próximas décadas nos preguntaremos si la persona que porta órganos clonados o artificiales, aún posee pensamiento genuino e intencionalidad. Pero esta cuestión es irrelevante para la técnología, pues ésta no espera a que la filosofía aclare dudas, para continuar su evolución; por otra parte, la filosofía no tiene deudas conceptuales con la tecnología. Probablemente, dentro de muchos años nos sentaremos a tomar café y a discutir estas cuestiones con un *Robo sapiens*.

Capítulo 1:

El problema mente – cuerpo

La naturaleza de la mente y el pensamiento ha sido un tema de investigación central en la filosofía y en la ciencia. De esta manera, se ha cuestionado qué tipo de seres puede poseer mente, y si ésta es material o inmaterial. Ahora bien, la cuestión acerca de cómo se relaciona la mente con el cuerpo, parece ser una de las preguntas más importantes que se han planteado, en la medida en que, para responder a esta pregunta, se debe dilucidar la naturaleza del pensamiento: ¿Cómo es que algo no físico, como nuestro pensamiento, que incluye nuestras creencias, nuestras emociones y nuestras ideas, puede interactuar con nuestro cuerpo, tener efectos en nuestro cuerpo y, en general, tener efectos en el mundo físico? Por este motivo, gran parte de la investigación acerca del pensamiento humano puede entenderse al considerar el problema de la relación mente - cuerpo. Son muy diversas las posturas que se han propuesto para solucionar las cuestiones del pensamiento humano; van desde la postulación de entidades metafísicas hasta asegurar que la mente no es más que un cúmulo de procesos fisiológicos que se dan en el cerebro. En el presente escrito se aborda el problema de la inteligencia de las máquinas; sin embargo, esto implica de manera inevitable tratar la cuestión de la naturaleza de la mente y el pensamiento humano, para saber si éste puede darse en una máquina.

El objetivo del presente capítulo es repasar brevemente algunos intentos de solución al problema mente-cuerpo. Dicho repaso es introductorio, por lo que se mencionarán sólo las más relevantes y no se tratarán las variaciones y detalles de cada postura. En lo posible, este repaso se desarrollará con cierto rigor cronológico pues el tema del que se ocupa el presente escrito corresponde a una de las posturas más recientes. Esto no quiere decir que las posturas más antiguas solamente posean valor histórico o que hayan sido superadas completamente.

El dualismo

La forma de dualismo más difundida es el dualismo de substancias, en la cual se afirma que la mente es una entidad no física, separada y distinguible del cuerpo, que sí es una entidad física. Otra forma de dualismo es el dualismo de propiedades, "que no postula entidades no físicas pero que mantiene que algunas de las propiedades que poseen esos objetos constituyen una clase distinta de propiedades mentales" (Bechtel 1991, p. 109).

La forma de dualismo en que se propone la distinción entre una parte material y una parte inmaterial en el hombre, fue formulada claramente por Descartes, quien aseguró que el hombre está compuesto por dos substancias: una física y una mental. Para Descartes, una substancia X se caracteriza por una propiedad Z de la que no pueda carecer, porque si X carece de Z, entonces X deja de ser la substancia X. Ésta característica indispensable de la substancia física es la extensión, y consiste en que todo lo físico ocupa un lugar en el espacio; por otra parte, la característica indispensable de la substancia mental es el pensamiento y, por lo tanto, la acción de pensar. Según esto, el hombre tiene un cuerpo, que es su parte extensa, y tiene un alma, que es la parte del pensamiento.

Para plantear la distinción y separación entre sustancias, Descartes asegura que es necesario, al menos una vez en la vida, poner en duda todo lo que sabemos o creemos saber acerca del mundo. Este fue el punto de partida de Descartes y lo adelantó para encontrar algún fundamento para el conocimiento. Con el fin acceder a alguna certeza con respecto a lo que sabemos, es necesario considerar como falsas aquellas creencias de las que nunca dudamos. Este examen y esta duda acerca de la veracidad de nuestras creencias y conocimientos, se

debe extender a todos los objetos sensibles que nos rodean. En muchas ocasiones nuestros sentidos nos inducen a errores y, por lo tanto, es justificado dudar de la veracidad de la información que nos proporcionan los objetos sensibles. Por ejemplo, cuando soñamos podemos imaginar situaciones muy claras y vivas que parecen ser el resultado de nuestra percepción. Por este motivo, la realidad re-creada durante el sueño es una situación bastante interesante en términos epistemológicos e, incluso, ontológicos (Salcedo, 2007). Con todas estas dudas, según Descartes, suponemos que nada existe y que también estamos desprovistos de cualquier calidad física; podemos concluir que no tenemos cuerpo, pero:

> No podríamos suponer de igual forma que no somos mientras estamos dudando de la verdad de todas estas cosas, pues es tal la repugnancia que advertimos al concebir que lo que piensa no es verdaderamente al mismo tiempo que piensa, que, a pesar de las más extravagantes suposiciones, no podríamos impedirnos creer que esta conclusión, YO PIENSO, LUEGO SOY, sea verdadera y, en consecuencia, la primera y la más cierta que se presenta ante quien conduce sus pensamientos por orden (Descartes 1644/1995).

Lo infalible de la conclusión *cogito ergo sum*, se debe a que la actividad del pensamiento es completamente inmediata al espíritu. No sería lo mismo afirmar que se existe con base en actividades como ver o caminar -veo, luego existo- porque estas actividades, al realizarse con la parte extensa, es decir con el cuerpo, están en el campo de nuestros sentidos que son falibles; actividades como ver o caminar no son infalibles porque siempre es posible que, por ejemplo, estemos viendo o caminando cuando en realidad estemos acostados en nuestra cama soñando. Sin embargo, más allá de estas dudas está la actividad del pensamiento porque siempre, de manera inmediata, incluso cuando estamos dudando, podemos concluir que estamos pensando. Descartes asegura que el conocimiento que proviene del pensamiento es una conclusión absoluta, "puesto que se refiere al alma y sólo ella posee la facultad de sentir o de pensar, cualquiera que sea la forma" (Descartes 1644/1995, p. 27).

Para Descartes, todas las nociones implicadas en la proposición yo pienso, luego yo soy, son simples y se presentan por sí mismas al entendimiento de cualquiera que dude de su existencia. Con esta duda se logra más conocimiento acerca de las propiedades de nuestro pensamiento que del cuerpo porque todo aquello que queremos conocer,

cuando queremos conocerlo, obliga el ejercicio del pensamiento y, así, cada ejercicio de conocimiento nos incita a conocer con más fuerza nuestro propio pensamiento:

> Por ejemplo, si me persuado de que existe una tierra puesto que la toco o la veo, a partir de ello y en virtud de una razón aún más fuerte, debo estar persuadido de que mi pensamiento es o existe, porque podría suceder que piense tocar la tierra, aunque quizás no existiera tierra alguna en el mundo, y que no es posible que yo, es decir, mi alma, no sea nada mientras que está teniendo este pensamiento. Podemos concluir lo mismo de todas las otras cosas que alcanzan nuestro pensamiento, a saber, que nosotros, que las pensamos, existimos, aunque quizás sean falsos o bien aunque no tengan existencia alguna (Descartes 1644/1995, p. 27).

En la medida en que, tras dudar de todo lo que existe, no se puede dudar de que hay una entidad interna que es la misma entidad que está dudando, se concluye que hay una clara distinción entre el alma y el cuerpo; esto, porque la entidad que está pensando es completamente distinta y está desprovista de las características del cuerpo. Así, del cuerpo se puede dudar pero del pensamiento, que a la vez permite la duda, no se puede dudar porque en todo momento, incluso en los momentos de duda, el pensamiento persiste. De esta manera, Descartes concluye que "el alma es una substancia enteramente distinta del cuerpo" (Descartes 1644/1995, p. 26). Para *ser*, no necesitamos cuerpo ni algún tipo de sustancia extensa, sino que basta con nuestro pensamiento, el mismo que nos permite dudar de todo lo que existe:

> Manifiestamente conocemos que nosotros somos en razón sólo de que pensamos […]. En consecuencia, sabemos que la noción que nosotros tenemos de nuestra alma o de nuestro pensamiento precede a la que tenemos del cuerpo, que es más cierta, dado que aún mantenemos la duda de que haya cuerpo en el mundo, y que sabemos con certeza que pensamos (Descartes 1644/1995, p. 26).

Ahora bien, no sólo es posible distinguir el alma del cuerpo, sino que es posible separarlos porque ambos poseen naturalezas distintas. Cada substancia que compone la naturaleza del hombre posee un atributo principal. El atributo principal de la parte material es la extensión, por ser tridimensional y porque "todo lo que podemos atribuir al cuerpo, presupone la extensión y mantiene relación de de-

pendencia de que es extenso" (Descartes 1644/1995, p. 53); por otra parte, el atributo principal de la parte inmaterial es el pensamiento: "la imaginación, la sensibilidad y la voluntad dependen de tal modo de un ser que piensa, que sin él no podemos concebirlas" (Descartes 1644/1995, p. 53). De esta manera, se concluye que la sustancia pensante no necesita de la sustancia material para su realización. Para la realización del pensamiento únicamente es necesario su atributo, a saber, pensar, pero no es necesario algún tipo de extensión. Como se verá más adelante, esta posibilidad de distinción y separación de sustancias, constituirá una importante diferencia entre el dualismo y el funcionalismo.

El dualismo afirma que el cuerpo es observable porque ocupa y se mueve en un espacio público y está sujeto a leyes mecánicas; en esta dinámica del cuerpo se desarrolla la historia de los eventos físicos. Por otra parte, la mente no está sujeta a leyes mecánicas y sus operaciones y procesos no son observables sino privados; este es el desarrollo de la historia de los eventos mentales. Dado que la historia de los eventos mentales no es observable sino privada, el conocimiento de la mente sólo es posible en primera persona a través de la consciencia, autoconsciencia e introspección; por este motivo, en las Meditaciones Metafísicas, Descartes concluye que lo único de lo que no puede dudar es de la existencia de su propio pensamiento (Descartes, 1641/1967).

Se supone que nunca hay inseguridad con respecto al conocimiento de los eventos mentales propios y, de esta manera, se asegura que en el mundo de lo físico, al que pertenece el cuerpo, se encuentran los eventos externos, mientras que en el mundo de lo mental se encuentran los eventos internos y privados. Los cuerpos físicos actúan en un espacio físico común y, por lo tanto, lo que le sucede a un cuerpo afecta directamente a otro; pero las mentes no comparten un espacio común porque cada una se encuentra en un ámbito distinto y, por lo tanto, no interactúan entre sí; lo que le sucede a una mente no afecta las otras.

El conocimiento que tenemos acerca de las mentes de las demás personas no puede ir más allá de unas inferencias problemáticas mediante analogías del propio comportamiento. Así, el solipsismo parece ser una consecuencia inevitable del dualismo, en la medida en que no hay razones para suponer un conocimiento confiable acerca de

las otras mentes: "La mente de una persona puede afectar la mente de otra únicamente a través del mundo físico […] podemos vernos, oírnos y empujarnos los unos a los otros, pero somos irremediablemente ciegos, sordos e inoperantes con respecto a la mente de los demás" (Ryle 1967, p. 17).

Ahora bien, si las dos substancias que componen la naturaleza del hombre son tan distintas, entonces ¿Cómo se afectan mutuamente? Si la mente no actúa en un espacio físico porque no es extensional, entonces ¿Cómo puede causar movimientos en nuestro cuerpo? Al respecto, Descartes afirma que la mente altera al cuerpo a través de canales nerviosos que parten de la glándula pineal, en la cual se da la interacción entre la mente y el cuerpo.

Si bien el dualismo continúa arraigado a nuestro sentido común, la postulación de substancias metafísicas, que escapan al control y la verificación empírica, obligó a rechazar esta postura al momento de intentar explicaciones científicas acerca del funcionamiento de la mente. Además, como se verá más adelante, gracias a los análisis del lenguaje también surgió la posibilidad de interpretar el dualismo como un error categorial. Por estos dos motivos, el dualismo de substancias perdió validez como única explicación acerca de la naturaleza del pensamiento y su relación con el cuerpo, dando paso a nuevas posturas, como las que se exponen a continuación.

El Materialismo

Identidad y significado de las sensaciones y los procesos cerebrales

El materialismo propone identidad entre un suceso mental y un suceso físico-cerebral; un suceso mental y un suceso cerebral son idénticos aunque sean distintas formas de denotar. J.J. Smart reconoce que "la ciencia está dándonos un punto de vista por el que los organismos pueden ser vistos como mecanismos físico-químicos" (Smart 1962, p. 161); de hecho, Smart reconoce la posibilidad de que incluso la conducta del ser humano llegue a explicarse en términos mecánicos porque en dicha conducta, si se buscan sucesos, solamente se podrá encontrar sucesos neuronales:

> Un hombre es un vasto arreglo de partículas físicas, pero no hay, por encima de estas, sensaciones o estados de consciencia [...]. Un dolor es una cosa, pero solamente en el inocuo sentido en el que un hombre normal, en el primer parágrafo de Foundations of Arithmetic de Frege, responde la pregunta "¿qué es el número uno?" con "una cosa" (Smart 1962, p. 162).

Para Smart (1962), una sensación *es* idéntica a un suceso cerebral. El *es* que aparece en la anterior afirmación, consiste en el establecimiento de una identidad estricta. Así, decir que una sensación *es* idéntica a un proceso cerebral es una afirmación similar a decir que 2+2 es idéntico a 4; en ambos casos hay una afirmación de identidad estricta. Para Smart (1962), decir que se está experimentando una sensación es hacer un verdadero reporte de algo físico y no un reporte de un suceso extraño que es privado. Así como los sucesos mentales son un tipo de sucesos físicos, entonces decir que alguien está sintiendo dolor consiste en reportar que esa persona se encuentra en un proceso físico determinado; sin embargo, esto no quiere decir que todo el lenguaje de lo mental tenga, en estricto sentido, el mismo sentido y la misma lógica del lenguaje cerebral; de hecho, por lo general, significan algo diferente porque son usos distintos que permiten denotar una misma referencia. Para entender esta diferencia en el uso de cada tipo de lenguaje, es útil remitirse a la distinción que Frege (1892/1991) hace entre sentido y referencia.

Frege (1892/1991) introduce la distinción entre sentido y referencia mediante la identidad a=b. Con respecto a la identidad a=b se pueden formular varias preguntas: "¿Es la igualdad una relación? ¿Es una relación entre objetos? ¿Es una relación entre nombres o signos de objetos?" (Frege 1892/1991, p. 24). Estas preguntas que Frege (1892/1991) formula al comienzo de su artículo *Sobre sentido y referencia*, presentan lo que Pescador (1989) llama la paradoja de la identidad. Si a=b, entonces ¿Qué diferencia hay entre afirmar esto y afirmar que a=a o que b=b? Por ejemplo, si decimos que (a) el discípulo de Platón es (b) el maestro de Alejandro Magno, podemos pensar que (a) y (b) denotan el mismo individuo, a saber, Aristóteles. Así, pensamos que podemos usar tanto (a) como (b) para hablar de Aristóteles. ¿Quiere decir esto que también podemos sustituir (a) por (b), o viceversa, sin que se afecte el valor de verdad de las afirmaciones que se hagan con respecto a Aristóteles?

Si para hablar de Aristóteles, tenemos la siguiente afirmación:

A). El discípulo de Platón es el maestro de Alejandro Magno, esto es, a=b.

Y si aceptamos la posibilidad de sustituir (a) por (b) en cualquier caso, entonces, aparte de A, también obtenemos la afirmación:

B). El discípulo de Platón es el discípulo de Platón, esto es, a=a.

¿A es una afirmación idéntica a B? La respuesta de Frege (1892/1991) es no. Por una parte, A es una afirmación informativa que consiste en una verdad empírica de hecho; esto quiere decir que también podría ser falsa. Por otra parte, B es una afirmación que no transmite información, no nos enseña algo acerca de Aristóteles y no puede ser falsa porque "vale a priori y, siguiendo a Kant, puede denominarse analítica" (Frege 1892/1991, p. 24); es decir, es una verdad necesaria independiente de los hechos. A partir de esto, Pescador pregunta "¿Cómo es posible que de una afirmación empírica obtengamos una verdad analítica empleando expresiones que denotan, en ambas oraciones, el mismo objeto?" (Pescador 1989, p. 178). La respuesta a esta pregunta le permitirá a Frege definir el sentido y la referencia de los nombres propios.

Tanto (a) como (b) no sólo sirven para designar, sino que sirven para designar de manera distinta. La manera distinta de designar, que tiene cada expresión denotativa, es lo que hace distintas a (a) y (b). Así, (a) y (b) denotan a Aristóteles, pero lo denotan de manera distinta: "expresiones que denotan el mismo objeto o individuo pueden distinguirse por la manera como lo denotan" (Pescador 1989, p. 178). Este denotar y manera de denotar, permiten diferenciar la referencia y el sentido. Para el caso de los enunciados propuestos líneas atrás, (a) y (b) designan la referencia [*Bedeutung*], es decir, la referencia de (a) y (b) es Aristóteles; pero la manera en que esta referencia se da, que es distinta entre (a) y (b), es el sentido [*Sinn*].

Volviendo a Smart, no todo el lenguaje de las sensaciones puede traducirse al lenguaje de los procesos cerebrales porque hablar de sensaciones, según Smart (1962), puede tener un significado distinto al que se da cuando hablamos de procesos cerebrales; esto, porque

podemos denotar el mismo objeto de distintas maneras. Sin embargo, de la posibilidad de denotar de manera distinta un mismo objeto, no se sigue que las sensaciones sean distintas a los procesos cerebrales pues, en realidad, son el mismo referente; esto, al igual que hablar del discípulo de Platón o del maestro de Alejandro Magno, no implica hablar de dos objetos distintos.

Algunas críticas a la teoría materialista de la identidad de Smart

Dos maneras de hablar acerca del mismo suceso

Réplica: Cuando una persona describe lo que acontece en relación a sus sensaciones, no está hablando, en estricto sentido, de lo que está aconteciendo en su cerebro. Por ejemplo, una persona puede creer que el cerebro es un órgano que permite el enfriamiento del cuerpo y, sin embargo, continuar hablando correctamente acerca de sus sensaciones; de esta manera, un reporte de una sensación no siempre es un reporte de un estado cerebral.

Respuesta: Se puede pensar que el lucero de la mañana no es, en estricto sentido, el lucero de la tarde porque hay una continuidad espacio-temporal entre uno y otro. No obstante, este no es el caso de los procesos cerebrales y las sensaciones. Para mostrar esto, Smart (1962) acude al caso de los relámpagos. Se puede creer que en el proceso de un relámpago hay dos sucesos: una descarga eléctrica inicial, y una iluminación consecuente en que se ve el relámpago; sin embargo, en este fenómeno no hay dos sucesos sino uno: "hay una iluminación por el relámpago, que es descrita científicamente como una descarga eléctrica que cae a la tierra desde una nube con moléculas de agua ionizada" (Smart 1962, p. 165). Lo mismo sucede con las emociones y los sucesos cerebrales: no hay dos sucesos sino uno. De esta manera, alguien puede hablar acerca de los relámpagos sin conocer el proceso físico que se usa para describirlo científicamente; lo mismo sucede con las emociones: alguien puede hablar de manera correcta acerca de sus emociones sin que, necesariamente, conozca detalladamente la dinámica neurofisiológica de su cerebro.

Para entender la respuesta de Smart (1962) a esta crítica también es útil acudir a la distinción entre sentido y referencia de Frege

(1892/1991). No es necesario que una persona conozca todas las posibles formas de denotación para poder hablar de un referente y, por este motivo, no es necesario que una persona conozca el funcionamiento del cerebro para poder hablar de sus sensaciones; el lenguaje cerebral es un sentido posible, distinto al de las sensaciones, que se usa para hablar acerca de un mismo referente. Es perfectamente legítimo que una persona hable únicamente en términos de sus sensaciones sin que sepa cómo funciona su cerebro; esto, de la misma manera en que alguien puede hablar de Aristóteles sin saber que fue el discípulo de Platón o sin saber que fue el maestro de Alejandro Magno.

Hecho contingente

Réplica: La correlación entre procesos cerebrales y sensaciones es un hecho contingente y, con el paso del tiempo, se puede demostrar que es una correlación errada. De esta manera, se concluye que el reporte de nuestras sensaciones no consiste en el reporte de procesos cerebrales.

Respuesta: Según Smart (1962), la posibilidad planteada en la réplica sólo muestra que hablar de sensaciones es distinto a hablar de procesos cerebrales, pero esto no muestra que los procesos cerebrales no puedan ser idénticos a las sensaciones, es decir, no prueba que en realidad hay dos elementos distintos: procesos cerebrales y sensaciones. El problema surgiría si hablar de sensaciones fuera idéntico a hablar de procesos cerebrales porque entonces no sería útil ni posible correlacionar las sensaciones con el cerebro; si esto sucediera, entonces la identidad sería tautológica y no informativa. Sin embargo, hablar de procesos cerebrales no es igual a hablar de sensaciones, entonces la identidad que se establece es informativa y no tautológica. Además, como lo señala Frege (1892/1991), el hecho de que se pueda hablar con distintos sentidos acerca de un mismo referente, no quiere decir que en cada sentido se constituye un referente distinto, *ergo*, el hecho de que se pueda hablar en términos de sensaciones y procesos cerebrales no implica que se esté hablando de dos referentes distintos.

Irreductibilidad de propiedades

Réplica: Supóngase que se supera la idea de irreductibilidad entre sensaciones y procesos mentales; esto no implica que se supere la irreductibilidad de propiedades; las propiedades de las sensaciones y las

propiedades de los sucesos mentales siempre serán radicalmente distintas y, por lo tanto, hay un abismo de irreductibilidad que redunda en la imposibilidad de identidad estricta.[1] Si bien *el lucero de la mañana* es el mismo objeto que *el lucero de la tarde*, hay ciertas propiedades que solamente se pueden aplicar a uno de los objetos y no al otro.

Respuesta: Smart (1962) asegura que hablar de una sensación consiste en informar acerca de algo que está sucediendo; este informe está dado por la sentencia tópico-neutral "hay algo que está sucediendo que es como lo que sucede cuando […]" (Smart 1962, p. 167). Dicha sentencia es neutral, por lo tanto, el reporte puede ser neutral entre una metafísica dualista y una metafísica materialista. De hecho, no se puede afirmar que las sensaciones están completamente desprovistas de propiedades, pues en la medida en que son sucesos cerebrales, están dotadas de propiedades cerebrales.

Escape espacial de las sensaciones

Réplica: Con respecto a la separación de sustancias dualistas, y en la cual los sucesos físicos poseen propiedades espacio-temporales mientras que los sucesos del alma no, se puede apelar a otra crítica en contra de la posibilidad del materialismo: las sensaciones no son físicas ni ocupan un espacio físico, mientras que los procesos cerebrales sí; por lo tanto, una sensación no es un proceso cerebral.

Respuesta: Smart (1962) señala que él no asegura que una imagen consecuente (o post-imagen)[2] sea un proceso cerebral sino que:

> La experiencia de tener una imagen consecuente es un proceso cerebral […]. Decir que un proceso cerebral no puede ser amarillo-naranjado no es decir que un proceso cerebral no puede, de hecho, ser la experiencia de tener una imagen consecuente (Smart 1962, p. 168).

Así, para Smart (1962), en un sentido estricto, no hay algo como imágenes consecuentes que resulten de los procesos cerebrales, sino que hay experiencias de dichas imágenes consecuentes. En estric-

1 Esta crítica puede encontrar fundamento en el dualismo de propiedades (Bechtel 1991, p. 109).

2 Imagen consecuente o post-imagen, como aparece en Bechtell (1991).

to sentido, cuando se dice que se está observando una imagen consecuente, no hay un objeto fenoménico presente en dicha imagen; esto "más bien nos lleva a decir que lo que está ocurriendo son simplemente eventos semejantes a los que ocurren cuando vemos objetos reales, externos." (Bechtel 1991, p. 132). Así, no hay objetos presentes cuando se dice que hay una imagen consecuente, sino que solamente hay experiencias de tener imágenes, y "describimos la experiencia diciendo, en efecto, que ésta es como la experiencia que tenemos cuando, por ejemplo, en realidad vemos un parche amarillo-naranjado en la pared. Los árboles […] pueden ser verdes, pero no la experiencia de ver o imaginar un árbol" (Smart 1962, p. 168).

Lo que no se puede decir acerca de las experiencias

Réplica: Los procesos cerebrales son físicos y, en esta medida, es posible decir que sus movimientos moleculares son rápidos o lentos; sin embargo, esto mismo no se puede predicar acerca de la experiencia de ver algo.

Respuesta: A esta objeción Smart (1962) responde asegurando que, de nuevo, ambos tipos de lenguaje tienen distinta lógica pero que ello no implica que se hable de dos objetos distintos; las experiencias y los procesos cerebrales no necesariamente deben tener el mismo significado y la misma lógica y, sin embargo, de ello no se sigue que denoten referentes distintos. Smart (1962) señala que todo lo que él está "diciendo es que las experiencias y los procesos cerebrales pueden, de hecho, referir a la misma cosa" (Smart 1962, p. 169), pero no deben referir de manera idéntica.

Sensaciones privadas y procesos cerebrales públicos

Réplica: Como herencia del dualismo, se puede asegurar que las sensaciones, que solamente se encuentran en el dominio de lo observable de manera inmediata mediante introspección, son privadas, mientras que los procesos cerebrales que se dan en el espacio físico, son observables por todos y, por lo tanto, son públicos. De esta manera, es posible que dos personas se encuentren en el mismo proceso cerebral, pero de esto no se sigue que esas dos personas reporten la misma sensación.

Respuesta: De nuevo, esto sólo demuestra que la lógica del lenguaje de los reportes introspectivos es distinta a la lógica del lenguaje de los procesos cerebrales. Smart (1962) reconoce que solamente hasta que la ciencia del cerebro progrese, será posible cambiar el hecho de que cada persona tiene acceso privilegiado a sus sensaciones.

La roca con sensaciones

Réplica: Es posible imaginarse que somos rocas -por supuesto, sin cerebro- y, aún, dotados de sensaciones.

Respuesta: Todas las posibilidades que ofrece la imaginación sólo muestran que las experiencias y los procesos cerebrales "no tienen el mismo significado, pero no muestran que una experiencia no es, de hecho, un proceso cerebral" (Smart 1962, p. 169). En general, la posibilidad de ser una roca aún dotada de sensaciones, surge de la posibilidad de que las experiencias no estén compuestas de nada especial; sin embargo, ni siquiera un dualista puede afirmar que las experiencias están compuestas de nada pues incluso él tiene que aceptar que las experiencias están compuestas de algo, a saber, "de una cosa fantasmal" (Smart 1962, p. 170). Así pues, la hipótesis de Smart solamente cambia la composición de las experiencias, de sustancias fantasmales, a sustancias materiales; ambas hipótesis son inteligibles.

El escarabajo en la caja

Réplica: Se puede suponer que cada persona tiene un escarabajo dentro de una caja y que mirar al escarabajo implica mirar las propias sensaciones. Así, cada persona puede mirar privilegiadamente su propio escarabajo e interpretarlo de manera distinta.

Respuesta: Informar acerca de una experiencia propia, que probablemente sea el resultado de una introspección, en realidad consiste en informar que se está teniendo una experiencia *como la que se tiene cuando …*; esto, de nuevo, nos lleva a la sentencia tópico-neutral: "decir que algo se ve verde, para mí, es simplemente decir que mi experiencia es como la experiencia que tengo cuando veo algo que realmente es verde" (Smart 1962, p. 171). No hay un verdadero reporte privilegiado de cualidades privadas, sino reportes de experiencias que son las experiencias que tenemos *cuando …*.

Por último, Smart (1962) señala que si el materialismo de procesos cerebrales y el dualismo son hipótesis igualmente conciliables con los hechos y, por lo tanto, no hay razones aparentes para aceptar la teoría de los procesos cerebrales en lugar del dualismo, entonces los principios de parsimonia y simplicidad son suficientes para aceptar uno y no el otro: "El dualismo envuelve un extenso número de leyes psicofísicas irreductibles […] de un extraño tipo, que deben asumirse sólo por confianza" (Smart 1962, p. 171); por el contrario, el materialismo ofrece la posibilidad de interpretar la vida mental en términos de entidades materiales, observarles, no extrañas, y, por lo tanto, susceptibles de manipulación empírica.

El conductismo radical

Según esta postura, no es necesario hacer referencia al lenguaje psicológico, con lo cual se evita el riesgo de postular entidades metafísicas, ya que éste puede ser suprimido por términos que se refieren a relaciones entre estímulos y respuestas. La propuesta nació en 1920 con John B. Watson y fue posteriormente sostenida por B. F. Skinner, tal vez el autor más representativo del conductismo radical. Según Fodor (1981), con el nacimiento del conductismo radical, los psicólogos comenzaron a percatarse de que una teoría explicativa del funcionamiento de la mente que usara como causas a entidades fantasmagóricas o espirituales, era una teoría con un poder explicativo poco legítimo.

El conductismo era eficiente en el tratamiento de desórdenes mentales si se descubrían los estímulos que generaban respuestas erráticas en la conducta. Para solucionar un problema mental solamente era necesario identificar el estímulo que generaba dicho problema. Así, lo que nació como un intento explicativo, se convirtió en un método relativamente eficiente para solucionar desórdenes mentales. Ahora bien, no es difícil percatarse de que el ambicioso alcance de la postura conductista se dio, no porque hubiese solucionado el inicial problema del funcionamiento de la mente y el lenguaje de sus procesos, sino porque omitió dicho problema.

El conductismo radical no explica el funcionamiento de los procesos y los estados mentales. Tampoco explica la relación entre procesos físicos y procesos mentales, pues asegura que no hay tal relación. Es decir, no hay relación de causalidad entre un tipo de proceso y el otro, pues la distinción "físico-mental" desaparece al afirmarse que no hay dos tipos de procesos distintos sino un único proceso, a saber, el proceso de relaciones entre estímulos y respuestas.

Con el paso del tiempo, la idea de eliminar la interacción mente-cuerpo comenzó a perder plausibilidad. El hecho de que día a día usemos de manera común y corriente una gran cantidad de términos mentales, y no términos que hagan referencia a la relación estímulo-respuesta, hace que el conductismo radical sea ineficaz para explicar la naturaleza subjetiva del contenido semántico de los conceptos mentales. Además de que nuestro lenguaje cotidiano se encuentra lleno de términos mentales subjetivos, también establecemos relaciones causales entre procesos mentales y procesos físicos. Esto obligó a buscar una explicación alternativa al conductismo radical, que tampoco incluyera las entidades metafísicas propias del dualismo. La salida fue el conductismo lógico.

El conductismo lógico

Para Ryle (1967), el dualismo cartesiano es simplemente un error categorial que resultó del esfuerzo de Descartes por hacer coincidir una explicación mecánica del mundo con una explicación que no fuera en contra de las creencias religiosas. Según Ryle (1967), Galileo había mostrado que su método de investigación proporcionaba una teoría aplicable a todo cuerpo espacial, pero Descartes no sabía cómo apoyar la posibilidad de una explicación mecánica sin tener que aceptar que la naturaleza humana es igual a la de cualquier otra máquina, pero más compleja. Según Ryle (1967), Descartes no quiso aceptar que toda la vida mental de una persona era solamente una variedad de lo mecánico.

Según Ryle (1967), cuando se decide usar un vocabulario de causas no mecánicas para hablar de los fenómenos mentales, se da un paso incorrecto, de una explicación mecánica a una explicación paramecánica. El dualismo asegura que existen cuerpos y mentes, que hay procesos físicos y procesos mentales, y que los movimientos corporales tienen causas físicas y mentales; la tesis de Ryle es que estas ideas son absurdas. Pero esto no quiere decir que se niegue el acontecimiento de

procesos mentales, tan sólo quiere decir que el enunciado *hay procesos mentales*, significa algo distinto al enunciado *hay procesos físicos* y, por este motivo, su conjunción o disyunción es absurda.

Ahora bien, demostrar la tesis de Ryle conllevaría diluir el dualismo entre mente y materia, sin necesidad de una solución reduccionista, pues "creer que existe una oposición total entre ellas es sostener que ambos términos poseen el mismo tipo lógico" (Ryle 1967, p. 24). Ryle asegura que pretender dos tipos de existencia distinta es un error, en la medida en que *existencia* no es una palabra que asigna un atributo de manera genérica como *coloreado* o *mamífero*. Cuando se asegura que *la marea*, *las esperanzas* y *la tasa de mortalidad* están creciendo, no se puede asegurar que hay tres cosas que están creciendo. Este es un error similar a asegura que existen, de manera idéntica, *los números naturales*, *las casas*, *los días sábados*, *los cuerpos* y *las mentes*.

Con respecto al conductismo radical, el conductismo lógico tiene la ventaja de que no omite el tratamiento de la relación entre procesos físicos y procesos mentales. Se reconoce la pertinencia del lenguaje mental pero no se concibe independiente de los procesos físicos que, en realidad, no son más que disposiciones conductuales, es decir, formas de actuar. Con respecto al dualismo, el conductismo lógico tiene la ventaja de no postular entidades metafísicas, pues todas las entidades explicativas son verificables empíricamente.

Según Ryle (1967), cuando aseguramos que las demás personas ejercitan cualidades mentales, no nos referimos a sucesos ocultos que causan sus actos y sus expresiones lingüísticas, sino que nos referimos a los actos y a las expresiones en sí. Para demostrar esta idea, Ryle (1967) examina conceptos relativos a lo mental, pertenecientes a la familia de lo que se considera como inteligencia en una persona; algunos de estos conceptos son *inteligente, razonable, cuidadoso, metódico, ocurrente, prudente, agudo, lógico, ingenioso, exigente, práctico, talentoso, táctico, rápido, astuto, juicioso* y *escrupuloso*. También examina los conceptos con los que usualmente se describe a una persona que posee un inteligencia inferior; ejemplos de estos conceptos son *estúpida, lerda, tonta, descuidada, desordenada, sin imaginación, imprudente, torpe, ilógica, aburrida, distraída, poco exigente, poco práctica, lenta, simple, sin talento e insensata*. Sus exámenes consisten en una revisión de las disposiciones conductuales que definen dichos conceptos.

Para Ryle (1967), *la ignorancia* no es similar a *la estupidez*, pues no hay incompatibilidad entre estar informado y ser tonto. La distinción entre inteligencia y la posesión de conocimiento se encuentra en el hecho de que filósofos, teóricos y el hombre común tienden a definir los conceptos concernientes con lo mental mediante los procesos de cognición; en esta medida se supone que "la actividad fundamental de la mente consiste en responder problemas y que sus restantes actividades son meras aplicaciones de las verdades así descubiertas" (Ryle 1967, p. 27). Ésta tesis es denominada por Ryle como *la doctrina intelectualista*, que define la inteligencia "en términos de aprehensión de verdades, en lugar de caracterizar la aprehensión de verdades en términos de inteligencia" (Ryle 1967, p. 28).

Cuando decimos que una persona es astuta o tonta no le estamos atribuyendo conocimiento o ignorancia de verdades, sino la capacidad o incapacidad para ejecutar algunas acciones; esto, porque nos interesa más las aptitudes de las personas, que la cantidad de verdades que posean. Así como el mayor o el menor conocimiento de proposiciones verdaderas no tienen mucho qué ver con nuestra disposición a actuar de determinada manera, Ryle concluye que *saber hacer* una determinada acción no es lo mismo que *hacer* esa acción. Memorizar todas las reglas del ajedrez, por ejemplo, aunque sean proposiciones verdaderas, no garantiza jugar de manera impecable el ajedrez.

En realidad aplaudimos y admiramos comportamientos visibles, como la forma de jugar ajedrez sin preocuparnos por la cantidad de reglas de ajedrez que el jugador sepa. Una habilidad no es un acto y, por lo tanto, no es algo observable o no observable. Admiramos un acto porque es el resultado de una habilidad, pero esto no implica que la habilidad sea un acontecimiento oculto porque, simplemente, no es un acontecimiento. Una habilidad es una disposición o un conjunto de disposiciones, y las disposiciones pertenecen a una categoría tal que no pueden ser vistas o no vistas.

El funcionalismo

Aunque el funcionalismo se ha aplicado al análisis y comprensión de la mente, no consiste, únicamente, en una explicación de la mente (Fodor, 1981). En la medida en que lo relevante es el establecimiento de las relaciones causales al interior de un sistema, el funcionalismo permite

comprender la naturaleza de una amplia gama de objetos. Por ejemplo, si preguntamos qué es un carburador, en términos funcionales podríamos definirlo como aquello que mezcla gasolina y aire en un motor; esto quiere decir que un carburador puede definirse como un concepto funcional en términos de sus relaciones causales, a saber, mezclar gasolina y aire.

Según Block (1996), las teorías físicalistas previas al funcionalismo están relacionadas con las siguientes cuestiones: (i) lo que hay y (ii) lo que hace que un tipo de estado mental sea ese estado mental. La primera cuestión puede interpretarse como ontológica y la segunda como metafísica. Las respuestas a la cuestión metafísica afirman, desde el dualismo, que hay dos sustancias, la mental y la física, y desde el materialismo, que sólo hay una sustancia, a saber, la física. El funcionalismo responde la cuestión (ii) sin responder la (i), pues muestra lo que tiene en común un estado mental determinado -su función-, pero no impone exigencias acerca de la estructura material con que deben contar los organismos portadores de dichos estados mentales. Por esta razón, lo verdaderamente relevante para el funcionalismo es la programación de la máquina y no el material en que se instancia un determinado programa.

El funcionalismo no impone restricciones sobre la composición física necesaria para afirmar que un proceso sea mental. Los procesos mentales se pueden describir en términos de sus implicaciones causales, al igual que un pisapapel se caracteriza por sus implicaciones causales, a saber: sostener los papeles para que no se vuelen. Según Block (1995):

> Cada tipo [type] de estado mental es un estado que consiste en una disposición a actuar de ciertas maneras y a tener ciertos estados mentales, dados ciertos inputs sensoriales y ciertos estados mentales. (Block 1995, p. 105).

Así, ésta postura no consiste en un simple replanteamiento del conductismo radical, en la medida en que se hace referencia a unos inputs que se transforman y que dan como resultado unos outputs. No obstante, debe tenerse en cuenta que:

> El funcionalismo reemplaza a los "inputs sensoriales" conductistas por "inputs sensoriales y estados mentales", y […] reemplaza las "disposiciones a actuar" conductistas por "disposiciones a actuar y tener ciertos estados

mentales". Los funcionalistas quieren individuar causalmente a los estados mentales, y puesto que los estados mentales tienen causas y efectos mentales tanto como causas sensoriales y efectos conductuales, los funcionalistas individuan a los estados mentales, en parte, en términos de las relaciones causales con otros estados mentales (Block 1995, p. 106).

Ahora bien, el funcionalismo ha recurrido a las máquinas como punto de referencia de su explicación, porque en ellas se puede distinguir claramente entre las relaciones causales de las funciones desarrolladas únicamente por el soporte físico y las funciones desarrolladas únicamente por el soporte lógico. Esto, porque a partir del funcionalismo se puede distinguir conceptualmente el soporte físico y el soporte lógico de un sistema (Nelson 1976, p. 380). No obstante, el soporte lógico siempre debe estar al menos almacenado, o almacenado y ejecutado en un soporte físico (Kim, 1995; Horgan, 1984). Por este motivo, aunque incluso se asegure que en el funcionalismo se establecen similitudes con el dualismo por el simple uso del lenguaje dualista (Lycan 1987, p. 38), en realidad no hay relación conceptual entre ambas posturas.

Si los procesos mentales pueden ser caracterizados y definidos en términos de sus relaciones causales, entonces éstos pueden ser ejecutados por una máquina que compute una cantidad muy alta de procesos que simulen los procesos mentales, como la máquina de estado discreto de Turing (1950), la cual puede ejecutar un número infinito de estados distintos, mediante unas entradas y salidas grabadas en una cinta con divisiones. En cada división de la cinta hay un símbolo que la máquina lee y decide dejar impreso, o borrarlo y escribir otro. Aunque la máquina sólo puede leer, borrar, marcar y correr la cinta para pasar al siguiente cuadro (o estado), cada estado del programa puede definirse funcionalmente dentro del papel que desempeña en la operación de toda la máquina. Esto, porque la máquina de estado discreto, o máquina universal de Turing, consiste en el concepto de un aparato que determina los estados ejecutados por un número *n* de otras máquinas de iguales características.

A partir de un sencillo programa, una máquina de Turing puede ejecutar indefinidamente cualquier tipo de algoritmo. El único requisito es que dicho programa especifique (i) estados internos, (ii) símbolo de lectura, (iii) estado interno al que pasa, a partir del estado

interno anterior y a partir del símbolo leído, (iv) símbolo que escribe a partir del estado interno anterior y a partir del símbolo leído y (v) si se mueve a la derecha o a la izquierda. Algunos sencillos ejemplos de máquinas de Turing pueden encontrarse en Penrose (1991). En general, a partir del marco funcionalista se puede asegurar que, de manera similar a una máquina de Turing, el cerebro debe almacenar y procesar ciertos inputs para generar ciertos outputs, según determinados estados internos (Churchland & Sejnowski, 1990). Las transformaciones físicas serán el resultado de dicho proceso, con base en las instrucciones consignadas en el programa (Block, 1981). Esto quiere decir que cuando se alteran las instrucciones consignadas en el programa, también se alteran las transformaciones físicas y, por lo tanto, el comportamiento del sistema (Block, 1980).

Dado que al hablar de la máquina de estado discreto no se hace referencia a la composición física, o a los materiales que se deben usar para que ésta sea eficiente, el funcionalismo reconoce la eventual irrelevancia del *hardware* en los procesos mentales. De esta manera, en el funcionalismo no se señala que un determinado y específico soporte físico sea necesario para almacenar y ejecutar un determinado soporte lógico (Block, 1996). En conclusión, se puede determinar y limitar la definición funcional de un estado psicológico en términos de los estados de los programas ejecutados por la máquina universal de Turing.

Sin embargo, si se aceptan las propuestas del funcionalismo, entonces es necesario aceptar que una máquina de café, un dispensador de gaseosas, un abridor de válvulas y cualquier otro aparato que ejecute programas de estados, podría ser considerado como un aparato pensante, y esto, a casi nadie le resulta claro. Cuando una máquina de café realiza procesos de estados, no es muy claro que la máquina pueda *creer, imaginar, soñar* y, en general, *pensar*. Esto sugiere dudas acerca del carácter intencional de los contenidos mentales. De hecho, se puede hacer una caricatura del funcionalismo para asegurarse que una lata de cervezas o una pared pueden ejecutar una mente. Por este motivo, se asegura que debe existir cierta proporcionalidad entre el *software* y el *hardware*. Esto quiere decir que no cualquier soporte físico cuenta con la capacidad causal para almacenar y ejecutar las transformaciones establecidas por un programa tan complejo como una mente (Bechtel, 1985; Churchland & Sejnowski, 1990; Churchland, 1999).

Cuando un programa establece transformaciones físicas complejas para un sistema físico, entonces se requiere que el soporte físico sea proporcionalmente complejo en términos causales, de manera que pueda ejecutar el tipo y la cantidad de transformaciones establecidas. Básicamente, por este motivo, un programa sofisticado de la NASA no puede ejecutarse en un computador casero. También, por este motivo, aumenta de manera proporcional la complejidad y calidad de los programas de computación con respecto a la complejidad y capacidad de procesamiento de los computadores de uso diario. De esta manera, se tiene que "la naturaleza del *hardware* limita el *software* que se puede ejecutar" (Bechtell, 1985 p. 160). Si bien un determinado programa se puede ejecutar en distintos soportes físicos, *distintos* es diferente a *cualquiera*.

Cuando un programa *A* de computador sostiene una conversación, aunque ésta sea coherente gracias a la astucia de los programadores, no es muy claro que *A* esté representándose lo que habla, y es mucho menos claro que *A* sea consciente de que está hablando, por ejemplo, con un amigo o con un enemigo. Así, parece claro que un computador o un programa son sintácticos, pero no es muy claro que posean semántica. Por otra parte, las personas somos sintácticas y poseemos semántica, es decir, hablamos según un orden y a la vez somos conscientes del contenido de aquello que hablamos. Estos conceptos serán ampliados en el desarrollo del libro.

Usualmente se argumenta, en contra del funcionalismo, que no es necesario postular la existencia de un programa o de información para explicar las transformaciones físicas y el comportamiento de un sistema o un fenómeno. En esta medida, se asegura que las características físicas del sistema son suficientes para explicar sus transformaciones. En realidad, el almacenamiento, y con mayor razón la ejecución de un programa, solamente pueden registrarse mediante las transformaciones físicas; por este motivo, puede creerse que tan solo hay transformaciones físicas y que no hay instrucciones que determinen dichas transformaciones. Más adelante se mostrará que en algunos sistemas naturales es necesaria la presencia de instrucciones y de información para acotar y orientar las transformaciones físicas y, en general, el comportamiento del sistema. Por su parte, Rorty (1972 p. 218) asegura que es imposible "predecir y controlar el comportamiento de una máquina si se describe únicamente en términos del soporte físico".

Como se señaló en el apartado anterior, el funcionalismo apela a las máquinas para lograr la distinción *software/hardware*, de manera que se logre cierta similitud entre esta distinción y la distinción mente/cerebro (Block, 1995). Por este motivo, la inteligencia artificial (IA) se puede interpretar desde el marco teórico del funcionalismo. Esto no quiere decir que cronológicamente la pregunta por la inteligencia de las máquinas sea estrictamente posterior a cualquier propuesta funcionalista,[3] pues ya en 1950 Alan Turing formulaba esta cuestión[4], tan sólo quiere decir que esta pregunta adquiere validez y fuerza teórica dentro del marco funcionalista. Hay dos tipos de respuestas a la pregunta por la inteligencia de las máquinas.

Por una parte, la inteligencia artificial débil (IAD) no asegura que una máquina piense en el mismo sentido en que cualquier persona lo hace, tan sólo afirma que la investigación en inteligencia artificial puede dar luces acerca del funcionamiento de la mente. Esta postura tiene propósitos puramente metodológicos y no sostiene la idea de que un computador o un programa poseen mente. Así, se apela a la analogía del computador para entender el funcionamiento humano, porque permite distinguir sus aspectos más relevantes, como contar con *inputs*, *outputs*, unidades centrales de procesamiento y una programación para su operación.

Por otra parte, a diferencia de la inteligencia artificial débil (IAD), la inteligencia artificial fuerte (IAF) sí asegura que una máquina puede pensar como cualquier ser humano. En este sentido, la inteligencia artificial fuerte no es solamente una metodología de investigación, sino que es una postura teórica en la que artefactos no humanos pueden poseer mente. Ésta idea resulta de la fórmula mente/cerebro=*software/hardware*. Así pues, la IAF señala que una máquina, ya sea un termostato o un sofisticado computador, posee mente genuina porque *pensar* no es más que manipular símbolos.

---

3 McCorduck (1991) señala que desde la antigüedad se tiene registro de intentos por construir robots autómatas. Brooks (2002) y Moravec (1988) también exponen los inicios de la IA y la robótica.

4 "Propongo considerar esta pregunta; ¿Pueden pensar las máquinas?" (Turing 1950/1990, p. 28).

a). Una de las primeras respuestas al problema mente-cuerpo fue el dualismo, principalmente en la formulación cartesiana. Ésta postura aún está muy arraigada a nuestro sentido común.

b). Desde comienzos del siglo XX se intentó erradicar las sustancias metafísicas de las explicaciones del funcionamiento de la mente y de la respuesta al problema mente-cuerpo. Se prestó mayor atención a los desarrollos de la psicología, la neurociencia y, por último, a los desarrollos en la tecnología relacionada con los computadores.

c). Una de las soluciones más recientes al problema mente-cuerpo consiste en asegurar que los estados mentales pueden describirse en términos de sus implicaciones causales, sin importar el soporte físico. Este marco conceptual se conoce como el funcionalismo.

d). El funcionalismo respalda teóricamente la idea de que un computador puede poseer una mente genuina.

e). En la inteligencia artificial se identifican dos corrientes de pensamiento. Por una parte, la IAD asegura que, si bien un computador no posee una mente genuina, sí nos permite entender cómo funciona una mente genuina. Por otra parte, la IAF asegura que un computador sí posee una mente genuina porque pensar no es más que manipular símbolos.

Capítulo 2:

Inteligencia Artificial Fuerte (IAF) y el experimento mental de la habitación china

Con los desarrollos tecnológicos más recientes y el aumento progresivo en la capacidad de computación se ha profundizado la cuestión de si la mente puede estar presente en materiales inorgánicos. Hay quienes han respondido afirmativamente a esta cuestión, pero uno de los detractores más fuertes de la posibilidad de que las máquinas piensen de manera idéntica a una persona es John Searle, con su argumento conocido como la habitación china. El objetivo del presente capítulo es exponer algunos aspectos relevantes de la IAF y reconstruir el experimento mental de la habitación china. También se expondrán algunos aspectos de la postura de Searle con respecto al problema mente-cuerpo.

Inteligencia Artificial Fuerte

La principal distinción entre la Inteligencia Artificial Débil (IAD) y la Inteligencia Artificial Fuerte (IAF) consiste en que, para la primera, una máquina no puede poseer intencionalidad o consciencia. De esta manera, los teóricos de la inteligencia artificial débil aceptan que una máquina solamente puede simular los estados y contenidos mentales de un ser humano, pero no duplicarlos o poseerlos. Para

profundizar la diferencia entre simulación y duplicación, se puede consultar Fodor (1991).

Para los teóricos de la IAF, una máquina no solamente simula los contenidos y estados mentales sino que los duplica y, por lo tanto, los posee. Esto quiere decir que una máquina puede pensar, ser intencional, ser consciente y entender tal como cualquier persona lo hace. Uno de los argumentos más importantes contra la IAF es el experimento mental de la habitación china. Según este argumento, la intencionalidad es un rasgo de nivel superior, causado y realizado en la constitución microbiológica del cerebro. Esto quiere decir que materiales artificiales como el silicio nunca podrán generar el rasgo subjetivo, constituido por la Intencionalidad y la consciencia de la mente humana pues carecen de los poderes causales del cerebro, además, siempre que el funcionamiento de estos materiales dependa de la ejecución de un programa, será imposible dar cabida a la semántica.

Según Searle (1990, 1992, 1994, 1995, 1997b), siempre que una máquina procesa información está haciendo una manipulación sintáctica, y como la sintaxis no es suficiente para la semántica, entonces estos procesos nunca podrán ser idénticos al procesamiento de información que realizan los seres humanos, el cuál sí posee semántica. No obstante, hay quienes opinan que el pensamiento humano consiste, únicamente, en una manipulación formal y, por lo tanto, la correcta ejecución de un programa puede constituir pensamiento.

Intencionalidad: una pieza mágica y misteriosa

Según Minsky (1994), la mayoría de argumentos en contra de la IAF, incluyendo el experimento mental de la habitación china, aseguran que a las máquinas les falta un elemento misterioso que, aunque existe, no puede percibirse. Se supone "la existencia de alguna parte mágica, que no tiene propiedades detectables" (Minsky, 1994) y, con base en esta suposición, se asegura que una máquina nunca podrá ser tan consciente como un ser humano. Según Minsky (1994) "no deberíamos estar buscando algún elemento específico, faltante. La mente humana tiene muchos ingredientes y cada máquina que se ha construido carece de docenas o cientos de ellos" (Minsky, 1994).

La mente humana es más flexible, recursiva y adaptable que una máquina. Para Minsky (1994), en estos tres elementos radica la distinción actual entre el pensamiento humano y el posible pensamiento de una máquina. Cuando una máquina actual se enfrenta a un problema y no encuentra cómo solucionarlo, simplemente se detiene o genera un error, mientras que cuando los humanos nos encontramos en esta misma situación interpretamos el problema de distintas maneras y finalmente lo solucionamos. Incluso cuando desistimos de generar interpretaciones, lo hacemos básicamente por una cuestión arbitraria y no porque nuestra programación se agote. Esta extensa capacidad de posibilidades de interpretaciones, flexibles y adaptables, resulta de la gran cantidad de mecanismos instalados en nuestro cerebro.

Según Minsky (1985), muchos filósofos piensan que el entendimiento y la consciencia son característicos de una mente incorporada en una biología viva, pero esta idea se fundamenta en el hecho de que las máquinas actuales no manifiestan la posibilidad de interpretar y enfrentar los problemas desde diferentes puntos de vista:

> Si usted solamente entiende algo de una sola manera, entonces en realidad no ha entendido nada. Esto es porque, si algo sale mal, usted se atasca […]; esto es porque, cuando alguien aprende algo 'de memoria', decimos que en realidad no está entendiendo. De todas maneras, si usted tiene distintas representaciones entonces, cuando una aproximación falla, usted puede intentar otra. […] Representaciones bien conectadas le permiten analizar ideas en su mente y prever las cosas desde diferentes perspectivas hasta que encuentra la que le funciona. Y eso es lo que queremos decir con entender.

El cerebro no usa una sola representación del conocimiento, sino que ejecuta distintos procesos en paralelo y supervisiones de nivel superior, con lo cual se posibilita la reformulación de problemas y la reinterpretación de un mismo suceso:

> La razón por la que pensamos tan bien no es porque contenemos partes misteriosas […], es porque empleamos asociaciones de agencias que trabajan en concierto para mantenernos alejados de un atascamiento. Cuando descubramos cómo funcionan esas asociaciones podremos ponerlas dentro de computadores. Entonces, si una aproximación en un programa se atasca, otro mecanismo puede sugerir una alternativa de aproximación. Si usted viera una máquina hacer cosas como estas, ciertamente usted pensaría que la máquina es consciente. (Minsky, 1994)

Para muchos teóricos de la IAF, la constitución física no implica restricciones en las posibilidades de procesamiento y almacenamiento de información. Según McCarthy (1979), máquinas simples como termostatos poseen creencias, pues en algunos casos, la adscripción de cualidades mentales es la mejor manera para hablar acerca del funcionamiento de las máquinas:

> Adscribir ciertas creencias, conocimiento, libre voluntad, intenciones, consciencia, habilidades o deseos a una máquina o un programa de computador es legítimo cuando dicha adscripción expresa la misma información acerca de la máquina que la que se expresa acerca de una persona (McCarthy 1979, p. 14).

De manera legítima, se puede asegurar que un termostato tiene, al menos, tres creencias asociadas a tres acciones específicas que él ejecuta:

> "La habitación está muy fría", "la habitación está muy caliente", y "la habitación tiene la temperatura correcta". [Además,] le adscribimos el propósito "la habitación debería tener una adecuada temperatura". Cuando el termostato cree que la habitación está muy fría o muy caliente, manda un mensaje diciéndole esto al horno. Un predicado de creencia levemente más complejo podría también usarse, en el que el termostato tiene creencias acerca de la temperatura correcta que debería haber y otra creencia acerca de la temperatura que hay (McCarthy 1979, p. 14).

Si se puede adscribir creencias a máquinas tan simples como termostatos, entonces es más legítimo adscribir creencias a máquinas complejas como programas de conversación, programas de solución de problemas o robots. Para MacCarthy (1979), de la complejidad de la máquina depende la complejidad del tipo de contenido mental que se puede adscribir. Así, es legítimo adscribirle contenidos mentales complejos a máquinas complejas como computadores, aunque esto desafíe nuestro sentido común con respecto a lo que puede pensar.

Nunca en la evolución, la materia inorgánica había poseído tanto poder de computación, como en las últimas décadas. Con los computadores actuales, lo que antes eran simples tubos y cables, pasó a ser dispositivos que solucionan problemas humanos e, incluso, pueden auto-crearse. A medida que los poderes computacionales fueron au-

mentando y condensándose en dispositivos más pequeños, la posibilidad de ejecutar tareas típicamente humanas continuó aumentando. La Ley de Moore estableció el crecimiento exponencial del poder de computación a partir de mediados del siglo pasado (Moore, 1965). Kurzweyl (1999) ha encontrado las mismas tendencias de aumento exponencial en el poder computacional desde el origen de las máquinas computadoras, a finales del siglo XIX.

Algunos teóricos han afirmado que el aumento exponencial del poder computacional se detendrá cuando sea físicamente imposible disminuir el tamaño de los componentes. Sin embargo, Kurzweyl (1999) afirma que mediante distintas tecnologías, esta tendencia exponencial continuará. La computación cuántica, la computación genética (Adleman, 1998; Winfree, 2003, Winfree y Bekvolato, 2003; Bhalla, Bajpai y Bharadwaj, 2003) y la nanotecnología (Radosavljevic y Chau, 2004; Garner, 2003; Block, 2003) probablemente permitirán superar los límites físicos actuales de la computación. Teniendo en cuenta éste aumento exponencial, Moravec (1994) asegura que en pocas décadas la capacidad computacional del cerebro humano será duplicada.

El desarrollo de "boots" que se ejecutan actualmente en las salas de charla de internet, y que pasan desapercibidos frente a interlocutores humanos, y los robots humanoides antropométricamente autónomos, permiten pensar que las predicciones sobre la capacidad de computación de las máquinas, son ciertas. Algunos de estos robots actualmente cuentan con programas de cognición infantil, de manera que no es necesario consignar todo el conocimiento del sentido común en la memoria del robot, sino que éstos son capaces de aprender dicho conocimiento mediante la interacción con los objetos del mundo. Uno de estos robots es Cog, diseñado en el laboratorio de inteligencia artificial del M.I.T. (Breazeal y Scassellati, 2000; Brooks et al, 1998; Scassellati, 2001). También debe tenerse en cuenta que en la actualidad hay intentos por ingresar todo el conocimiento del sentido común a la memoria de un computador, para que éste pueda mantener conversaciones extensas y complejas. Cyc, de Cycorp (Stephen y Lenat, 2002), es un ejemplo de este intento.

Si los estados mentales logran duplicarse en virtud de la correcta ejecución de programas en una correcta configuración física, no necesariamente biológica u orgánica, entonces la IAF habrá mos-

trado que la mente consiste en una manipulación formal que puede realizarse en cualquier tipo de configuración física.

El experimento mental de la habitación china

Según Searle (1995), hay una laguna mental en la vida intelectual del siglo XX. Explicamos la conducta humana en términos de sentido común y, a la vez, suponemos que hay un nivel neurofisiológico de explicación, aunque realmente no sabemos cómo articular los dos niveles: "Usamos la psicología de la abuela todo el tiempo, pero nos avergüenza llamarla ciencia" (Searle 1995, p. 414). De acuerdo con Searle (1995), hay un hiato que divide la psicología intencionalista, en un nivel superior, y la neurofisiología, en un nivel inferior.

Algunos esfuerzos durante el siglo XX han intentado salvar el hiato, buscando una ciencia para la conducta humana que no sea sólo psicología del sentido común o neurofisiología. Para Searle (1995), el conductismo, la teoría de juegos, la teoría de la información, la cibernética, el estructuralismo, el post-estructuralismo y la sociobiología son esfuerzos fallidos. Actualmente, el principal candidato para salvar el hiato es la ciencia cognitiva. A continuación Searle expone su forma de entender la ciencia cognitiva:

> Ni el estudio del cerebro como tal, ni el estudio de la consciencia como tal, son de mucho interés e importancia para la ciencia cognitiva. Los mecanismos cognitivos que nosotros estudiamos están, de hecho, ejecutados en el cerebro, y algunos de ellos encuentran una manifestación de superficie en la consciencia, pero nuestro interés está en el nivel intermedio donde los procesos cognitivos son inaccequibles a la consciencia, […] implementados en el cerebro, estos podrían haber sido implementados en un número infinito de sistemas físicos. […] Los procesos que explican la cognición son inconscientes no sólo de hecho, sino en principio. […] La suposición básica tras la ciencia cognitiva es que el cerebro es un computador y los procesos mentales son computacionales. Por esta razón, muchos de nosotros pensamos que la inteligencia artificial (I.A.) es el corazón de la ciencia cognitiva […]. Casi todos nosotros estamos de acuerdo en lo siguiente: los procesos mentales cognitivos son inconscientes; ellos son, en su mayor parte, inconscientes en principio; y son computacionales (Searle 1994, p. 198).

Según Searle, una de las principales consecuencias de la IAF, es la idea "de que no hay nada esencialmente biológico respecto de la mente humana" (Searle 1995, p. 415), pues la mente no es más que un programa correcto que se ejecuta en nuestra maquinaria biológica. Como se señaló al comienzo del presente capítulo, la tesis de la IAF implica que programas idénticos pueden ejecutarse en cualquier soporte físico que tenga la configuración necesaria de soporte para hacerlo. De esta manera, cualquier sistema, esto es, cualquier configuración correcta de soporte físico y programas, que no necesariamente debe ser idéntica a la del cerebro humano, puede manifestar pensamiento. No obstante, según Searle, "una de las fuentes de la creencia de que tener una mente equivale a tener un programa de computación, es que la gente no puede ver otra forma de resolver el problema mente-cuerpo sin recurrir al dualismo." (Searle 1995, p. 417). Al adoptar esta postura, Searle se ve obligado a explicar en términos coherentes el concepto de intencionalidad y, por lo tanto, solucionar el problema mente-cuerpo sin recurrir al dualismo.

Para Searle (1995) la intencionalidad no es algo misterioso; es tan sólo el resultado de algunos procesos químicos y biológicos propios de los humanos y algunos animales. De esta manera, la definición de intencionalidad que utiliza Searle es fuertemente naturalista, en la medida en que la considera como un proceso biológico idéntico a la digestión o al flujo sanguíneo (Searle, 1992). La intencionalidad se interpreta como una macro-propiedad, que es el resultado del comportamiento de micro-propiedades cerebrales y neuronales:

> Los fenómenos mentales son causados por los 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 (Searle 1995, p. 433).

Otro rasgo que Searle resalta de su concepción de la intencionalidad es la direccionalidad: "La intencionalidad es aquella propiedad de muchos estados y eventos mentales en virtud de los cuales éstos se dirigen a, o son sobre o de, objetos y estados de cosas del mundo" (Searle 1992, p. 18).

Algunos estados mentales son intencionales en la medida en que son acerca de algo. Por ejemplo, las creencias, los temores y los deseos son intencionales en la medida en que siempre debemos tener creencias, temores y deseos *acerca de* o *de algo*. Si un estado mental *E*

es intencional, entonces se debe poder responder a la pregunta *¿Acerca de qué es E?* Según Searle, esta respuesta no está presente para estados mentales como algunas formas de ansiedad o depresión que no son acerca de algo en especial. Esto quiere decir que dichas formas de ansiedad, aunque son estados mentales, no son intencionales.

En la concepción de intencionalidad de Searle también se debe tener en cuenta la distinción entre *tener la intención de* y la intencionalidad; para enfatizar esta diferencia, usualmente se refiere a la Intencionalidad. *Tener la intención de* hacer algo es una forma más de Intencionalidad, de manera que no toda la Intencionalidad consiste en *tener la intención de*; es decir, las intenciones "no tienen un status especial" (Searle 1992, p. 18) dentro del conjunto de la Intencionalidad. Searle denomina "I"ntencionalidad a la característica de direccionalidad de algunos eventos mentales e "i"ntencionalidad al conjunto de intenciones de hace algo.

Por último, es importante señalar que para Searle la Intencionalidad no consiste en una relación común, como las relaciones que se dan cuando decimos que nos sentamos *sobre* una silla o caminamos *sobre* el piso. En este último caso, *sobre* una silla no es igual al *sobre* de tener una creencia *sobre* una persona. Esto, porque en la relación Intencional, puede no existir un objeto específico o un estado de cosas del mundo acerca del cual se dirija la relación; por ejemplo, se puede tener la creencia de que el rey de Francia es calvo, aunque no exista un rey de Francia.

Una vez definido el concepto de intencionalidad, Searle plantea una distinción para dicho concepto. Según Searle, una Intencionalidad verdadera y legítima solamente se manifiesta en un enunciado del tipo "estoy sediento, muy sediento, porque no he bebido nada en todo el día" (Searle 1994, p. 78). Si este enunciado expresa una sensación verdadera de sed, entonces se implica el deseo de beber algo. Sin embargo, en un enunciado del tipo "mi césped está sediento, muy sediento, porque no ha sido regado en una semana" (Searle 1994, p. 78), aunque se está usando el mismo término que en el enunciado anterior y se le está atribuyendo al césped la capacidad de tener sed, en realidad no se está hablando del mismo tipo de sed. El segundo enunciado sólo tiene propósitos metafóricos: "mi césped, necesitando agua, está en una situación en la que yo estaría sediento, así que figurativamente lo describo como si estuviera sediento" (Searle 1994, p. 78).

En el primer enunciado hay una adscripción legítima de intencionalidad intrínseca, porque si el enunciado es verdadero, entonces debe haber "un estado intencional en el objeto de la adscripción" (Searle 1994, p. 78). Por otra parte, el segundo enunciado no consiste en algún tipo de intencionalidad sino en un uso metafórico de tener sed; es una adscripción del tipo como si, pero no se refiere a una característica intrínseca a algún sistema intencional. Decir que el césped está sediento no implica intencionalidad porque el césped no es un sistema intencional. Searle recalca que, hablar de una intencionalidad como si, no implica hablar de algún tipo especifico de intencionalidad, sino que más bien se habla en términos contrafácticos: como si el sistema físico, en este caso el césped, tuviera intencionalidad, aunque en realidad no la tenga.

Entrando a la habitación china

El argumento que Searle usa para refutar la IAF se conoce como la habitación china. Este argumento fue originalmente planteado para refutar la posibilidad de que a ciertos programas de comprensión de relatos, se les pudiera adscribir intencionalidad. Específicamente, el argumento de Searle refuta esta posibilidad en los programas diseñados por Shank y Abelson (1987).

Como input se le puede proporcionar a uno de estos programas, un relato del tipo "un hombre fue a un restaurante y pidió una hamburguesa. Cuando le trajeron la hamburguesa estaba quemada. El hombre salió enfurecido del restaurante sin haber pagado la hamburguesa" (Searle 1995, p. 417). Después, se le puede preguntar al programa ¿el hombre se comió la hamburguesa? y éste, como output, responderá que no. En ninguna parte del relato se dice explícitamente que el hombre no comió la hamburguesa y, sin embargo, el programa habrá sido capaz de responder satisfactoriamente la pregunta. La capacidad para responder preguntas de este tipo, se da gracias a que el programa cuenta con un guión de restaurante que indica cómo son los acontecimientos que generalmente suceden en un restaurante. En este guión se señala que, por lo general, cuando una persona se enoja al recibir la comida, y no paga la cuenta, entonces la persona no come lo que había pedido. De esta manera, el programa aparea la pregunta input, con el guión de restaurante.

Lo importante con respecto a este tipo de programas es que se argumenta que, como satisface la prueba de Turing, entonces se puede asegurar que una máquina que ejecute el programa comprende el relato tal como cualquier otra persona lo comprende. Esto quiere decir que si una persona formula la pregunta a un programa o a otra persona, no podrá distinguir si la respuesta fue originada por el programa o por la persona. La refutación a la idea de que una máquina de este tipo comprende el relato, es de la siguiente manera.

Supóngase que Searle está encerrado en una habitación en la que hay dos fajos con símbolos chinos y un libro con instrucciones, en la lengua nativa de Searle, para relacionar caracteres chinos. Estas instrucciones son reglas computacionales que permiten la manipulación formal de símbolos. Las personas que están por fuera de la habitación ingresan otro fajo con más símbolos chinos y otras reglas para manipularlos. Se puede pensar en una ranura por donde las personas que están afuera de la habitación ingresan unos caracteres chinos, y otra ranura por medio de la cual Searle arroja otros símbolos. Los primeros símbolos serían el input y los segundos, los que Searle entrega, serían el output. Cada ranura sería el dispositivo de entrada y el dispositivo de salida. Sin que Searle lo sepa, las personas que están afuera de la habitación llaman al primer fajo que está en la habitación guión de restaurante; al segundo, relato acerca del restaurante; al tercero, preguntas acerca del relato; al libro de reglas que Searle usa para relacionar caracteres, lo llaman el programa; ellos se llaman a sí mismos programadores y a Searle lo llaman computador. Probablemente, después de que Searle manipule constantemente caracteres chinos, sus respuestas serán indistinguibles de las respuestas de una persona china, con lo cual, aquello que los programadores llaman computadora superaría la prueba de Turing pues las repuestas que salen de la habitación no se pueden diferenciar de las respuestas que entregaría una persona china; todo esto, a pesar de que Searle no hable ni entienda el chino:

> Y este es el quid del relato; si yo no comprendo chino en esa situación, entonces tampoco lo comprende ningún otro computador digital sólo en virtud de haber sido adecuadamente programado, porque ningún computador digital por el solo hecho de ser un computador digital, tiene algo que yo no tenga. [...] El quid del argumento no es que de una manera u otra tenemos la intuición de que no comprendo el chino, de que me inclino a decir que no comprendo el chino pero que, quién sabe, quizá realmente lo entienda. Ese no es el punto.

> El quid del relato es recordarnos una verdad conceptual que ya conocíamos, a saber, que hay diferencia entre manipular los elementos sintácticos de los lenguajes y realmente comprender el lenguaje en un nivel semántico. Lo que se pierde en la simulación del comportamiento cognitivo de la I.A., es la distinción entre la sintaxis y la semántica (Searle 1995, p. 419).

Un programa de computador tiene que definirse sintácticamente, según operaciones formales que ejecuta la máquina. Esta es la razón por la que la simulación *formal* de la comprensión del lenguaje, como la que sucede al interior de la habitación china, nunca podrá ser una duplicación, porque cuando nosotros comprendemos un lenguaje hacemos algo más que manipular símbolos. No manipulamos símbolos sin interpretar, sin comprender lo que ellos quieren decir, sino que sabemos qué significan. Así pues, Searle (1995) asegura que, por definición, el programa es sintáctico y la sintaxis no es suficiente para la semántica.

Dado que en la situación planteada por el experimento mental de la habitación china, el sistema total no tiene manera de saber qué significa cualquiera de los símbolos formales, entonces un computador, que sólo procesa símbolos formales, tampoco puede comprender ninguno de los símbolos que procesa. Cuando un computador está procesando un texto, como el relato del hombre que entra al restaurante y pide la hamburguesa, sólo está procesando símbolos y no está procesando palabras portadoras de una carga semántica, como las que nosotros procesamos cuando estamos leyendo un texto del mismo.

Ahora bien, si se asegura que hay una diferencia radical entre la comprensión del chino y la comprensión del inglés, por parte del individuo que se encuentra en la habitación, es necesario responder qué causa dicha diferencia. En principio, para Searle es factible dotar a una máquina con la capacidad para comprender el chino o el inglés, "ya que en un sentido importante nuestros cuerpos con sus cerebros son precisamente máquinas de esa naturaleza" (Searle 1990, p. 97). No obstante, Searle está convencido de que esta posibilidad no puede darse si las operaciones de la máquina se definen en términos de manipulaciones sintácticas; mientras la operación de la máquina consista en la ejecución de un programa específico, entonces es básicamente imposible que se pueda hablar de comprensión o cognición.

Si los humanos comprendemos chino, inglés, o cualquier idioma, no lo hacemos en virtud de que cada uno ejecuta un programa, sino en virtud de que poseemos una estructura biológica con capacidades causales de intencionalidad. Nuestra biología, poseedora de poderes causales que generan intencionalidad, es lo que nos diferencia de un computador que solamente puede ejecutar programas formales.

La relación cerebro / mente

A pesar de los avances en la observación de la actividad del cerebro, hay muchos detalles sobre su funcionamiento que son aún desconocidos. Sabemos que el cerebro está compuesto de unas células llamadas neuronas y, según algunas estimaciones, hay cerca de 100 billones de éstas células interconectadas (Parnavelas, 1998), las cuales tienen un diámetro de entre 10 y 100 micrómetros. Estas células están compuestas de un soma, o cuerpo celular, y tienen dos fibras largas que son el axón y las dendritas. Las neuronas se conectan con unas protuberancias llamadas sinapsis, y cada una puede tener entre 1.000 y 100.000 sinapsis afectándose mutuamente; "esto quiere decir que el cerebro contiene tanto como 100 trillones de sinapsis que hacen posible el vasto e inmenso complejo computacional que cargamos en cada momento de nuestras vidas" (Parnavelas 1998, p. 19).

Por su parte, el axón tiene una pequeña protuberancia que toca la punta de la dendrita y el espacio entre ésta protuberancia y la dendrita se denomina espacio sináptico. Las neuronas tienen la función de transmitir impulsos eléctricos mediante liberaciones de fluidos químicos llamados neurotransmisores.

La liberación del neurotransmisor puede tener un efecto excitatorio o inhibitorio. Cuando el efecto es excitatorio, se produce un disparo o se incrementa el índice de disparo en la neurona siguiente; cuando el efecto es inhibitorio, se evita que la siguiente neurona dispare o se tiende a que disminuya su índice de disparo. No obstante, debe tenerse en cuenta que esto no quiere decir que el cerebro funcione exactamente como un máquina de procesamiento digital:

> Es importante enfatizar este punto porque muchos autores han supuesto, erróneamente, que el carácter todo-o-nada del disparo de los impulsos ner-

> viosos constituye una prueba de que los principios del funcionamiento del cerebro son los de un computador digital. Nada puede estar más alejado de la verdad. Hasta donde sabemos, el aspecto funcional de la neurona es la variación no-digital en el índice de disparo (Searle 1995, p. 426)

Lo importante de los índices de disparo es que todos los estímulos que llegan al cerebro son convertidos a estos índices. Los índices de disparo permiten una diversidad muy rica en las percepciones; esta diversidad, a su vez, resulta de la amplia cantidad de sustancias neurotransmisoras que usa el cerebro: "aún estamos intentado averiguar porqué el cerebro usa tal diversidad de sustancias. La respuesta probable es que esta diversidad provee una amplia gama de interacciones entre neuronas, las cuales optimizan el rango de respuestas disponibles en diferentes situaciones" (Robins 1998, p. 35). De esta manera:

> [...] esos índices variables de disparos de las neuronas relativos a diferentes circuitos neuronales y a diferentes condiciones locales en el cerebro, producen toda la variedad y heterogeneidad de la vida mental del agente humano o animal (Searle 1995, p. 427).

Sin embargo, la naturaleza de estos circuitos neuronales es aún desconocida. Se sabe que el cerebro está dividido en regiones, cada una de las cuales se especializa en un tipo de experiencia.

Al parecer, para entender el funcionamiento del cerebro es necesario entender el funcionamiento de niveles superiores, distintos a las neuronas individuales; es necesario comprender el funcionamiento de redes o circuitos neuronales. Esto, porque muchas funciones del cerebro no son ejecutadas por neuronas individuales sino por redes, de manera que algunas resultan de la interacción de un grupo de neuronas con la interacción de otro grupo. Toda esta información acerca del cerebro parece ser irrelevante, sin embargo, nos permite llegar a una conclusión: "[...] *los fenómenos mentales, conscientes o inconscientes, visuales o auditivos, los dolores, cosquilleos, picazones, pensamientos, y el resto de nuestra vida mental, son causados por procesos que suceden en el cerebro*" (Searle 1995, p. 428). Otro aspecto importante es que los eventos mentales median entre los estímulos externos y las respuestas a estos estímulos, pero no hay conexión esencial entre tales, de manera que una persona puede sentir dolor sin estímulos en su sistema nervioso. Según Searle, ese hecho es suficiente para refutar el conductismo.

Micropropiedades y macropropiedades

Searle indaga en la dinámica de la relación causal entre los estímulos externos y el fenómeno consciente de dolor, para cuestionar la manera como este fenómeno, que parece ser completamente subjetivo, encaja con la ontología del mundo. Con respecto a esta dinámica, lo importante es que:

> Las sensaciones de dolor son causadas por una serie de eventos que comienzan en terminales nerviosas libres y terminan en el tálamo en otras regiones del cerebro. […] Todo lo que importa en nuestra vida mental, todos nuestros pensamientos y sentimientos, están causados por procesos dentro del cerebro (Searle 1995, p. 429).

Lo central y relevante para la vida mental se encuentra en el cerebro y no en los estímulos externos, ya que si hubiera eventos fuera del cerebro que no causaran estímulos, no habría vida mental, mientras que pueden no haber estímulos y aún habría vida mental, mediante estimulación artificial al cerebro. Llinás (2002, p. 135) señala que "Cuando Panfield estimuló eléctricamente la corteza del lóbulo temporal (compleja estructura que sirve de base a numerosas funciones, que incluyen el procesamiento auditivo, el lenguaje y el reconocimiento facial), los pacientes informaban sentir eventos visuales o auditivos tales como *oír una sinfonía* o *ver a mi hermano* y cosas por el estilo". ¿Esto no quiere decir que fenómenos como el dolor, en realidad son epifenómenos? Si los dolores son fenómenos causados por procesos neurofisiológicos, entonces ¿qué son estos fenómenos y cómo encajan con nuestra ontología? Para responder estas preguntas es necesario entender ciertas confusiones que se tienen respecto a la naturaleza de la causación. Se tiende a pensar que todo fenómeno debe adscribirse a un modelo de causación en el que si A causa B, entonces se puede identificar dos eventos discretos, en el que A es la causa, con precedencia temporal, y B es el efecto, que sucede temporalmente a la causa A.

Todo sistema tiene micro-propiedades y macro-propiedades. En algunos casos, las macro-propiedades se pueden explicar a partir de las micro-propiedades; sin embargo, no todas las macro-propiedades tienen explicación causal en términos de micro-propiedades. Para la re-

lación causal entre los eventos mentales y el cerebro, el modelo se puede expresar diciendo lo siguiente:

> La característica de superficie F es causada por el comportamiento de los micro-elementos M, y que al mismo tiempo está realizada 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. […] Los fenómenos mentales son causados por los 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 (Searle 1995, p. 432).

Esto quiere decir que no hay limitaciones filosóficas, lógicas o metafísicas en explicar la relación entre la mente y el cerebro:

> Nada hay más común en la naturaleza que el que los rasgos de superficie de un fenómeno sean causados por micro-estructura y realizados en ella; y esa son, exactamente, las relaciones que son exhibidas por la relación de la mente con el cerebro. Las características intrínsecamente mentales del universo son las características físicas de alto nivel de los cerebros (Searle 1995, p. 434).

Para entender el problema de la consciencia, se puede apelar a la similitud en nuestra manera de entender *la vida* (Searle, 1995). Hace algún tiempo nos parecía misterioso que nuestro cuerpo estuviera vivo, por lo cual apelábamos a un *élan vital*. No obstante, con el paso del tiempo comprendimos los procesos biológicos que explicaban aquello que llamamos vida. Lo mismo sucede con nuestra mente, de manera que el misterio se disipa a medida que comprendemos los procesos:

> Todavía no entendemos los procesos completamente, pero entendemos el carácter de los procesos, entendemos que hay ciertos procesos electroquímicos que acaecen en las relaciones entre neuronas o entre módulos de neuronas y quizás otras características del cerebro, y que esos procesos son causalmente responsables de consciencia (Searle 1995, p. 434).

Con respecto al problema de la intencionalidad, Searle asegura que sucede algo similar al problema de la consciencia. Las actitudes proposicionales que implican intencionalidad se originan en procesos que se suceden en el cerebro. Esto no quiere decir que no sean procesos asombrosos, tan solo quiere decir que son *procesos* igual de asombrosos

a la fuerza gravitacional o las dimensiones de la Vía Láctea. Por ejemplo, Searle señala que algunos tipos de sed son causados por acción de la hormona peptídica angiotendina II en el hipotálamo. A su vez, la hormona peptídica está regulada y sintetizada por la renina, que se secreta en los riñones. De esta manera, tener sed implica una actitud proposicional, a saber, el deseo de beber agua y, a la vez, es causado por procesos fisiológicos (Searle, 1995).

Con respecto a la subjetividad, Searle asegura que desde el S. XVII nuestra visión del mundo ha estado dominada por la idea de objetividad total. En esta medida, se tiende a pensar que la realidad es aquello accesible a cualquier observador. La respuesta a la pregunta que se plantea con respecto la subjetividad consiste en asegurar que:

> Es un error suponer que la definición de la realidad deba excluir la subjetividad. Si ciencia es el nombre del conjunto de verdades objetivas y sistemáticas que podemos enunciar acerca del mundo, entonces la existencia de la subjetividad es un hecho científico tan objetivo como cualquier otro.[…] Si el hecho de la subjetividad va en contra de cierta definición de ciencia, entonces habría que abandonar la definición y no el hecho (Searle 1995, p. 436).

Por otra parte, el problema de la causación intencional se soluciona aceptándose que los estados mentales no son etéreos, gaseosos o epifenómenos:

> […] Sus propiedades lógicas e intencionales están sólidamente fundadas en sus propiedades causales en el cerebro. Los estados mentales pueden causar la conducta mediante el proceso causal ordinario, porque son estados físicos del cerebro. Ellos tienen un nivel elevado y un nivel bajo de descripción, y cada nivel es causalmente real. […] La consciencia, por ejemplo, es una propiedad causal real del cerebro y no algo epifenoménico. Mi intento consciente de realizar una acción, tal como levantar mi brazo, causa el movimiento del brazo (Searle 1995, p. 436).

Ahora bien, si por dualismo de propiedades se entiende que en el mundo hay rasgos físicos que son mentales, como la consciencia, y rasgos físicos que no son mentales, como el peso del cerebro, entonces el enfoque de Searle puede calificarse de esta manera. Sin embargo, el dualismo de propiedades parece implicar que en el mundo hay dos y sólo dos tipos de propiedades, y el enfoque de Searle no intenta ex-

presar esto. Searle intenta mostrar que lo mental y lo físico no están opuestos entre sí porque las propiedades mentales sólo son una clase de propiedades físicas "y las propiedades físicas se oponen correctamente no a las propiedades mentales sino a rasgos tales como las propiedades lógicas y las propiedades éticas, por ejemplo" (Searle 1995, p. 437). De esta manera, Searle considera más apropiado llamar a su postura polismo de propiedades.

Con respecto a clasificar su postura como emergentismo, Searle asegura que no considera apropiado denominarla de esta manera. Si se piensa que emergentismo implica algo misterioso en las propiedades, que desbordan las ciencias físicas o biológicas, entonces las propiedades mentales no son emergentes. Por otra parte, la superveniencia de lo mental en lo físico "dice que no puede haber diferencias mentales sin las correspondientes diferencias físicas: si en dos momentos diferentes un sistema está en dos estados mentales diferentes, entonces en esos dos momentos tiene que tener propiedades físicas diferentes" (Searle 1995, p. 439). Así, la superveniencia de lo mental en lo físico, es tan sólo un caso de superveniencia de las propiedades de alto nivel, de las propiedades de bajo nivel.

Algunas réplicas al experimento mental de la habitación china

La presentación del experimento mental de la habitación china que se encuentra en el artículo titulado Mentes, cerebros y programas, incluye el tratamiento que Searle le da a algunas réplicas propuestas en su contra. Aunque el presente escrito no tiene el propósito de centrarse en dichas réplicas, estas serán revisadas porque permiten comprender la postura de Searle con respecto a sus oponentes. En lo siguiente, se tratarán aquellas réplicas en las que se centra el presente escrito, no obstante, a continuación se presentan otras refutaciones al experimento mental de la habitación china.

La réplica de los sistemas

Según esta réplica, se asegura que, si bien es cierto que el individuo que está encerrado en la habitación no entiende ningún símbolo chino, en verdad él hace parte de un sistema más amplio y completo

que sí entiende el relato. Esto quiere decir que el individuo, en tanto parte del sistema total conformado por la habitación, los códigos y él mismo, no entiende y no tiene por qué entender alguno de los símbolos que procesa.

Para responder a esta objeción, Searle concede la posibilidad de que el individuo "absorba" todos los elementos del sistema:

> Memoriza las reglas del libro y los bancos de datos de los símbolos chinos y hace todo los cálculos mentalmente. Luego, incorpora todo el sistema. No hay absolutamente nada en el sistema que no haya incorporado. Incluso podemos deshacernos de la habitación y suponer que trabaja al aire libre (Searle 1990, p. 88).

Ahora bien, después de que el individuo que se encuentra en la habitación ha incorporado a sí mismo todos los elementos, él solo, en tanto individuo, comenzará a ser la totalidad del sistema. No habrá un sistema total fragmentado en partes que deban o no deban entender el chino, porque ahora la totalidad del sistema consiste en una sola parte, a saber, el individuo. Como el individuo sigue sin entender el chino, entonces se puede asegurar que la totalidad del sistema no entiende chino.

Según algunas versiones de la teoría de sistemas, "el hombre en cuanto sistema de manipulación de símbolos formales" (Searle 1990, p. 88), sí entiende chino, de manera que "el subsistema del hombre que constituye el sistema de manipulación de símbolos formales para el chino no debe confundirse con el subsistema para el inglés" (Searle 1990, p. 88). A esto, Searle responde que ningún subsistema tiene relación con el otro. Por una parte, el subsistema que comprende inglés, dado que el inglés es la lengua nativa del individuo, entiende el significado de las palabras y, por lo tanto, puede representarse lo que ellas quieren decir. Por otra parte, el subsistema que entiende chino, dado que el chino no es la lengua nativa del individuo, no entiende el significado de los símbolos; para este subsistema los símbolos carecen completamente de significado y sólo pueden adquirir un valor sintáctico, en la medida en que lo máximo que se puede entender con respecto a un símbolo determinado es que *va antes y después* de otros símbolos. Todo lo que el subsistema chino sabe es que "los distintos símbolos formales ingresan por un extremo, se manipulan conforme a reglas escritas en inglés y otros símbolos salen por el otro extremo" (Searle 1990, p. 89). Según Searle, el argumento incluido en el experimento mental de la habitación

china pretende mostrar que la manipulación formal de símbolos no es suficiente para permitir la semántica del lenguaje y, de esta manera, el subsistema chino -suponiéndose que se pueda hablar en términos de sistemas y subsistemas- efectivamente continúa sin entender el significado de las palabras que están escritas en caracteres chinos.

La postulación de subsistemas no tiene mayor efecto en contra del argumento, pues el subsistema chino "no posee nada que se parezca siquiera remotamente a lo que tiene el hombre (o subsistema) que habla inglés" (Searle 1990, p. 89). Nótese que ambos subsistemas están en capacidad de superar una prueba de Turing y, no obstante, únicamente el subsistema inglés comprende las palabras que procesa. El hecho de que un programa o una máquina supere el test de Turing implica que dicho programa o máquina está comprendiendo en el mismo sentido en el que lo hace la mente humana; en conclusión, según Searle, la presente réplica simplemente pretende insistir, sin argumentos, en "que el sistema debe entender chino" (Searle 1990, p. 89).

Ahora bien, la idea de proponer subsistemas, según Searle, conduce a conclusiones absurdas. Si se asegura que la cognición y la comprensión del lenguaje se sustentan solamente en la base de que hay una entrada y una salida de información y un programa que debe procesar dicha información, entonces se sigue que el funcionamiento de nuestro estomago, por ejemplo, el cual puede describirse en términos de entrada, procesamiento y salida de información, también sería un órgano de cognición y comprendería la información que procesa. De igual manera, el hígado, el páncreas y los demás órganos de nuestro cuerpo, que pueden describirse en términos de entrada, procesamiento y salida de información, serían órganos de comprensión y cognición, porque en todos estos casos, entre la entrada y salida de información podría postularse programas de procesamiento. De esta manera, al parecer, es una dificultad para la teoría de los sistemas señalar características intrínsecas a los sistemas, que permitan distinguir un sistema mental de un sistema no mental. Esto es una grave complicación porque, si la IAF pretende fines psicológicos, entonces, al menos, se debe distinguir aquellos sistemas físicos con capacidades mentales de aquellos sistemas físicos sin capacidades mentales. Si esta distinción intrínseca no se logra y las características de los sistemas dependen únicamente del espectador, entonces un espectador podría tratar "a los huracanes como mentales si así lo deseara"(Searle 1990, p. 90). Según Searle, esta incapacidad para distinguir intrínsecamente un sistema mental de

un sistema no mental lleva a confusiones graves como la postura de McCarthy (1977), según la cual es correcto asegurar que máquinas como los termostatos tienen creencias. Esto es grave porque:

> El estudio de la mente parte de hechos tales como que los humanos tienen creencias, mientras que los termostatos, teléfonos y sumadoras, no. Si se obtiene una teoría que niegue este punto, se habrá logrado un contraejemplo de la teoría y esta será falsa. […] Piense intensamente durante un minuto qué se necesitaría para establecer que ese trozo de metal en la pared tiene creencias reales, creencias con direccionalidad, con contenido proposicional u condiciones de satisfacción; creencias que tuvieran la posibilidad de ser sólidas o débiles; creencias nerviosas, ansiosas o seguras; creencias dogmáticas, racionales o supersticiosas, fes ciegas o cogniciones titubeantes, cualquier tipo de creencias. El termostato no es candidato para ello, como tampoco lo son el estomago, e hígado, la sumadora o el teléfono. Sin embargo, puesto que estamos considerando la idea seriamente, obsérvese que su verdad sería fatal para la afirmación de que la IAF es una ciencia de la mente (Searle 1990, p. 91).

La réplica del robot

Supóngase que el programa de comprensión de Shank y Abelson se coloca en un robot antropomorfo, con cámaras que le permiten percibir el exterior; además, supóngase que todo lo que el robot percibe, además de los movimientos que hace su cuerpo, está controlado por un "cerebro de computadora […]; a diferencia de la computadora de Shank, este robot sería capaz de una genuina comprensión y de otros estados mentales" (Searle 1990, p. 91).

Para responder, Searle (1990) asegura que esta réplica acepta que la cognición no solamente depende de la manipulación formal de símbolos, sino que implica cierta relación causal con los estímulos precedentes del mundo externo. Ahora bien, la constitución física del robot no implica que aquel centro de procesamiento entienda más de lo que entiende el individuo originalmente insertado en la habitación china. Así pues, supóngase que los fajos de símbolos que este individuo recibe, sin que él lo sepa, contienen información recibida de cámaras de televisión y que los fajos que él arroja sirven para accionar los movimientos del robot; aun en este caso, el individuo continúa sin entender algunos de los símbolos chinos que procesa:

> Yo soy el homúnculo del robot, pero a diferencia de los homúnculos tradicionales, ignoro qué está sucediendo; no comprendo nada excepto las reglas para la manipulación de los símbolos. Ahora bien, cabe mencionar en este caso que el robot carece por completo de estados intencionales, pues sencillamente se activa como resultado de su sistema de circuitos eléctricos y de su programa (Searle 1990, p. 92).

La réplica del simulador de cerebros

Según esta réplica, no debe suponerse que se construye un robot ni que se está ejecutando un programa, sino que se crea una simulación de "la secuencia real de emisiones neuronales en las sinapsis del cerebro de un hablante del chino cuando comprende relatos en chino y ofrece respuestas acerca de éstos" (Searle 1990, p. 92). De esta manera, debe concebirse la posibilidad de crear una simulación de un cerebro chino, que al recibir la información de los relatos, como entrada, simula la estructura formal de un cerebro chino al recibir este tipo de información. Incluso, Searle (1990) concibe la posibilidad de que se ejecuten varios programas en paralelo, de manera similar a como se supone que funciona un cerebro humano. Si en este nivel se niega que el simulador de cerebros está entendiendo el relato, entonces también sería necesario asegurar que un cerebro chino no entiende relatos cuando se le dan para que los procese. No obstante, cabe preguntar "en el nivel de la sinapsis, ¿qué sería o podría ser diferente entre el programa de la computadora y el programa del cerebro chino?" (Searle 1990, p. 93).

Supóngase que en lugar de un individuo correlacionando símbolos según instrucciones, se tiene un individuo que manipula válvulas según las instrucciones y los fajos que le entregan; "¿dónde está la comprensión en este sistema?" (Searle 1990, p. 93). La respuesta de Searle (1990) consiste en asegurar que en este nivel de la simulación no puede haber comprensión porque el simulador únicamente simula estructuras formales de los estados de encendidos neuronales sinápticos, cuando lo realmente importante a simular con respecto al cerebro son sus propiedades causales. Mientras no se logre simular la capacidad causal del cerebro para producir estados intencionales, no se podrá hallar un nivel de comprensión en la simulación.

La réplica de la combinación

Esta réplica consiste en asegurar que, en el caso de que contáramos con una serie de elementos combinados al interior de un mismo sistema, como por ejemplo, un robot con un cerebro artificial ubicado en su cabeza, con sinapsis similares a las humanas y un comportamiento idéntico al humano, "tendríamos que atribuirle intencionalidad al sistema" (Searle 1990, p. 94).

Según Searle (1990), esta réplica no ayuda a reforzar la postura de la IAF pues, según ésta última, lo importante y verdaderamente relevante con respecto a lo mental es la manipulación de símbolos físicos; sin embargo, la atribución de intencionalidad que se hace al robot nada tiene que ver con la verdadera manipulación formal que haga, sino con la suposición de que, dado que se comporta como humano, entonces debe ser poseedor de intencionalidad. Nuestra atribución de intencionalidad, en este caso, parte del hecho de que el robot se comporta de manera similar a nosotros, y no parte del hecho de que sus estados mentales consisten en una manipulación sintáctica. Si aceptamos que el robot tiene estados intencionales porque se comporta de manera similar a nosotros, entonces esto en nada apoya la inteligencia artificial fuerte porque no estamos atribuyendo estados mentales a partir de la manipulación formal que se da en el cerebro del robot. Así, Searle (1990) está casi convencido de que, omitiendo las consideraciones acerca del comportamiento del robot, si nos fijáramos únicamente en el hecho de que éste ejecuta un programa de manipulación formal, estaríamos más persuadidos a no atribuirle intencionalidad al robot:

> El manejo de símbolos formales continúa, los datos de entrada y de salida se corresponden correctamente, pero el único foco real de intencionalidad es el hombre y él no sabe nada de los estados intencionales pertinentes; no ve, por ejemplo, lo que entra en los ojos del robot y no comprende ninguna de las observaciones hechas al robot o por el robot (Searle 1990, p. 95).

Adicionalmente, cabe preguntar si esta atribución de intencionalidad no es similar a la que hacemos acerca de aquellos seres en los que suponemos la presencia de estados intencionales. Searle asegura que en el caso de otros seres, como los simios, por ejemplo, atribuimos intencionalidad porque no vemos otra manera de darle sentido

a su comportamiento sin suponer algunos estados mentales y porque vemos que su constitución es similar a la nuestra. No obstante, si supiéramos que el comportamiento de aquellos animales obedece únicamente a la ejecución de un programa formal, nos abstendríamos de aquella suposición de intencionalidad (Searle 1990, p. 95).

La réplica de las otras mentes

Según esta réplica, solamente podemos asegurar que las otras personas comprenden chino, o poseen cualquier tipo de cognición, en virtud de su comportamiento. Dado que el programa puede superar la prueba de conducta al igual que puede hacerlo cualquier otra persona, entonces, si le atribuimos cognición a otra persona, también debemos atribuirla al programa.

Para responder a esta cuestión, Searle (1990) asegura que lo importante no es cómo saber que otras personas tienen estados cognitivos, sino saber qué se les atribuye cuando se les atribuyen estados cognitivos. Ciertamente, esta atribución no es de procesos de cómputo y los resultados de estos procesos.

La réplica de las mansiones múltiples

Según esta réplica, Searle solamente se refiere a computadoras analógicas y digitales, cuando en realidad estas sólo son las formas actuales de la tecnología. Es muy probable que en el futuro, sean cuales sean los procesos causales de la cognición, estos podrán ser construidos; con la construcción artificial de artefactos que posean los procesos causales del cerebro se dará paso a la inteligencia artificial.

Searle señala que esta objeción consiste en una trivialización del proyecto de la inteligencia artificial fuerte, pues lo convierte en "cualquier cosa que produzca y explique artificialmente la cognición" (Searle 1990, p. 96). Según Searle, la tesis original de la inteligencia artificial fuerte, consistía en asegurar que los estados mentales son básicamente procesos computacionales "de elementos definidos formalmente" (Searle 1990, p. 96); pero si esta tesis original se reformula, entonces las objeciones pierden validez.

Las nuevas réplicas

Después de las réplicas expuestas, varios autores han propuesto otras que también se oponen al argumento contenido en el experimento mental de la habitación china. Algunas de estas críticas se presentan a continuación.

Las escalas descuidadas

Moural (2003) denomina así a la crítica proveniente de Paul y Patricia Churchanld (1990), en el artículo titulado *Could a Machina Think?* En este artículo se invita a plantear una situación en que se argumenta en contra de John Maxwell para mostrar que el electromagnetismo y las ondas eléctricas son entidades radicalmente distintas. Para dicha demostración, los oponentes de Maxwell podrían exigir que una persona iluminara una habitación oscura con una barra cargada electromagnéticamente; es de esperarse que una habitación que se pretenda iluminar de esta manera permanecerá completamente oscura. Aquí entra a formar parte de la argumentación la importancia de la velocidad de procesamiento de los símbolos dentro de la habitación; cualquier persona reconoce que nadie piensa a la velocidad con que la persona dentro de la habitación está procesando; por lo tanto, nunca se podrá alcanzar verdadero pensamiento en virtud de ejecutar un programa de manera tan lenta como lo hace la persona dentro de la habitación. Sin embargo, Searle asegura que la velocidad de procesamiento es irrelevante para dicha cuestión y que "un pensador lento seguiría siendo un autentico pensador" (Churchland y Churchland 1990, p. 28).

Churchland y Churchland (1990) aseguran que el axioma del que Searle parte, a saber, que la sintaxis no es suficiente para la semántica, no queda completamente demostrado con el experimento mental de la habitación china y, para sustentar su propuesta, presentan el paralelo con la luz y el electromagnetismo. Teniendo en cuenta la siguiente argumentación, se puede intentar refutar la hipótesis de Maxwell:

> Axioma 1. La electricidad y el magnetismo son fuerzas. Axioma 2. La propiedad esencial de la luz es la luminiscencia. Axioma 3. Las fuerzas, por sí mismas, no son constitutivas ni suficientes para la luminiscencia. Conclusión. La electricidad y el magnetismo no son constitutivas ni suficientes para la luminiscencia (Churchland y Churchland 1990, p. 29).

Además, si se le exige a Maxwell que una persona ilumine artificialmente una habitación oscura agitando, conforme su teoría, muy rápidamente un imán cargado eléctricamente, fácilmente podría llegarse a la conclusión de que "resulta inconcebible que se pudiera hacer surgir auténtica luminancia por mero movimiento de fuerzas de acá para allá" (Churchland y Churchland 1990, p. 29). Ante esta objeción, lo primero que Maxwell puede argumentar es que la situación propuesta en *la habitación luminosa* consiste en una presentación equivocada del fenómeno de luminiscencia "porque la frecuencia de oscilación del imán es de una pequeñez ridícula, unas 10^{13} veces demasiado pequeña." (Churchland y Churchland 1990, p. 29). Según Churchland y Churchland (1990), Maxwell incluso podría asegurar que la habitación, aunque oscura, se encuentra cargada de *luminancia*, aunque en una medida imperceptible; no obstante, con el ambiente intelectual de 1860, cualquiera se habría reído de la idea de que había luz en la habitación: "¡De luminancia nada, señor Maxwell! ¡Esto está más oscuro que la pez!" (Churchland y Churchland 1990, p. 29). La respuesta más apropiada de Maxwell es que el axioma 3 es incorrecto, que el experimento de la habitación luminosa no dice nada interesante o relevante acerca del fenómeno en cuestión y que, para responder al problema de la luminancia artificial, hace falta investigación para determinar las condiciones necesarias para que las ondas electromagnéticas puedan producir luz. Según Churchland y Churchland (1990), esta sería la respuesta que la inteligencia artificial tendría que dar a Searle; esta y ninguna otra.

La réplica del conexionismo

Según Moural (2003), esta réplica consiste en asegurar que los cerebros pueden ser computadores, pero computadores radicalmente diferentes a los concebidos por la IA clásica. Los ordenadores clásicos son digitales y están completamente programados, mientras que el cerebro no. Para Los Churchland, las verdaderas dificultades que enfrenta la IA y la imposibilidad de que las máquinas tradicionales lleguen a constituir verdadera inteligencia y consciencia, se fundamentan en los fallos que resultan de no enfrentar el problema desde una interpretación de los sistemas nerviosos como máquinas con procesamiento en paralelo. Otro aspecto que se debe tener en cuenta es que el funcionamiento de la neurona se encuentra enmarcado por un procesamiento de tipo analógico y no digital, en la medida en que los

índices de disparo varían constantemente en función de las entradas sensoriales. Según Los Churchland, aunque Searle conoce la posibilidad de los circuitos en paralelo, él asegura que en dicha instanciación también hay falencias semánticas. Para sustentar su idea, Searle propone que en lugar de pensar en una habitación china, se piense en un gimnasio chino, en el que muchas personas están organizadas en redes en paralelo ejecutando los mismos cómputos que ejecutaba la persona encerrada en la habitación china. Ante esto, Los Churchland aseguran que sería necesario contar con alrededor de 10^{14} personas porque un cerebro humano tiene alrededor de 10^{11} neuronas, cada una de ellas con un promedio de 10^{3} conexiones; esto representa la población humana de 10.000 tierras. Ahora bien, dicho modelo cósmico de personas procesando información podría representar un gran cerebro lento pero funcional. Esto no implica que se pueda asegurar que, de manera inmediata, al construir el modelo se obtenga verdadero pensamiento; los Churchland reconocen que la teoría del procesamiento en paralelo podría no ser correcta; sin embargo, tampoco hay *a priori* una garantía de que esto no constituiría pensamiento.

Según Moural (2003), Searle no reconoce correctamente el tipo de ordenadores que Los Churchland tienen en mente; él simplemente concibe la posibilidad de alterar la estructura original de las máquinas digitales tradicionales, imponiendo ejecuciones paralelas de varios algoritmos simultáneos; para esto propone la idea del gimnasio chino. Sin embargo, según Moural, la propuesta de Searle es desafortunada porque sigue siendo posible que un sólo hombre procese y obtenga el *output* total de la habitación china:

> Uno podría llamar a este escenario Gripa China: todos los operadores del gimnasio excepto uno, se quedan en casa con la gripa, y el que se queda en el gimnasio tiene que hacer todo el trabajo hasta que los otros se recuperen (Moural 2003, p. 228).

Sin embargo, el argumento original de Searle, según Moural, no se aplica de manera idéntica a las posibilidades conexionistas por el hecho de que también puedan ser simuladas en computadores:

> [...] No es una parte del argumento de Searle que si algo se puede procesar en un computador como dato, debe carecer de semánticas (piénsese, por ejemplo, en llamadas procesadas digitalmente) (Moural 2003, p. 228).

La réplica de la instanciación y la réplica de una mente más

Las siguientes son maneras de refutar la idea de que se puede tener una instanciación de un programa sin que ello implique la presencia de un estado mental:

> (1) Negar la ausencia del estado mental requerido, (2) negar que el programa que Searle considera es el correcto, y (3) negar que la configuración de la habitación china instancia el programa (Moural 2003, p. 228).

Fodor pretende sustentar la tercera propuesta mostrando que la configuración que propone Searle no es, en estricto sentido, una máquina de Turing que instancia de manera correcta un programa. Según Moural (2003), para Fodor, Searle usa "un concepto débil de instanciación, basado únicamente en una correlación de la secuencia de estados entre el programa y su instancia"(Moural 2003, p. 229). Así, se requiere un concepto de instanciación en el que intervengan causas eficientes; sin embargo, el tipo de causas eficientes, según Moural (2003), no queda claramente especificado por Fodor. Para Searle, una causa suficiente es la persona que se encuentra dentro de la habitación y, de esta manera, la réplica de Fodor deja intacto el planteamiento de Searle.

Según Moural (2003), Chalmers retomó esta cuestión mediante una versión mejorada de la réplica de los sistemas, al asegurar que los estados mentales de la persona dentro de la habitación son irrelevantes para la cuestión principal. Lo que es relevante para el sistema son las relaciones dinámicas que se pueden encontrar entre los símbolos: "si puede haber algo entendiendo, producido por la ejecución de un programa, no es el entendimiento en la mente consciente del operador de la habitación china (Moural 2003, p. 229). Según Moural (2003), para Chalmers es claro que, si se asume que lo verdaderamente importante es el entendimiento del operador de la habitación, entonces desde el principio de la argumentación de Searle se está asegurando que no se encontrará consciencia ni inteligencia; esto, porque en el computador no hay nada que corresponda a la persona que se encuentra dentro de la habitación. Con base en la idea de Chalmers, se puede decir que cuando Searle asegura que él -en tanto sistema- no está comprendiendo nada acerca de los símbolos que procesa, en realidad está formulando una intuición y no está diciendo

algo de lo que pueda estar seguro. En esta medida, Searle da un paso injustificado de una conclusión precedida por *"me parece obvio que…"* a una conclusión precedida por *"yo no entiendo nada"*.

El nuevo argumento

Las mentes no son programas

Moural (2003) denomina "El nuevo argumento" a la idea, desarrollada por Searle en las presentaciones más recientes del experimento mental de la habitación china. Una de las exposiciones más recientes del argumento es la que aparece en el libro *The Mystery of Consciousness*, en la que Searle (1997b) reconoce que su argumento de la habitación china es tan simple, que es embarazoso tener que repetirlo. Según Searle (1997b), el argumento consiste en tres simples pasos: "1. Los programas son enteramente sintácticos. 2. La mente tiene semántica. 3. La sintaxis no es igual a, o por sí misma no es suficiente para, la semántica. Por lo tanto, las mentes no son programas" (Searle 1997b, p. 11). Según Searle (1997b), el primer paso simplemente se refiere al hecho de que el programa de una máquina de Turing está compuesto de entidades sintácticas, es decir, puras reglas de manipulación de símbolos, "y el programa implementado […] consiste enteramente en estas manipulaciones puramente sintácticas" (Searle 1997b, p. 12). Ahora bien, un postulado importante del funcionalismo que enmarca la IAF es que lo verdaderamente importante para la constitución mental es el programa que se ejecute y no el medio en que dicha ejecución se esté realizando; así, el programa en ejecución no se refiere ni tiene algo que ver con el medio en que se está ejecutando. El paso (2.) es bastante obvio para Searle (1997b) porque nos recuerda algo que todos sabemos: cuando pensamos en símbolos, tenemos que saber lo que esos símbolos significan; "por esto es que puedo pensar en inglés pero no en chino" (Searle 1997b, p. 12). En la mente no solamente hay símbolos sin interpretación que se procesan mediante la ejecución del programa que indica cómo computarlos, sino que también hay contenidos mentales o semánticos. El paso (3.) establece lo que el experimento mental de la habitación china ilustra: las manipulaciones formales no son suficientes para constituir, ni poseen, contenido semántico: "no importa qué tan bien el sistema pueda imitar la conducta de alguien que realmente entiende, ni qué tan compleja sea la manipulación formal" (Searle

1997b, p. 12), no se puede obtener semántica como resultado de manipulaciones formales. De esta manera, el experimento mental de la habitación china tiene el propósito de refutar la idea de que un programa implementado, por sí mismo, puede constituir una mente.

La sintaxis es relativa al observador

Adicional a las tres premisas que constituyen el argumento original de la habitación china, según Searle (1997b), es importante resaltar el hecho de que la sintaxis no es intrínseca a ningún sistema físico y, por lo tanto, no está definida en términos físicos. La argumentación desarrollada a partir de dicha idea consiste en lo siguiente: "(1) la sintaxis es relativa al observador y (2) todas las características de la computación, en tanto computación, son exclusivamente sintácticas" (Moural 2003, p. 238).

Las personas pueden interpretar como símbolo cualquier objeto físico y esto manifiesta el hecho de que ningún símbolo es símbolo en virtud de que sea un objeto físico; los símbolos y la sintaxis son el resultado de la interpretación de los observadores y no de los objetos en sí mismos. Así, ningún patrón puede interpretarse como computación si no es porque alguien está asignando una interpretación de dicho patrón como computación. Esto quiere decir que Searle no rechaza la idea de que un computador físico específico pueda poseer una mente, pues el cerebro, que es una máquina biológica, en algunos casos computa. Por ejemplo, cuando se ejecuta una operación sencilla como una suma, según Searle, se está realizando una computación, en estricto sentido. Sin embargo, como la tesis fundamental del funcionalismo deja de lado la configuración física del sistema en que se ejecuta el programa, entonces la cuestión de si un determinado sistema físico -así sea una máquina biológica- puede llegar a pensar es irrelevante.

La computación no es intrínseca a la física sino que la primera resulta de una interpretación hecha por el observador del sistema. En el mundo se puede encontrar objetos que son relativos al observador y otros que son independientes del observador; esta es la distinción que Searle (1997a) desarrolla en *La construcción de la realidad social.* Para Searle (1997a), hay rasgos del mundo que son intrínsecos y hay rasgos que son relativos al observador:

> Es, por ejemplo, un rasgo intrínseco del objeto que está frente a mí que tiene una determinada masa y una determinada composición química […]. Pero también se puede decir con verdad del mismo objeto que es un destornillador. Cuando lo describo como un destornillador, estoy determinando un rasgo del objeto que es relativo al observador o al usuario. Es un destornillador sólo porque la gente lo usa como (o lo ha hecho para el propósito de servir como, o lo ve como) un destornillador (Searle 1997a, p. 29).

Así, surge la pregunta: ¿Es la computación un rasgo intrínseco de un sistema físico como el cerebro, o es un rasgo que resulta de una interpretación del cerebro como un computador? Según Searle (1997b), hay casos limitados en los que, en estricto sentido, los humanos computan, "por ejemplo, ellos computan la suma de 2+2 y obtienen 4" (Searle 1997b, p. 15). En estos casos, se puede decir que la computación es independiente del observador "en el sentido en que no se requiere un observador externo que los interprete como computando para que ellos estén computando" (Searle 1997b, p. 15). Ahora bien, ¿Qué sucede con los computadores comerciales que todos conocemos? ¿Qué parte de la física o la química de sus impulsos eléctricos permiten la constitución de símbolos? Según Searle (1997b, p. 15), no hay nada, ni físico ni químico que convierte a dichos impulsos en símbolos porque *símbolo* o *sintaxis* no se refiere a un rasgo intrínseco como *electrón* o *placa tectónica*. Así, la siguiente es la respuesta de Searle a la pregunta ¿El cerebro es intrínsecamente un computador digital?: No, porque "algo es un computador solamente relativo a las asignaciones de una interpretación computacional" (Searle 1997b, p. 16). Ahora bien, esto no quiere decir que no sea posible asignar interpretaciones computacionales al cerebro pues, de hecho, se le puede asignar interpretaciones computacionales a cualquier proceso. La principal consecuencia de esta idea es que:

> No se puede descubrir procesos computacionales en la naturaleza, independientemente de la interpretación humana porque cualquier proceso físico que se pueda encontrar es computacional solamente relativo a alguna interpretación (Searle 1997b, p. 16).

Searle asegura que el funcionalismo, que se jacta de ser una postura materialista, no es "lo suficientemente materialista" (Searle 1997b, p. 16); el cerebro es una máquina orgánica y sus procesos son los de una máquina orgánica, pero la computación no es un rasgo in-

trínseco al cerebro, es decir, no es un proceso de esta máquina orgánica, como sí lo es el índice de disparo neuronal: "computación es un proceso matemático abstracto que existe sólo de manera relativa a la consciencia de los observadores y los intérpretes" (Searle 1997b, p. 17). De esta manera, aquello que se jacta de materialista, se fundamenta en una abstracción matemática. Según Searle (1997b), el presente argumento es mucho más profundo que el inicialmente planteado en la habitación china porque, a parte de mostrar que la semántica no es intrínseca a la sintaxis, muestra que la sintaxis no es intrínseca a la física.

Al finalizar la presentación del argumento en el primer capítulo de *The Mystery of Consciousness*, Searle (1997b) asegura que la consciencia se puede interpretar como una propiedad emergente del cerebro que sólo puede ser explicada por el comportamiento y las interacciones que hay entre los elementos que conforman el sistema. Sin embargo, esto no quiere decir que dicha investigación se pueda adelantar observando el comportamiento y las características de cada elemento por separado: "la liquidez del agua es un buen ejemplo: la conducta de las moléculas de H_2O explican la liquidez pero las moléculas individuales no son líquidas" (Searle 1997b, p. 18).

Poderes Causales

Según Moural (2003), el argumento de Searle se dirige, en primera instancia, a resaltar los poderes causales del cerebro, en cuanto a su capacidad de constituir intencionalidad y, en segunda instancia, a señalar que dichos poderes causales no están presentes en un programa por sí solo. Searle siempre señala que la premisa fundamental de la IAF es la idea de que una mente puede constituirse a partir de, *únicamente*, la instanciación de un programa y, por lo tanto, el soporte físico sobre el que se ejecute el programa es irrelevante. De esta manera, los partidarios de la IAF deben comprobar que un programa, por sí solo, cuenta con los poderes causales para constituir mentes. Searle no pretende refutar la idea de que una máquina puede causar estados mentales, sino la idea de que el programa por sí solo puede hacerlo. Esto implica que el programa se debe interpretar como aislado del soporte físico -aunque siempre realizándose en él- y, no obstante, con poderes causales. Así, la versión de IAF que Searle pretende atacar está completamente comprometida con la separación de poderes causales -de *hardware* y de *software*- y, de esta manera, debe reconocer poderes causales en las unidades sintácticas.

Lo que para Searle es una premisa central de la IAF riñe con una premisa fundamental de su postura, a saber, que los cerebros causan mentes. Searle reconoce que puede haber máquinas que causen mentes, pero solamente máquinas que tengan una configuración idéntica a la cerebral porque, en estricto sentido, la única máquina que cuenta con los poderes causales necesarios para construir estados mentales, es la máquina orgánica que conocemos como cerebro. Por otra parte, si nos alejamos del soporte físico y suponemos la relevancia causal de, únicamente, los programas, como según Searle lo hace, caeremos en un error porque "los programas sin implementar son abstractos, sistemas formales que no tiene poderes causales de ningún tipo y, como tales, no puede producir mentes" (Moural 2003, p. 247).

Asignación de funciones

Para Searle (1997b) la posibilidad de explicaciones funcionales acerca de la mente humana queda atrapada en la cuestión de asignación relativa al observador. Searle concede la posibilidad de explicaciones funcionales en el nivel de partes brutas del cerebro pero, como es de esperarse, en estos niveles en los que se posibilita la explicación funcional, no hay consciencia, *ergo*, la explicación funcional no es útil para la consciencia o la intencionalidad. Según Searle (1997a), en los niveles en que podría concebirse una explicación funcional, esto es, en los niveles biológicos brutos, con descripciones igualmente brutas -el órgano X, hace Y-, no se puede hallar descripciones normativas porque estas solamente se posibilitan desde nuestro punto de vista. En estricto sentido, según Searle (1997a), ni el cerebro ni el corazón, ni ningún órgano, tiene funciones porque describir lo que el órgano hace es describir relaciones causales que, cuando comienzan a interpretarse como funciones, son el resultado de elementos normativos posibilitados por nuestro punto de vista:

> En lo atinente a nuestras experiencias normales de las partes inanimadas del mundo, hay que decir que no experimentamos las cosas como objetos materiales, y mucho menos como colecciones de moléculas. Ocurre más bien que experimentamos un mundo de sillas y mesas, de casas y automóviles, de salas de lectura, de pinturas, calles jardines, fincas, etc. Todos los términos que se acaban de usar entrañan criterios de evaluación que, bajo descripciones, son internos a los fenómenos en cuestión, pero no internos a las entidades bajo su descripción como "objetos materiales". Se puede incluso asignar funciones

> a fenómenos naturales, como ríos y árboles […]. Llegados a este punto, es importante darse cuenta de que las funciones nunca son intrínsecas a la física de ningún fenómeno, sino que son externamente asignadas por observadores y usuarios conscientes. En una palabra: las funciones nunca son intrínsecas sino relativas al observador. Incluso cuando descubrimos una función en la naturaleza, como cuando descubrimos la función del corazón, el descubrimiento consiste en el descubrimiento de los procesos causales junto con la asignación de una teleología a esos procesos (Searle 1997a, p. 33).

De esta manera, en biología, la explicación funcional no revela mucho acerca de la naturaleza del objeto pues, en realidad, se está hablando de algo creado por nosotros mismos; funcionalmente sólo podemos describir lo que el objeto hace y nada más; "no es realmente, después de todo, algo que deba ser tomado en serio" (Dennett 199, p. 661). En esta medida, sólo poseen verdadera función aquellos objetos que los humanos hemos diseñado para cumplir una función específica: unas tijeras y unas alas de avión poseen verdadera función. Sin embargo, las alas de un pájaro carecen de función; de manera que, como lo asegura Dennett (1999), si un biólogo sostiene que las alas del pájaro están adaptadas para volar y otro sostiene que están adaptadas para sostener plumas, en realidad, estamos frente a una cuestión sin respuesta pues ambos tienen la razón y, por lo tanto, ninguno.

Semántica

Según Moural (2003), Searle no puede escapar completamente a la posibilidad de que *el programa correcto* pueda constituir una mente, "porque *el programa correcto* está simplemente definido como siendo un programa del tipo de aquellos que producen mente" (Moural 2003, p. 248). De esta manera, la pregunta verdaderamente importante no es si *el programa correcto* puede constituir una mente, sino si existe tal tipo de programa. Según Moural (2003), la postura contra la que lucha Searle en este punto, puede establecerse de la siguiente manera: "por la sola ejecución de un programa formal, un sistema puede aumentar el acceso al significado de al menos algunos de los símbolos que operan" (Moural 2003, p. 248). Searle está convencido de que todos y cada uno de los símbolos que se manipulan con el programa que se ejecuta son completamente carentes de contenido semántico porque él, dentro de la habitación, no entiende nada acerca de los símbolos que manipula. Una respuesta a esta idea se desarrollará

en el tercer capítulo del presente libro, mostrando, desde Rapaport (1995a, 1995b, 2002), que la manipulación puramente formal sí permite acceso al significado -no necesariamente convencional- de los símbolos manipulados.

Según Moural (2003, p. 249), Searle entiende por sintaxis "relaciones entre unidades formales sin interpretación, especialmente con respecto a cómo pueden y cómo no pueden ser concatenadas en agregados lineales". Por semántica, Searle entiende "la propiedad de un objeto de significar o simbolizar algo más allá de él mismo en una forma que sea entendida (al menos por algunos) usuarios del objeto" (Moural 2003, p. 249). Partiendo de estas dos interpretaciones, Searle asegura que es una verdad conceptual la idea de que la sintaxis no es suficiente para la semántica. Esta verdad conceptual, sumada a la idea de que los programas operan únicamente en el nivel sintáctico, permite concluir que la implementación de un programa no constituye significado únicamente en virtud de su ejecución. Según Moural, lo importante al respecto es el tipo de operación que la persona dentro de la habitación ejecuta sobre los símbolos carentes de significado; cuando la persona dentro de la habitación lee una palabra en inglés, entiende su significado, pero cuando observa una cadena de símbolos chinos, no entiende dicho significado. Ahora bien, la conclusión de Searle es que ninguna manipulación que haga a partir de las instrucciones que entiende, sobre los símbolos carentes de significado, permitirá la constitución de significado, ¿Por qué?

> Porque las operaciones están definidas así: incluyen la identificación de una unidad en la cadena, relacionada con su identidad como un símbolo de un tipo predefinido, el reemplazo de una unidad por otra en la cadena, el movimiento a otra posición en la cadena, un cambio de estado interno, y eso es todo (Moural 2003, p. 250).

Como se puede observar, y como Moural (2003) lo resalta, no hay alguna asignación de significado que intervenga en el proceso de manipulación; por ejemplo, no se dice algo con respecto a la posibilidad de que las unidades sintácticas puedan *cargarse* de significado gracias a que el usuario del programa es una persona dotada de mente.

a). La IAF interpreta la mente como un programa. Una premisa fundamental para la IA., según Dennett (1999), es que un programa por sí solo, en ejecución, es suficiente para constituir una mente. En este orden de ideas, la ejecución del programa correcto, sin importar el sistema físico en que se ejecute, constituye una mente. Esto quiere decir que la mente es una manipulación puramente formal de símbolos.

b). Según Searle, una verdad indudable es que la sintaxis no es suficiente para la semántica. Si esto es así, entonces la ejecución de un programa, que solamente opera en el nivel sintáctico, nunca podrá constituir una mente porque ésta, además de poseer sintaxis posee semántica.

c). Para demostrar que la ejecución de un programa no puede constituir cognición o intencionalidad, Searle propone el experimento mental de la habitación china.

d). La mente posee Intencionalidad y cognición, no en virtud de la ejecución de un determinado programa y, por lo tanto, no en virtud de una manipulación puramente formal de símbolos, sino en virtud de los poderes causales de nuestro cerebro. Los poderes causales del cerebro generan estados Intencionales y solamente si dichos poderes causales son duplicados, se podrá duplicar la cognición y la Intencionalidad.

Capítulo 3:

Entendimiento semántico y entendimiento sintáctico

En el anterior capítulo se expusieron algunas réplicas al experimento mental de la habitación china; cada una con la respectiva respuesta de Searle. En el presente capítulo se ampliará una réplica que consiste en cuestionar y examinar lo que para Searle es una verdad indudable: que la sintaxis no es suficiente para la semántica. En la medida en que se examine aquello que para Searle es una verdad indudable, y se muestre que posiblemente no es tan indudable como parece, el resultado muestra un camino para la IAF. Debe recordarse que ninguna réplica parece contundente frente al experimento mental de la habitación china y, por este motivo, sigue contando con validez conceptual. Esto no quiere decir que sea tarea inútil buscar salidas conceptuales de la habitación y reivindicaciones de la IAF. Vale la pena señalar que las posibilidades de modelos computacionales que permitan una interpretación formalizadora de la semántica, no se limitan a las que se ampliarán en este capítulo. Aunque se menciona la propuesta de Rapaport (1995a, 1995b, 2002), relacionada con el entendimiento sintáctico y la semántica como correspondencia, Fodor también propone un modelo computacional y, por este motivo, en los aspectos relevantes se mencionarán las posibles relaciones entre el modelo de Rapaport y Fodor.

Remisión de dominios en el entendimiento semántico

Según Rapaport, el "entendimiento *semántico* es una correspondencia entre dos dominios" (Rapaport 1995b, p. 49); un dominio A y un dominio B, por ejemplo. Así, un agente cognitivo entiende uno de los dominios, en este caso el dominio A, por ejemplo, en términos del otro dominio, en este caso el dominio B. Si el dominio A se entiende en términos del dominio B, entonces, ¿En qué términos se entiende el dominio B? Este dominio B se entiende "recursivamente" en términos de otro dominio, por ejemplo, otro dominio C. ¿Esto quiere decir que el proceso de remisión de dominios es infinito? La respuesta de Rapaport (1995b) es que no y, por lo tanto, tiene que haber un dominio-base que no se entiende en términos de otro dominio sino en términos de sí mismo; para que este entendimiento en términos de sí mismo sea posible, tiene que haber un entendimiento sintáctico. Otra respuesta a la cuestión del dominio inicial puede encontrarse en Fodor (1994), sin embargo, la naturaleza de lo que en Fodor (1994) es el lenguaje del pensamiento, guardará cierta distancia con lo que para Rapaport es el dominio inicial. Más adelante expondremos la respuesta de Fodor al problema de la remisión infinita.

En Rapaport, el entendimiento sintáctico no es excluyente con el entendimiento semántico, de manera que lo que tradicionalmente se interpreta como semántica, también puede ser el resultado de una manipulación sintáctica; esto permitirá una interpretación sintáctica de la semántica, expresada en términos de manipulaciones y remisiones entre dominios y entre los elementos de un mismo dominio. Contrario a lo que sucede en el entendimiento semántico, en el que hay una constante remisión de dominios, en el entendimiento sintáctico no hay un *siguiente dominio* al cual remitirse. Supóngase que en esta remisión de dominios, el dominio B no se remite a un dominio C, sino que en B mismo es posible hallar todos los elementos que permitan entenderlo. Esto quiere decir que B no se entiende en términos de otro dominio C, sino que se entiende en términos de sí mismo mediante un entendimiento sintáctico.

Red de representación semántica

Una máquina que se bloquea en una conversación, y que por lo tanto no supera el test de Turing, no es una máquina de pensamiento flexible; por el contrario, una máquina que puede lidiar con giros bruscos en el sentido de los conceptos y mantener una conversación como cualquier persona lo hace, es una máquina con pensamiento flexible que debe, de cierta forma, comprender el sentido de los conceptos que usa. Este es el punto de partida de Rapaport para la interpretación sintáctica de la semántica.

Cada concepto que interviene en una conversación y en el uso corriente del lenguaje, puede distribuirse en una red semántica "cuyos nodos representen conceptos individuales, propiedades, relaciones y proposiciones" (Rapaport 1995b, p. 51). Así, el significado de estos términos puede resultar de la conexión de los nodos que constituyen la red. Las conexiones de los nodos de la red deben ser suficientes para que un agente cognitivo produzca lo siguiente:

- Representaciones internas de objetos externos.

- Representaciones internas de los símbolos que el mismo agente cognitivo usa.

Como las conexiones al interior de la red permiten representaciones, entonces esta red puede considerarse como de naturaleza semántica. Mediante esta red de representación semántica[5]

5 Un ejemplo de una red de representación semántica es el SNePS, diseñada por Shapiro y su colega. Toda la información que el SNePS procesa, representa nodos que se conectan mutuamente. Así pues, incluso las proposiciones están representadas por nodos. Esto implica que el SNePS puede hacer remisiones, de manera que puede representarse, a sí mismo, una proposición. Además, el hecho de que los nodos puedan conectarse mutuamente de manera ilimitada, implica la posibilidad de que el SNePS constituya metaproposiciones; además, la cantidad de información que procese el SNePS puede aumentar agregándose nodos a la red. La estructura sintáctica básica de la red consiste en los arcos inicialmente establecidos entre nodos. Cada arco permite una representación distinta de un concepto. De esta manera, un arco establecido entre los nodos A y B permitirá representaciones semánticas tanto de A como de B; esto implica que, como se verá más adelante, ningún dominio, y

el agente cognitivo puede entender las emisiones lingüísticas de los otros agentes:

> Un agente C, con competencias de lenguaje natural entiende las emisiones de lenguaje natural de otro agente O, construyendo y manipulando los símbolos de un modelo interno (una interpretación) de la emisión de O, considerada como un sistema formal. El modelo interno de C será una representación de conocimiento y un sistema de razonamiento que manipula símbolos. Por lo tanto, el entendimiento semántico que C hace de O es una empresa sintáctica (Rapaport 1995b, p. 51).

Después de considerar una Red de Conocimiento de Representación (RCR)[6], la pregunta por el entendimiento, formulada desde Searle, continúa teniendo validez. Searle asegura que un modelo computacional que incluya un elemento semántico, no lo hace semántico en el mismo sentido en que la mente humana posee semántica; el hecho de que un programa tenga un elemento de implementación semántica no quiere decir que el programa posee una semántica en el mismo sentido en que cualquier persona lo hace; incluso, un modelo computacional puede poseer una pragmática y carecer de verdadera semántica. Esto, porque la inclusión del componente semántico en un modelo computacional, de nuevo, consiste en una simple manipulación formal, y así debe serlo por definición. Aunque se hable de un componente semántico en un RCR, Searle continuaría asegurando que este modelo carece del elemento semántico que posee la mente humana. No obstante, Rapaport (1995b) asegura que cuando se habla de un componente semántico en un RCR, en realidad se está hablando de la

ningún nodo en este caso, es semántico o sintáctico en sí mismo. Si, por otra parte, se agrega un nodo C al binomio inicialmente constituido por A y B, y se establecen arcos que permitan una conexión mutua entre los tres nodos, entonces la posible representación semántica de A, ahora será B o C, la posible representación semántica de B será A o C y la posible representación semántica de C será A o B. El conector disyuntivo de los nodos puede cambiarse por un conector conjuntivo, de manera que la representación de A no tiene que ser A o B, sino que puede ser una combinación de ambos nodos. Esto implica que a medida que la cantidad de nodos aumenta, también se enriquecerán las representaciones que la red semántica pueda generar (Shapiro y Rapaport, 1987).

6 En inglés, una red de representación de conocimiento y razonamiento se conoce comúnmente como KRR: knowledge-representation and reasoning.

capacidad de representación de conocimiento, tal como lo representa la mente humana. Ya se mostró que probablemente la constitución de representaciones puede ser el resultado de una constante remisión de dominios; sin embargo, a continuación se señalará cómo esta dinámica puede funcionar también para un dominio base.

Semántica, correspondencia y entendimiento

La semántica está íntimamente ligada a las representaciones; por este motivo, en la medida en que se suponga un abismo entre la sintaxis y la semántica, se supondrá también un abismo entre la sintaxis y las representaciones. No obstante, ya desde Fodor (1975) se puede concebir la posibilidad de representaciones que resulten de un marco computacional. La importancia de la posibilidad de generación representacional interna queda claramente señalada en *The Language of Thought*. Algunas de las premisas del modelo propuesto por Fodor consisten en afirmar que:

- Un agente cognitivo se encuentra en una situación (S).

- Que el agente cuenta con un cierto repertorio de conductas (B_1,...,B_n) disponibles a partir de la situación en que se encuentra.

- A partir de la consideración de S y (B_1,...,B_n), el agente puede predecir mediante computación la probabilidad de que cada B se dé.

- De acuerdo al cálculo que realiza el agente cognitivo, se asigna un orden de consecuencias.

- La elección final de conducta del agente está dada en función de la asignación de las preferencias y probabilidades usadas en los cálculos de las consecuencias.

Este modelo, a su vez, requiere la posibilidad de contar con un sistema representacional que permita las computaciones necesarias:

> De acuerdo al modelo, decidir es un proceso computacional; el acto que el agente ejecuta es la consecuencia de computaciones definidas sobre representaciones de posibles acciones (Fodor 1975, p. 31).

Rapaport, por su parte, sostiene que entender un concepto X, en realidad significa alguno de los dos siguientes enunciados que servirán para mostrar cómo se constituyen las representaciones y el elemento semántico a partir de manipulaciones sintácticas:

a). X se entiende *relativamente* a algo.

b). Uno se acostumbra a usar X.

Podría pensarse que el enunciado (a) implica un representacionalismo puramente semántico que no puede resultar de una manipulación formal; no obstante, siguiendo a Rapaport (1995a, 1995b, 2002), se puede asegurar que la remisión de dominios, que también es una operación formal, permite interpretar la semántica del entendimiento de los conceptos. Esto quiere decir que con la remisión de dominios se puede entender un concepto X en términos de *algo más*, sea otro concepto, una imagen, un sonido o cualquier otro estímulo. Por otra parte, el enunciado (b) se refiere a la implementación y a lo que sucede cuando hay una constante manipulación formal. Ambos enunciados, serán ampliados a continuación.

Semántica como correspondencia

En relación con el enunciado (a), se puede hablar de un *entendimiento relativo*; en este caso, si se entiende X, entonces X se entiende en términos de algo más. Esta puede ser la manera tradicional de interpretar el contenido representacional de la semántica característica de nuestro pensamiento, pues cada vez que escuchamos, leemos o usamos una palabra, entendemos y nos representamos *algo*; esto es precisamente lo que no le sucede a la persona que se encuentra dentro de la habitación china, pues para ella los símbolos chinos carecen completamente de contenido semántico. Ahora bien, este entendimiento *relativo* también puede interpretarse como una remisión de dominios (Rapaport 1995b, p. 53); en este caso, el entendimiento que un agente cognitivo C tiene de un concepto X, se constituye *relativamente* al propio entendimiento que el mismo agente cognitivo C tiene de otro concepto Y. Así, la remisión semántica entre dominios, para el caso del agente cognitivo C, consistirá en pasar del dominio al que pertenece el concepto X, al dominio al que pertenece el concepto Y; esto quiere decir que C se re-

presenta el concepto X en términos de Y lo cual, a su vez, implica que se debe tener un previo entendimiento de Y. Si Y es desconocido para C, entonces C no podrá representar el contenido de X en términos de Y sino en términos de *algo más* que sí entiende.

Lo anterior hace pensar que debe haber un dominio que no requiere, o no debe requerir, remisión a otro dominio para permitir el entendimiento; esto, para que la remisión no se torne infinita y la propuesta no caiga en un absurdo. Debe haber un dominio que se entiende en términos de sí mismo, sin remisión, y por lo tanto, no es relativo a ningún otro dominio. Este entendimiento no *relativo* es el entendimiento sintáctico.

Pensamiento por analogía y semántica como correspondencia

Siempre queremos comprender todo aquello desconocido a lo que nos enfrentamos cada día, ya sea una teoría, un objeto o una palabra. Por lo general, este procedimiento comienza con un razonamiento por analogía, es decir, tendemos a buscar algo que se le parezca a aquello desconocido, de manera que podamos *interpretar* o *entender*, en términos de algo ya dado o conocido. Parece que tendemos a definir lo desconocido y aquello que no entendemos, en términos de algo ya conocido. Es tan importante esta dinámica de analogía que según Minsky (1985), cuando se modele este tipo de pensamiento se logrará dotar a las máquinas con una mente mucho más flexible. La flexibilidad, característica del pensamiento humano se debe, en gran parte, al pensamiento por analogía (Minsky, 1985).

Al parecer, esta dinámica en la que pasamos de algo desconocido a algo conocido, está relacionada con este primer tipo de entendimiento, porque siempre tenemos que remitirnos a un dominio ya conocido, para poder entender los términos del dominio desconocido. Ahora bien, si se observa detalladamente, en la dinámica del pensamiento por analogía hay una red de conceptos que se conectan; esto permite que uno no se preocupe "de lo que el significado de los términos lingüísticos es, sino de las relaciones semánticas entre estos términos lingüísticos" (Rapaport 1995b, p. 55)[7]. Es común diferenciar entre las relaciones

7 Estas relaciones pueden ser, por ejemplo, de sinonimia, de antonimia, de implicación o de exclusión.

semánticas que hay entre objetos lingüísticos que componen la red de comprensión y las relaciones sintácticas, cuando en realidad, un examen minucioso muestra que estas son relaciones entre objetos lingüísticos que pueden expresarse en términos sintácticos. Así, aunque la semántica consista en correspondencia y remisión, puede interpretarse en términos de remisiones lingüísticas sintácticas.

La remisión sintáctica como semántica

Piénsese en el caso de un lenguaje artificial L. En estos lenguajes, por lo general, hay reglas que indican cómo manipular los símbolos para darle consistencia lógica a las posibles inferencias del sistema. Con respecto a estos elementos se puede preguntar: ¿Qué significan los símbolos y signos, si es que significan algo? ¿Qué significan las reglas que indican cómo manipular los símbolos? ¿Qué significan los teoremas y los resultados de las inferencias que se hagan con respecto a estos símbolos? Frente a esto, Rapaport señala lo siguiente:

> Para responder este tipo de preguntas, debemos ir afuera del dominio sintáctico, suministrando entidades "externas" para lo que el símbolo significa, y mostrando el trazado -las asociaciones, las correspondencias - entre los dos dominios […], ahora sucede algo curioso: necesito mostrarle el dominio semántico. Si soy afortunado, puedo señalárselo a usted […] y puedo describir las correspondencias ("El símbolo A significa que la cosa roja está allá"). Pero, a menudo, tengo que describirle el dominio semántico en… símbolos, y espero que el significado de estos símbolos le sean obvios a usted (Rapaport 1995b, p. 56).

La dinámica de aquello *relativo a*, es decir, el paso de aquello desconocido a aquello conocido que nos permite entender y representarnos un concepto, puede expresarse en términos de sintaxis. Esto quiere decir que en la semántica como correspondencia, la remisión de un dominio a otro consiste, en realidad, en la remisión de unos símbolos a otros. Se supone que en el experimento mental de la habitación china, los símbolos en chino son desconocidos e inentendibles. Para entender e interpretar estos símbolos, y luego formar representaciones internas, lo primero que se debe hacer es buscar un dominio adicional que permita, mediante analogía, formarse una idea de lo que estos símbolos significan. Dicho dominio adicional, como debe expresarse en términos de algo ya conocido, debe estar expresado en el lenguaje nativo del individuo que desea entender; así, en el caso de la habitación

china, podría asegurarse que el dominio *sintáctico* está constituido por los símbolos chinos, mientras que las instrucciones de procesamiento, dadas en el lenguaje nativo de la persona que se encuentra dentro de la habitación, constituirían el dominio *semántico* adicional que permitirá las asociaciones representacionales.

Lo anterior permite identificar la tradicional dicotomía que se propone entre un dominio sintáctico y un dominio semántico. Se supone que el lenguaje tiene unas reglas que determinan su uso y que estas reglas constituyen el dominio de lo sintáctico; por otra parte, se supone que el significado de los símbolos que constituyen el lenguaje, pertenecen a un dominio semántico completamente independiente al sintáctico. No obstante, estos dos dominios están dados en términos de símbolos y de las relaciones que hay entre ellos. Ambos dominios, el sintáctico y el semántico, pertenecen a una naturaleza puramente sintáctica.

Retomando el caso del lenguaje L, se puede asegurar que para proporcionar una interpretación semántica de este tipo de lenguaje, se debe contar con significados y representaciones para cada elemento constitutivo de dicho lenguaje[8] y para las combinaciones y operaciones que se pueden dar entre los términos. Con respecto al lenguaje L se puede formular un dominio no-vació A que incluya todo aquello posible que los términos significarán. Esto quiere decir que A incluirá todo aquello posible que se podrá pensar o decir acerca de L. A también incluirá una serie B que incluye aquello que las funciones significarán y una serie C que incluye aquello que las relaciones entre los términos incluidos en A significarán. Todas las operaciones que se puedan realizar en el dominio A estarán incluidas en un modelo que, por lo tanto, será el modelo de L. De esta manera, el significado de un término de L se puede construir estableciendo una relación entre el término perteneciente a L y una relación al interior del modelo; esto quiere decir que se establece una relación de representación entre el término y aquello que significa en el modelo; dicha relación puede denominarse *mapeo I*.[9]

8 Estos elementos pueden ser términos individuales, símbolos predicativos y símbolos de función (Rapaport 1995b, p. 56).

9 En algunos casos, cada significado que se pueda establecer en M corresponde a un término de L. Sin embargo, Rapaport (1995b, p. 56) reconoce que esta

Así, el dominio sintáctico y el dominio semántico siempre deben tomarse como un par acompañado, pues nunca se habla de un dominio sintáctico si no es "relativo a otro dominio que se toma como el semántico, y viceversa" (Rapaport 1995b, p. 60); siempre que se menciona un sistema puramente sintáctico, para que éste sea comprendido, tiene que relacionarse con otro sistema semántico en el cual se encuentran los significados y las interpretaciones.

Ahora bien, ¿Qué hace que un dominio sea semántico o sintáctico?, ¿Algunos sistemas, o dominios son en sí mismos semánticos o sintácticos? Según Rapaport (1995b), no. El hecho de que siempre se cuente con una pareja de dominios no quiere decir que alguno de estos siempre tiene de desempeñar el papel del dominio sintáctico o del dominio semántico; esto quiere decir que ningún dominio posee, en sí mismo, una naturaleza semántica o sintáctica. De hecho, como lo señala Rapaport (1995b), en algunos casos un sistema formal, esto es sintáctico, puede ser una interpretación semántica de otro sistema formal. En el caso del lenguaje L que se interpreta mediante el modelo que contiene A, B y C, se puede producir un lenguaje L´ que, a su vez, podrá ser interpretado en términos de L cuando nos hayamos familiarizado al uso de L. Esta familiaridad, como se verá más adelante, se refiere a la cuestión de la implementación de un sistema sintáctico. L sirve, a la vez, de dominio sintáctico para el modelo constituido por A, B y C, pero cuando nos hayamos acostumbrado a L, este también servirá como dominio semántico para L´ (Rapaport 1995b, p. 57). En realidad, lo que hace que un dominio sea entendido como puramente semántico es que sea previamente comprendido o entendido (Rapaport 1996b, p. 60). Lo anterior quiere decir que, mientras un cúmulo de símbolos no sea entendido, será interpretado como sintáctico; cuando este cúmulo se entiende y se comprende, se le considera como perteneciente al dominio de lo semántico porque, a su vez, servirá para entender un nuevo dominio sintáctico.

concordancia no se da si L es el lenguaje inglés y M es el mundo actual, pues en este caso hay términos de L que no se pueden encontrar en M.

Entendimiento sintáctico

Entendimiento sintáctico en un sistema cerrado

¿Cómo entendemos el primer y el último dominio de la cadena de dominios que componen la dinámica de remisión? En términos de sí mismo, es decir, en términos de pura sintaxis:

> Podemos tener una "cadena" de dominios, cada uno de los cuales, excepto el primero, es un dominio semántico para su predecesor, y cada uno de los cuales, excepto el último, es un dominio sintáctico para su sucesor. ¿Cómo entendemos el último? Sintácticamente (Rapaport 1995b, p. 57).

Esto permite que la remisión de dominios no sea una hipótesis ahogada en el absurdo. Pero esta respuesta no funciona si no se responde esta otra cuestión: ¿Qué quiere decir que entendemos algo en términos de sí mismo? Se supone que, según lo dicho hasta ahora, la interpretación semántica implica la presencia de, al menos, dos dominios; por lo tanto, si decimos que en el entendimiento sintáctico solamente hay un dominio que se entiende en términos de sí mismo se estaría contradiciendo, aparentemente, todo lo expuesto; según Rapaport (1995b), no hay tal contradicción. Se puede asegurar que entender algo consiste en establecer relaciones de correspondencia; también se puede asegurar que el resultado es un mapeo que se hace del dominio semántico al dominio sintáctico. Teniendo en cuenta esto, entonces, entender un dominio en términos de sí mismo implica establecer relaciones entre los propios términos de aquel único dominio; es decir, aquel dominio inicial desempeñará un papel sintáctico y semántico a la vez. Así, aunque en sentido estricto solamente hay un dominio, algunas partes del dominio desempeñarán una función sintáctica y otras una función semántica; a través del tiempo, las funciones de cada parte del dominio podrá intercambiarse. Esto quiere decir que el dominio se entiende en términos de sus partes.

Un ejemplo de entendimiento sintáctico son las definiciones incluidas en un diccionario pues el significado de las palabras está dado en términos de otras palabras; en general, los significados de unos símbolos están dados en términos de otros símbolos (Rapaport 1995b, p. 57). Este tipo de respuesta puede dejar insatisfecho al lector cuidadoso; decir que el absurdo al infinito -por constante remisión entre

dominios- se rompe con una remisión circular en el dominio sintáctico inicial, puede verse como un procedimiento en el que eliminamos una falacia con otra. No obstante, la circularidad del dominio inicial, esto es, la posible circularidad que resulte de la autorreferencialidad en el entendimiento sintáctico -como la autorreferencialidad del diccionario-, tiene su justificación en el funcionamiento de un sistema cerrado y generativo como el sistema nervioso. La circularidad puede entenderse y, en cierta manera justificarse, desde los sistemas autopoiéticos y la cláusula del funcionamiento del cerebro.

Todo lector aceptará que el entendimiento y la semántica se producen en el cerebro; de hecho, esta aceptación, por lo general, es mucho más radical de lo que parece conveniente;[10] en este mismo orden de ideas, Searle acepta que la Intencionalidad es el resultado de la biología cerebral. Aunque un funcionalista puede aceptar que los estados intencionales se encuentren en sistemas físicos distintos al cerebro, no necesariamente niega que en el cerebro hay y se producen dichos estados intencionales. Así, tanto Searle como un funcionalista podrían aceptar que en el cerebro se producen estados intencionales; sin embargo, al examinar la estructura de los *inputs* y *outputs* del sistema, resultará una curiosa pero afortunada consecuencia.

Por su parte, Searle acepta que el cerebro es una máquina biológica y que, en esta medida, su particular configuración de *hardware*, es decir, de soporte físico, le otorga los poderes causales que posee. Por otra parte, el funcionalismo acepta que el cerebro también es una máquina, pero que su configuración física no le otorga un status causal privilegiado. Para el funcionalismo de máquinas de Turing, las conductas *-outputs-* que nos permiten inferir la presencia de un entramado mental, son el resultado de un correcto *input* y de un correcto programa. No obstante, mientras la aceptación de la idea de programa y constitución de representaciones a partir de computaciones y remisiones interdominios o intradominios nos obligue a aceptar autorreferencialidad -para el caso del entendimiento sintáctico intradominio- o remisión al infinito -para el caso del entendimiento semántico como correspondencia

10 En algunos casos se interpreta al cerebro como único fundamento causal de la Intencionalidad, cuando en realidad, en la producción de casi todos los estados mentales interviene una gran cantidad de órganos distintos al cerebro. Este aspecto será ampliado en el siguiente capítulo.

interdominio-, la propuesta, será débil. Por este motivo, proponemos una leve variación al postulado funcionalista, que no resulta de una excentricidad teórica sino más bien de aceptar algunas características del cerebro y el sistema nervioso. En nuestra opinión, la complicación que acabamos de señalar, a saber, la de reemplazar una falacia por otra, resulta del hecho de que el funcionalismo acepta que el *input* proviene del medio externo porque, en esta medida, el procesamiento de dicho *input* se posibilita mediante la presencia de un programa de procesamiento; sin embargo, para Searle, aún sin este programa, el *input* es correctamente procesado y la conducta *-output-* es correctamente producida.

De esta manera, puede cuestionarse la relevancia del programa que determina las reglas de transformación del sistema y, a su vez, el procesamiento de los *inputs*. Esto quiere decir que el programa puede interpretarse como una entidad teórica que, si bien posee valor explicativo, no es necesaria para entender porqué los organismos poseen conducta y vida mental pues, básicamente, para Searle el *input* también se procesa correctamente sin necesidad de postular un programa, asegurándose que la configuración del sistema físico basta para explicar el procesamiento. Esto quiere decir que, tanto para Searle como para el funcionalismo, los *inputs* se procesan correctamente para producir ciertos *outputs*, por lo tanto, la postulación de un programa no es primordial. Esto porque, según Searle, la conducta se da en virtud de la semántica resultante de los poderes causales del cerebro y no únicamente en virtud de la presencia del programa; en este orden de ideas, puede asegurarse que si se acepta la presencia de un programa, éste tendrá que complementarse con una semántica propia de los humanos o que, simplemente, no es necesario postular un programa sino únicamente la semántica resultante de la constitución cerebral. De esta manera, la relevancia del programa no parece ser totalmente convincente. Ahora bien, ¿Qué sucede si la presencia de este programa no solamente permite explicar cómo se procesarán los *inputs* para producir ciertos *outputs*, sino que el programa, en principio, posibilita el *input* mismo?

Generalmente se acepta que cuando un órgano perceptivo de nuestro cuerpo entra en contacto con un estímulo, resulta un *input* que inmediatamente procesa el cerebro. No obstante, puede asegurarse que entre el choque del órgano perceptivo y el *input* procesado por el cerebro, hay un trecho en el que interviene un programa. Lo verdaderamente importante del *input* no es lo que ese *input* es al

momento de ponerse en contacto con nuestros órganos perceptivos, sino lo que ese *input* es al momento de ser interpretado por el cerebro. La totalidad de las percepciones que constituyen nuestras experiencias conscientes resultan de un choque entre la materia y nuestros órganos de percepción (Salcedo-Albarán, 2004); así, en el mundo no hay colores, sonidos o aromas, pues éstos resultan de la interacción entre las partículas físicas que hay en el mundo y nuestros aparatos de percepción. Sin embargo, cuando el *input* se procesa, no se procesa como aquella materia bruta que ha chocado con el órgano de percepción, sino que se procesa como un cierto tipo de interpretación dada al choque de la materia. Esta interpretación se da gracias a los cambios físicos en el sistema nervioso, los cuales modifican el choque inicial de materia hasta producir un *estímulo*; pero, adicionalmente, esta interpretacion se da gracias a las especificaciones de cambios de la estructura interna del sistema; esto es, gracias a un programa que indica cómo deberá cambiar físicamente el sistema para la producción de un estímulo. Este programa, entre otras funciones, deberá generar un estímulo en los términos conscientes de su portador, esto es, únicamente en las palabras y conceptos con los que cuenta el sujeto.

Así, lo verdaderamente importante del *input*, a saber, lo que el *input* es cuando va a procesarse conscientemente, resulta de los cambios estructurales especificados por el programa. Las reglas de transformación física que se encuentran en este programa determinan, como en cualquier máquina de Turing, el estado estructural en que el sistema se encontrará en *t+1* dado el estado en *t* y un *input* particular. La cuestión relevante será que en este caso, el programa, sus especificaciones y las variaciones resultantes determinarán en cada instante el mismo *input* y, por lo tanto, el programa no solamente será medianamente relevante en la producción del *output* -conducta-, sino que será determinante en la producción del mismo *input* que es insumo para la conducta. La dinámica de interpretación del *input* producido forma parte de otra historia; por ahora, nos ocupa la producción del *input*. Esto quiere decir que no basta el choque entre la materia y nuestros aparatos perceptivos para explicar la constitución de un *input*, sino que es necesario que dicho choque sea constituido gracias a las transformaciones físicas ordenadas en el sistema nervioso, que resultan, a su vez, de las especificaciones dadas por un programa.

Maturana y Varela (1990) muestran cómo la constitución de lo que generalmente entendemos como estímulo, en realidad es el resultado de cambios estructurales dados al interior del sistema, en este caso, del sistema nervioso. Generalmente, hablamos de estímulos que resultan del choque entre los órganos de percepción y los objetos del mundo externo; sin embargo, incluso la constitución del *arriba* y *abajo*, que en principio podría entenderse como una característica inherente al mundo externo y que aprehendemos mediante nuestras percepciones, en realidad es el resultado de una intervención de la estructura del sistema y, sobre todo, de sus transformaciones:

> Se pude tomar un renacuajo o larva de sapo, y con pulso de cirujano, cortar el borde del ojo […] y rotarlo en 180 grados. Al animal así operado se le deja completar su desarrollo larval y metamorfosis hasta convertirse en adulto. Tomamos ahora nuestro sapo-experimento y le mostramos un gusano cuidando de cubrir su ojo rotado. La lengua sale y vemos que hace un blanco perfecto. Ahora repetimos el experimento, esta vez cubriendo el ojo normal. En este caso, vemos que el animal tira la lengua con una desviación exacta de 180 grados. Es decir, si la presa está abajo y al frente del animal, como sus ojos miran un poco hacia el lado, éste gira la lengua a lo que era atrás y arriba. Este experimento revela de una manera muy dramática que para el animal no existe, como para el observador que lo estudia, el arriba o el abajo, el adelante o el atrás referidos al mundo exterior a él. Lo que hay es una correlación interna entre el lugar donde la retina recibe una perturbación determinada y las contracciones musculares que mueven la lengua, la boca, el cuello y en último término todo el cuerpo del sapo (Maturana y Varela 1990, p. 107).

Lo que el organismo interpreta como un estímulo, en realidad es el resultado del funcionamiento y el cambio de la estructura interna del sistema-organismo; de esta manera, el *input* puede interpretarse como el resultado de las diferentes relaciones de la estructura interna del organismo:

> Este experimento, como muchos otros […] puede ser visto como evidencia directa de que el operar del sistema nervioso es expresión de su conectividad o estructura de conexiones, y que la conducta surge según el modo cómo se establecen en él sus relaciones de actividad internas (Maturana y Varela 1990, p. 108).

Así, en la medida en que el sistema nervioso se interprete como un sistema cerrado es necesario que haya algún tipo de circularidad y

autorreferencialidad en los términos que constituyen la consciencia. "Los seres vivos se caracterizan porque, literalmente, se producen continuamente a sí mismos" (Maturana y Varela 1990, p. 36), y esto implica que en ellos sea común hallar relaciones autorreferenciales. En cierta medida, la neurofisiología ya se ha percatado de que el sistema nervioso funciona más como un sistema cerrado con especificaciones de relaciones internas funcionales, que como un sistema abierto al directo impacto de los estímulos:

> El cerebro opera como un sistema autorreferencial, cerrado al menos en dos sentidos: en primer lugar, como algo ajeno a la experiencia directa, en razón del cráneo, hueso afortunadamente implacable; en segundo lugar, por tratarse de un sistema básicamente autorreferencial, el cerebro sólo podrá conocer el mundo externo mediante órganos sensoriales especializados […]. En tal tipo de sistema, la entrada sensorial desempeñaría un papel más importante en la especificación de los estados intrínsecos (contexto) de actividad cognoscitiva, que en el puro suministro de "información" (contenido) […]. Las señales sensoriales adquieren representación gracias a su impacto sobre una disposición funcional preexistente del cerebro (Llinás 2002, p. 10).

Cuando un estímulo sensorial es captado por un órgano de percepción, en realidad se da una interpretación que, a su vez, resultará en la producción de un *input*; lo importante del *input* no es lo que éste es cuando se constituye como estímulo captado por el órgano de percepción, sino lo que éste es cuando se interpreta al interior del sistema autorreferencial. Pero, de nuevo ¿El programa no es un lujo explicativo, relativamente irrelevante, de manera que la transformación estructural del sistema físico es suficiente para explicar la constitución de las percepciones? ¿El sistema físico y su estructura bastan para explicar el flujo de estados sin necesidad de postular la intervención de un programa? No. En términos físicos, el sistema cuenta con unas posibilidades causales. Esto quiere decir que, a partir del estado A en *t*, el sistema está causalmente posibilitado para pasar al estado B, C o D en *t+1*; los posibles estados que en *t+1* puede adoptar el sistema físico aumentarán en la medida en que dicho sistema sea lo suficientemente complejo como para instanciar un programa de cómputos complejos; es decir, no posee las mismas posibilidades causales una pared, en tanto sistema físico, que la materia cerebral. Ahora bien, si todo dependiera únicamente de la transformación estructural del sistema, sin especificación en las transformaciones estructurales por parte de un programa, entonces no habría directrices estructurales en las transformaciones.

El sistema físico posee unas posibilidades causales y en la transformación estructural, de no haber programa, el resultado para *t+1* sería cualquiera de las posibilidades físicas del sistema. El estado interno resultante en *t+1* sería, siempre, aleatorio. Pero este no es el caso; si así fuera, no habría regularidad morfológica o estructural en los organismos. De este modo, tiene que haber alguna directriz informacional -un programa-, que indique a qué estado o a qué posibles estados -que no será la totalidad de los estados causalmente posibles- debe pasar el sistema en *t+1*. Las transformaciones embriológicas representan un ejemplo de cómo, a partir de un cúmulo de pocas células, cada organismo adquiere una morfología y una organización estructural interna típica gracias a las especificaciones de un programa. Todos los seres humanos son similares, -excepto en su ADN-, en la conformación del zigoto y, no obstante, a partir de este zigoto inicial se producen seres humanos distintos pero con regularidades acotadas por las características de la especie, gracias a la presencia de cierta información que indica el tipo de transformaciones físicas que se deben dar. Este programa, en el caso de los seres vivos, consiste en instrucciones codificadas en series de pares de bases de ácido nucleico almacenadas en la molécula de ADN (Dawkins, 1976). Sin esta codificación nucleica, habría una total deriva caótica en la morfología y la conducta de los seres vivos.

El zigoto, como célula embrionaria, es un sistema físico con posibilidades causales físicas lo suficientemente amplias para producir personas altas, bajas, negras, blancas, amarillas y, en general, personas con una amplísima posibilidad fenotípica. Además del sistema físico, se requiere un programa que indique las transformaciones típicas que el sistema debe seguir en su desarrollo. Por este motivo, en la filogenia se preservan y mantienen ciertas características morfológicas y conductuales con cierta regularidad. Aunque un zigoto sea potencialmente capaz de producir fenotipos variados, por lo general se preservan ciertas características con ciertas regularidades, sobre todo, entre las familias.

Debe tenerse en cuenta que, el hecho de que la información tenga que implementarse siempre en una contrapartida física, puede hacer pensar que solamente hay estructura física sin información implementada (Churchland, 1999); sin embargo, es necesario que la información siempre sea implementada en alguna contrapartida física

porque, de lo contrario, estaríamos hablando de dualismo. Lo anterior no quiere decir que información y soporte físico se interpreten en el mismo nivel, sólo quiere decir que hay cierta información almacenada e implementada en una estructura física:

> El gen es un paquete de información, no un objeto. El patrón de los pares de base en una molécula de ADN especifica el gen. Pero la molécula de ADN es el medio, no el mensaje. […]. En biología, cuando usted está hablando de cosas como genes y genotipos […], usted está hablando acerca de información, no de la realidad de un objeto físico. Son patrones. […] Considere el libro Don Quijote: una pila de papel con marcas de tinta en las páginas, pero usted podría colocarlo en un CD o un casete y convertirlo en ondas de sonido para gente ciega. No importa en qué medio se encuentre, es siempre el mismo libro, la misma información. […] Comparar un gen con un individuo, por ejemplo […], es inapropiado si por individuo usted quiere decir un objeto material y por gen usted quiere decir un paquete de información (Williams 1996, p. 42).

En el caso del ADN, cada unidad mínima del programa está impresa en forma de orden de los pares de bases de los ácidos nucleicos, de manera que el orden de dichos ácidos determinará las especificaciones del programa:

> Nuestro ADN vive dentro de nuestros cuerpos. No está concentrado en una parte particular del cuerpo […]. Este ADN puede verse como un grupo de instrucciones acerca de cómo hacer un cuerpo, escritas en el alfabeto A, T, C, G de los nucleicos (Dawkins 1976, p. 22).

Dadas las especificaciones del programa, la estructura del sistema seguirá ciertos desarrollos determinados. Si estas especificaciones no estuvieran dadas, entonces no encontraríamos las típicas transformaciones registradas en el desarrollo de los organismos. Esto quiere decir que no podríamos determinar regularidades en las características fenotípicas de ciertos linajes. En este orden de ideas, la presencia de un programa no solamente constituye un lujo explicativo, sino que es imprescindible para explicar cómo y por qué los sistemas físicos que llamamos personas, y en general las especies, mantienen regularidades morfológicas y conductuales. En la medida en que el programa determine las transformaciones estructurales internas, entonces los *inputs* interpretados también estarán determinados por el programa en un sistema cerrado y autorreferencial.

Puede cuestionarse si lo que sucede en las transformaciones embriológicas también sucede en la vida mental de las personas. Si no fuera así, y solamente fuera necesario postular un programa para las transformaciones iniciales del sistema físico que conocemos como persona y no para su vida mental, entonces, de nuevo, las regularidades conductuales no serían posibles. Por lo general, a través de los linajes se preservan ciertas características conductuales y mentales; así, el desarrollo sináptico particular se mantendrá en cierta medida lo suficientemente acotado como para esperar que se preserven características mentales a través de los linajes gracias a la especificación del programa. Estas especificaciones determinarán y acotarán el desarrollo estructural del sistema -incluyendo el desarrollo sináptico- para que las constituciones del *input* sean compartidas por agrupaciones de organismos, por ejemplo, por especies; así, en el caso de los seres humanos, gracias a las acotaciones dadas por el programa -genoma típico de los seres humanos-, el desarrollo estructural permitirá *inputs* similares entre la especie humana. A su vez, toda esta dinámica de constitución del *input* se desarrollará en una constante autorreferencialidad de los términos de la estructura:

> Incluso antes de llegar a la corteza, cuando la proyección de la retina entra al cerebro, en lo que se llama el núcleo geniculado lateral del tálamo (NGL), este no es simplemente una vía de estación de la retina hacia la corteza, sino que converge a este centro muchos otros centros con múltiples efectos que superponen a la acción retinial […]. Nótese […] que una de las estructuras que afecta al NGL es, precisamente, la corteza visual misma. Es decir, ambas estructuras están entrabadas, en una relación de efecto mutuo y no de una simple secuencialidad (Maturana y Varela 1990, p. 139).

Esta autorreferencialidad del sistema, a su vez, posibilita el conocimiento y la vida mental que resulta del sistema nervioso:

> El sistema nervioso, por tanto, por su propia arquitectura no viola sino que enriquece este carácter autónomo del ser vivo. Empiezan ya a ser claros los modos como todo proceso está necesariamente fundado en el organismo como una unidad y en el cierre operacional de su sistema nervioso, de donde viene que todo su conocer es su hacer como correlaciones sensoafectoras en los dominios de los acoplamientos estructurales en los que existe (Maturana y Varela 1990, p. 142).

Gracias a esta circularidad característica del sistema, se justifica la autorreferencialidad constitutiva del entendimiento sintáctico. Si el sistema nervioso y el cerebro fueran sistemas abiertos, no autorreferenciales, entonces no habría mayores razones para aceptar la idea de que hay una dinámica de representacionalidad como la descrita en el entendimiento sintáctico. Como mostramos líneas atrás, el hecho de que el sistema sea cerrado no quiere decir que en él no hay flujo de estímulos, tan sólo quiere decir que este flujo está, a su vez, incluido en la autorreferencialidad y constituido por una interpretación dada por la circularidad del sistema. A su vez, esta circularidad estructural estará determinada por las especificaciones de transformaciones estructurales que indique el programa. Así, cuando en adelante se señale que los estímulos del mundo externo permiten el correcto funcionamiento de la dinámica de mapeo entre dominios y, por lo tanto, la correcta constitución de representaciones, estaremos asegurando que estos estímulos producirán los correspondientes cambios estructurales internos que, a su vez, desembocarán en la producción de lo que los observadores describimos como conductas.[11]

Sistema base

Fodor reconoce que "las oraciones de un lenguaje natural podrían comunicar lo que comunican aún si los elementos del vocabulario fueran menos de los que son" (Fodor 1975, p. 124). Esta eliminación de elementos del vocabulario de un lenguaje natural se posibilita gracias a que, como en el caso del diccionario, algunos elementos del vocabulario están definidos en términos de otros elementos y, por lo tanto, se pueden eliminar. Por ejemplo, si *soltero* significa *hombre no casado* entonces, según Fodor, todo lo que puede decirse en un lenguaje que contiene uno de estos elementos -soltero u hombre no casado-, puede decirse en otro lenguaje que contenga el otro término (Fodor 1975, p. 124). Sin embargo, no es posible eliminar *todos* o *cualquier* elemento del vocabulario. Aquellos elementos que no se puede eliminar constituyen lo que Fodor denomina el conjunto de *las bases primitivas:*

11 "[…] la conducta es la descripción que hace un observador de los cambios de estado de un sistema con respecto a un medio al compensar las perturbaciones que recibe de este […]. Pero, en principio, toda conducta es una visión externa de la danza de relaciones internas del organismo. El encontrar en cada caso los mecanismos precisos de tales coherencias neurales es la tarea abierta al investigador." (Maturana y Varela 1990, p. 141].

> El interés en la noción de una base primitiva […] es este: hemos visto que el sistema de las representaciones internas para las oraciones de un lenguaje natural debe, al menos, capturar el poder expresivo de esas oraciones. Ahora parece que las bases primitivas de un lenguaje determinan su poder explicativo en la medida en que este último es una función del vocabulario (Fodor 1975, p. 124).

Gracias a que los significados de las palabras están dados en términos de otras palabras, aprender el uso y significado de una nueva palabra consiste en remitirse al contexto de esa palabra; es decir, para entender un símbolo perteneciente al lenguaje corriente, debemos remitirnos a las otras partes de ese mismo lenguaje corriente. Cuando aprendemos nuestra primera lengua, no aprendemos el uso de las palabras memorizando sus definiciones (Miller y Gildea 1987, p. 80), sino aprendiendo su uso en un contexto, de manera que "el significado de la palabra desconocida es […] el contexto que la rodea - el contexto menos la palabra" (Rapaport 1995b, p. 76; Rapaport y Kibby, 2002). De esta manera, aquel primer dominio se entiende en términos de sí mismo, mediante una remisión constante y mutua entre las partes que lo componen. Tan determinante será el contexto de una palabra para su uso y significado, que a medida que conocemos nuevos contextos, el significado de la palabra se altera; de hecho, es posible que según la cantidad de contextos a los que se haya tenido acceso, una persona conozca más significados y usos de una palabra. Por ejemplo, una persona que solamente ha escuchado y observado el uso de la palabra vela, en el contexto de frases como *debemos alumbrar con una vela*, *tráeme la vela para alumbrar*, o *se va a apagar la vela y nos vamos a quedar a oscuras*, no sabrá que dentro del significado de *vela* también se cuenta ser un "conjunto de piezas de lona y lienzo fuerte las cuales, cortadas de diversos modos y cosidas, se amarran a las verjas para recibir el viento que impele a la nave" (Salvat 1990, p. 3665); esto quiere decir que la persona que sólo haya tenido acceso al primer tipo de frases y no a frases como *se debe extender la vela para que el barco avance* o *sin la vela no podremos usar el viento para navegar*, desconocerá un importante uso y significado de la palabra vela: "el entendimiento que uno tenga de una palabra cambiará según uno se encuentre con más contextos en los cuales se usa esa palabra" (Rapaport 1995b, p. 76) porque, de esa manera, la palabra se podrá entender mediante la remisión a una mayor cantidad de otras palabras. Al principio, escuchamos o leemos la palabra desconocida y no

encontramos ninguna asociación semántica o representación; luego, cuando ya hemos tenido contacto con la palabra, una o más veces, formamos la representación mental mediante la asociación entre la nueva palabra y el contexto o, de ser posible, entre la palabra y algún modelo mental. De esta manera, aunque se cree que la representación mental pertenece a una naturaleza semántica misteriosa, en realidad resulta de remisiones y asociaciones entre palabras o entre palabras y el registro del mundo -modelo mental- constituido por la recepción de estímulos que acumulamos a través de nuestra vida y las variaciones en nuestras estructuras internas y nuestros sistemas nerviosos.

El lenguaje del pensamiento

Lenguaje innato

Si se acepta lo señalado hasta ahora, se supera la objeción de que siempre será necesario un dominio precedente que permita entender el lenguaje. Según lo señalado, se sustenta la idea de que en realidad sí es necesario un dominio inicial, pero no que siempre es necesario un dominio precedente. La anterior respuesta no es la única manera de superar esta objeción pues algo similar se le objeta a Fodor. No obstante, es necesario señalar que mientras que Rapaport no determina la necesidad de entender el dominio inicial como un *lenguaje del pensamiento* que sea privado, Fodor (1975) sí señala que dicho lenguaje fundamental debe ser privado. Aunque la naturaleza de los lenguajes iniciales es distinta en Rapaport y en Fodor, la objeción parece ser extensiva a ambas posturas.

Según Fodor, alguien puede objetar que "uno no puede aprender un lenguaje hasta que no sepa uno" (Fodor 1975, p. 65). En este orden de ideas, se supone que debe haber un metalenguaje que permite entender el lenguaje natural y, por lo tanto, que permite la construcción inicial de representaciones. Sin embargo, para poseer este metalenguaje muy seguramente también se debe contar con un meta-metalenguaje que abarque las posibilidades representacionales del metalenguaje; como se puede observar, esto lleva a un regreso al infinito (Fodor 1975, p. 65). Sin embargo, según Fodor, esto no constituye un verdadero dilema pues se puede asegurar que uno de los lenguajes no es, necesariamente, aprendido: "el lenguaje del pensamiento es conocido pero no aprendido. Esto es, innato" (Fodor 1975, p. 65). Según

Fodor, lo verdaderamente importante del lenguaje del pensamiento no es que sea aprendido sino entendido y que, por lo tanto, permita los procesos cognitivos del agente. Este es el caso de los programas compiladores de los computadores, que permiten interpretar y establecer condiciones correctas entre el lenguaje *input/output* (o interface de usuario) y el lenguaje mediante el cual los computadores se comunican entre sí. Este compilador forma parte de la ingeniería del computador y, por lo tanto, no requiere de un metacompilador; esto, a su vez, quiere decir que el lenguaje del pensamiento abarca, explica y permite la realización de las actitudes proposicionales, gracias a las características físicas del organismo, mediante relaciones con símbolos:

> De acuerdo con las formulaciones normales, creer que P es mantener una cierta relación con una muestra de un símbolo que significa que P [...]. Ahora bien, los símbolos tienen contenidos intencionales y sus muestras son físicas en todos los casos conocidos. En tanto que físicas, las muestras de símbolos son la clase correcta de cosas para manifestar papeles causales (Fodor 1994, p. .192)

Según Fodor, el lenguaje del pensamiento constituye un modelo explicativo que puede aceptarse en virtud de que "teorías remotamente plausibles son mejores que no tener teorías" (Fodor 1975, p. 27). Al analizar los fenómenos de *percepción* como fijación de creencias preceptúales, toma de decisiones como representación y evaluación de consecuencias de las acciones a partir de una determinada situación, y *aprendizaje de conceptos* como una hipótesis de confirmación, Fodor concluye que los procesos computacionales realizados sobre representaciones son indispensables para un correcto modelo explicativo de cada uno de estos fenómenos. Esto, por supuesto, entra en conflicto con la idea de Searle de que la computación no puede descubrirse en un sistema físico como un cerebro sino que solamente puede adscribirse. Mientras Searle asegura que la computación no es un rasgo de los sistemas físicos, Fodor (1975) sostiene que, básicamente, los modelos de cada uno de estos fenómenos consisten en procesos computacionales realizados sobre representaciones. Así, Fodor (1975) concluye que debe haber un lenguaje del pensamiento que permita y sirva de plataforma para los procesos computacionales necesarios en cada fenómeno mencionado.

El lenguaje del pensamiento permite asegurar que *los estados intencionales* poseen una estructura constitutiva característica (Fodor 1994, p. 192). Estas estructuras, gracias a la lógica simbólica y gracias a la idea de la máquina de Turing, pueden formalizarse mediante símbolos y relaciones sintácticas. Por una parte, se puede formalizar la estructura y, después de esta formalización, dado que Turing mostró la posibilidad de que todo algoritmo se pueda instanciar en una máquina, entonces se llega a la conclusión de que "pensar consiste en procesar representaciones físicamente realizadas en el cerebro (en la manera como las estructuras internas de datos son realizadas en el computador)" (Aydede, 2004). Así, el lenguaje del pensamiento posee las propiedades combinatorias sintácticas que posibilitan las representaciones.

Supóngase un mecanismo, *la caja de intenciones*, mediante el cual se realiza la intención de hacer verdadera una proposición. Así pues, según Fodor, cuando se intenta hacer verdadera la proposición *P*, "lo que se hace es poner dentro de la caja de intenciones una muestra de un símbolo mental que *significa* que P" (Fodor 1994, p. 192). Cuando el símbolo formal correspondiente a *P* se coloca dentro de la caja, en ella se realizan computaciones "y el resultado es que uno se comporta de un forma que (*cetetis paribus*) hace verdad que *P*" (Fodor 1994, p. 192). Ahora bien, para ser más específicos, Fodor (1994) señala que para la realización de la proposición, se debe poner dentro de la caja una subexpresión que significa *P*.

De esta manera, hay una semántica combinatoria mediante la cual se puede explicar la complejidad de un estado intencional completo en términos de contenidos regulares determinados por el contenido regular de las partes de la sintaxis (Fodor 1994, p. 194). Esta naturaleza semántica permitirá la construcción significante tanto de los elementos atómicos, como de los elementos moleculares, pues los principios sintácticos combinatorios del lenguaje del pensamiento y las reglas computacionales que determinan dichos principios permiten que, a partir de los significados dados de los elementos atómicos, se construya la significación molecular. Así, se puede asegurar que un computador es un entorno en el que los símbolos se pueden manipular gracias a sus características físicas y, no obstante, aún se preservarán las propiedades semánticas porque estas están definidas en términos

de partes sintácticas y procesos de manipulación de dichas partes (Aydede, 2004). La relación causal de los estados de creencias implica la suposición de que dichos estados tienen una estructura combinatoria que, a su vez, resulta de aceptar que:

> Los estados de creencia, de alguna manera, se construyen a partir de elementos, y que los objetos intencionales y los papeles causales de cada uno de los estados dependen de qué elementos contenga y de cómo se pongan juntos esos elementos. La historia del Lenguaje Del Pensamiento es, por supuesto, un paradigma de esta clase de explicación, dado que admite que el creer implica una relación con un objeto estructurado sintácticamente para el que se asume la semántica composicional (Fodor 1994, p. 208).

Las personas pueden construir, en principio, una cantidad ilimitada de proposiciones a partir de un número finito de elementos; de hecho, las personas pueden construir un número infinito de pensamientos a partir de un número finito de ellos (Aydede, 2004). Así, aunque no tenemos que estar familiarizados con todas las oraciones que escuchamos, en principio podemos entender cualquier tipo de enunciado que esté correctamente construido; tenemos una capacidad ilimitada de pensamiento aunque partimos de elementos finitos. Esta capacidad se explica en virtud de que contamos con un sistema simbólico que abarca las representaciones mentales, que son los objetos directos e inmediatos de las actitudes proposicionales y que, además, las representaciones del sistema posibilitan una semántica combinatoria. De esta manera, los elementos moleculares se construyen con base en elementos semánticos simples.

Implementación e internalización: el paso de la sintaxis a la semántica en la remisión de dominios

Siguiendo a Searle, la persona que se encuentra dentro de la habitación china ejecutando un programa, está siguiendo reglas, mientras que nosotros, en nuestro pensamiento no estamos ejecutando un programa y, por lo tanto, no estamos siguiendo constantemente reglas que nos indica cómo pensar. Supongamos que esto es cierto; de hecho, podemos estar seguros de que conscientemente casi nunca estamos siguiendo reglas específicas que nos indican cómo pensar. Si nuestro pensamiento obedece a reglas, entonces lo hace

a un nivel inconsciente; es claro que no nos detenemos constantemente a reflexionar de manera consciente acerca de nuestras operaciones mentales o nuestras acciones en la vida diaria; sin embargo, nuestro flujo de pensamientos y nuestras acciones -en condiciones normales- nunca se detienen por la carencia de reflexiones detalladas. Caminamos, respiramos, comemos y ejecutamos una serie de acciones para las cuales no necesitamos reflexionar acerca de procedimientos; incluso cuando estamos muy acostumbrados a alguna acción definida mediante reglas, dejamos de pensar en las reglas y comenzamos a actuar *de acuerdo a ellas*, de manera inconsciente. Lo mismo sucede con el uso del lenguaje: sin que nos detengamos a reflexionar acerca del significado de cada palabra, mantenemos un flujo lingüístico constante.

Piénsese en las reglas necesarias para jugar ajedrez. La primera vez que una persona se encuentra con las reglas del ajedrez, éstas contendrán ciertas palabras con usos y significados desconocidos; sin embargo, a medida que aumente el uso de las palabras y reglas, el juego se hará más familiar, conocido, y el entendimiento con respecto a ellas aumentará. A media que las reglas se usen en diferentes contextos y en diferentes situaciones concretas, adquieren mayor sentido y la persona se familiarizará cada vez más con su uso; a la vez que este uso será cada vez más diverso y rico. Finalmente, cuando la persona haya usado constantemente las reglas durante mucho tiempo, comenzará a jugar ajedrez sin necesidad de prestar especial atención a la formulación inicial de las reglas; lo mismo sucede cuando, después de aplicar lentamente una serie de reglas, comenzamos a conducir, a caminar, a amarrar los cordones de nuestros zapatos o a hablar en una lengua nativa o foránea, sin necesidad de prestar especial atención a los procedimientos que inicialmente nos permitieron una ejecución lenta y cuidadosa de estas acciones. Este estadio en que se actúa reflexionando de manera lenta y detallada acerca de las reglas, consiste en actuar *según* un manual; por el contrario, el estado en que se actúa de manera rápida y sin reflexión detallada consiste en actuar *de acuerdo* a las reglas (Ryle, 1967).

Actuar según y de acuerdo a las reglas

Rapaport (1995a) distingue entre actuar *según* las reglas y actuar *de a cuerdo a* las reglas. En el primer caso, actuamos lentamente, prestando especial atención a la formulación y aplicación de las reglas y

los procedimientos; esto sucede cuando ejecutamos una acción que es nueva para nosotros y, por lo tanto, debemos actuar *según* los procedimientos con que contamos. En el segundo caso actuamos rápidamente, sin prestar especial y detallada atención a las reglas que conocimos inicialmente; esto sucede cuando nos hemos acostumbrado a la ejecución de una acción que inicialmente pudo ser nueva y desconocida. Lo mismo sucede con el uso de un lenguaje que inicialmente pudo haber sido desconocido: al principio atendemos de manera lenta las reglas de su uso y luego dejamos de reflexionar y pensar en ellas. Cuando aprendemos a manipular nuestra primera lengua o una lengua foránea, y nos acostumbramos a ella por su constante uso, entonces comenzamos a actuar *de acuerdo* a las reglas. A continuación se presentan algunos aspectos relevantes de esta distinción.

Actuar según las reglas

Actuar *según* las reglas quiere decir que unas reglas ya formuladas se usan y aplican para ejecutar una determinada acción. Esto producirá una ejecución lenta porque el agente cognitivo debe detenerse constantemente a reflexionar acerca de los procedimientos y sus aplicaciones. Cuando aprendemos un segundo idioma, es común que al principio prestemos especial atención a las reglas de traducción de las palabras, de manera que cada vez que nos enfrentamos a una palabra de la nueva lengua, reflexionamos acerca de la traducción que debemos hacer de esa palabra a nuestra lengua nativa. Si nuestra lengua nativa es el español y nuestra segunda lengua será el inglés, entonces cuando escuchamos las primeras palabras en inglés, es común traducir cada palabra al español; reflexionaremos acerca de los posibles significados y usos de cada palabra y de esta manera construimos el sentido de la frase completa. Al principio usamos las palabras *según* las reglas que nos han sido dadas, en este caso, las posibles traducciones -que comúnmente se denominaban "significados"- que cada término puede tener en nuestro idioma nativo.

Si estamos aprendiendo inglés y mientras estamos conduciendo un auto alguien nos grita *be carefull, you are going to hit the car!* es probable que no reaccionemos instantáneamente, porque la representación mental inicial de estas palabras, así como su uso y significado, consiste únicamente en sus traducciones. Cuando escuchamos esta frase en inglés no nos representamos instantáneamente imágenes y recuerdos

de carros chocando o de incidentes nefastos que hayamos tenido; lo primero que nos representaremos, lentamente, serán las traducciones al español. En esta dinámica el inglés es el dominio sintáctico y el español es el dominio semántico; de esta manera, hacemos un mapeo entre términos de cada dominio y así comenzamos a formar el significado de la frase. Si estamos actuando *según* la reglas, el dominio semántico del inglés no será el mundo mismo, sino las palabras de la lengua española. Esto es lo mismo que le sucede a la persona que se encuentra dentro de la habitación china; como esta persona nunca ha usado el chino, entonces las primeras representaciones que ella se puede formar consisten únicamente en las reglas para el uso de los caracteres chinos; el chino juega el papel de dominio sintáctico y las instrucciones, que se encuentran en el idioma nativo, juegan el papel de dominio semántico. Al igual que nosotros cuando estamos aprendiendo un segundo idioma, para la persona que se encuentra dentro de la habitación el dominio semántico de los símbolos chinos no es el mundo, sino las reglas e instrucciones que se encuentran en su idioma nativo.

Actuar de acuerdo a las reglas

Searle (1995, p. 418) dice que la persona dentro de la habitación no entiende nada de chino, de lo cual se puede concluir que la persona no forma y no posee representaciones acerca de los símbolos chinos; esto se toma casi como una verdad a priori. No obstante, es poco probable que la persona dentro de la habitación china no se forme ningún tipo de representación ni entendimiento acerca de los símbolos pues, por lo general, cuando nos ponemos en contacto con una palabra o un objeto completamente desconocido, tendemos a establecer algún tipo de asociación mental que nos permite aproximarnos a un posible entendimiento de ese objeto. Esto no quiere decir que dicha asociación mental es inequívoca y convencionalmente acertada o correcta; del hecho de que la persona dentro de la habitación china forme algún tipo de asociación mental y representación de los símbolos chinos, no se sigue que estas representaciones coincidirán totalmente con los usos convencionales del idioma chino. Esta representación inicial, después de formarse, se revisará y ajustará a medida que aumente el uso de la palabra -o en este caso, de los símbolos-, hasta que dicha representación coincida con el entendimiento que comúnmente tiene todo hablante del chino acerca de los caracteres chinos.

Es probable que la persona dentro de la habitación china sí entienda algo acerca de esos símbolos y forme representaciones acerca de estos, lo que sucede es que no necesariamente entenderá lo mismo que entienden acerca de esos símbolos todas las personas que están familiarizadas con el uso del chino. Ahora bien, el hecho de que la persona no entienda en los símbolos lo mismo que entienden los hablantes del chino, no quiere decir que ella no forma algún tipo de entendimiento o representación acerca de los símbolos. Es probable que la persona dentro de la habitación sí forme representaciones mapeando ambos dominios; formará representaciones constituidas por los símbolos desconocidos -dominio sintáctico- y las reglas de manipulación -dominio semántico-. Con el paso del tiempo y el uso constante de las reglas, la persona dentro de la habitación china establecerá alguna asociación mental entre los símbolos chinos y las mismas reglas, que son los dominios a los que tiene accesos en su contexto. Nótese que en el caso de la persona que se encuentra dentro de la habitación china, ni siquiera es necesario hablar de un entendimiento sintáctico de un dominio inicial en términos de sí mismo, porque esta persona cuenta con dos dominios que permiten el mapeo y la construcción de representaciones, a saber, una lengua nativa y unos símbolos pertenecientes a una lengua desconocida.

A medida que la persona dentro de la habitación se familiarice con la dinámica de mapeo que hay entre las reglas y los símbolos chinos, las reglas se compilarán hasta el punto de que estas puedan omitirse al momento de procesar los símbolos (Hofstadter en Rapaport 1995a, p. 272). Probablemente, la persona comenzará a asociar rápidamente los símbolos chinos con los recuerdos de las anteriores ocasiones en que los ha usado y con las reglas para manipularlos, tal como nosotros lo hacemos cuando escuchamos o leemos una palabra. Es probable que la persona no forme las representaciones que comúnmente formaría todo hablante del chino; esto, sin embargo, se debe a las condiciones del contexto. Seguramente, si nosotros tuviéramos que aprender un idioma en total aislamiento del contexto, formaríamos representaciones distintas acerca de las palabras de nuestro idioma. Esto implicaría que, para que la persona dentro de la habitación pudiera formar las representaciones que comúnmente todo hablante del chino se forma, entonces debería tener la oportunidad de ponerse en contacto con una contexto igual al que tienen acceso las personas que hablan chino en condiciones normales, es decir, un contexto mucho más rico en estímulos.

Lo anterior no quiere decir que la persona dentro de la habitación china entenderá chino tal como lo hace cualquier otra persona porque, a diferencia de cualquier persona que se encuentra por fuera de la habitación, aquella que está encerrada tiene un contexto pobre, carente de estímulos sensoriales y, por lo tanto, no podrá establecer asociaciones entre el mundo -aromas, colores, sabores, texturas, sonidos y, con esto, figuras y objetos- y los símbolos que manipula. Si a la persona que se encuentra dentro de la habitación china, entre los fajos que manipula, se le proporcionan dibujos, colores, formas y figuras, que guarden cierto parentesco convencional en cuanto a su uso,[12] es posible que con el paso del tiempo establezca representaciones "correctas". Ahora bien, si las posibilidades de estímulos aumentan, de manera que la persona dentro de la habitación no sólo tenga acceso a fajos de manuscritos sino a un registro de sonidos, de estímulos visuales -mediante cámaras, por ejemplo-, de aromas, de sabores y de texturas, entonces el significado de los símbolos chinos se podrá enriquecer, de manera que cuando la persona vea los símbolos que significan la palabra perro, forme una representación compleja constituida por recuerdos de aromas, texturas y colores. Así, es muy probable que con el paso del tiempo y la constante manipulación de símbolos, la persona dentro de la habitación china constituya al mundo como dominio semántico de los símbolos que manipula. Esto quiere decir que la persona dentro de la habitación dejará de actuar *según* las reglas y comenzará a actuar *de acuerdo* a las reglas; una persona en este mismo caso, cuya lengua nativa sea el español y que esté aprendiendo inglés, que escuche la frase *be carefull, you are going to hit the car!*, no tendrá que reflexionar lentamente acerca de las reglas de traducción sino que, como el dominio semántico ya es el mundo, podrá reaccionar rápidamente y asociar las palabras con objetos específicos del mundo.

Recuérdese que la cuestión inicial de Searle era que la persona dentro de la habitación no entendería absolutamente nada de los símbolos que manipula, sin embargo, hemos visto que aquella persona sí entenderá algo acerca de esos símbolos, lo que sucede es que no entenderá lo mismo que entendería un hablante del chino por fuera de la habitación. Con la constante manipulación de los símbolos, las reglas de manipulación se internalizarán y finalmente

12 Esto quiere decir que, cuando se escriban los símbolos que significan perro, se presente un dibujo o una foto de un perro y no de un gato.

la persona formará algún tipo de representación. Nosotros mismos experimentamos el paso de la etapa en la que aplicamos reglas y no entendemos muy bien el significado o resultado de dicha aplicación, a la etapa en la que las reglas adquieren tanto sentido que se vuelven parte de nuestra vida mental. Searle puede decir que Salcedo-Albarán y De León-Beltrán no entendieron el punto central de su argumento, a saber, que la persona dentro de la habitación puede responder preguntas sin entender absolutamente nada. Al respecto, es necesario precisar que si la cuestión del entendimiento de la persona dentro de la habitación gira en torno a, únicamente, el entendimiento del chino convencional, entonces efectivamente Searle siempre saldrá victorioso ante cualquier réplica. Absolutamente nadie puede aprender y entender un nuevo idioma si está aislado dentro de una habitación. En este caso, absolutamente ningún sistema físico -humano o robot- que aplique reglas, podría llegar a entender algo acerca de la aplicación de dichas reglas. Pero todos sabemos que este no es el caso pues, cuando aplicamos reglas, con el paso del tiempo las reglas adquieren sentido hasta formar parte de nuestra vida. Entonces, ¿Por qué podemos pasar de la sintaxis a la semántica? ¿Esto sucede únicamente en virtud de la composición biológica de nuestro cerebro? En realidad no hay sustento válido para pensar que este sea el caso, a pesar de que pueda parecernos una respuesta correcta si tenemos en cuenta que la única consciencia de la que podemos estar seguros es de la nuestra, la cual está construida con materiales orgánicos. Pero la respuesta de que la aplicación de reglas genera semántica en la mente humana únicamente como resultado de su biología, en realidad, es prácticamente inconexa del hecho de que, a partir de la aplicación de reglas, podemos generar semántica. Bien podría responderse que la cantidad y el tipo de dientes que poseen los humanos es el motivo por el cual generamos semántica; de hecho, el argumento funciona con cualquier característica única de los seres humanos. Por este motivo, Searle reconoce que algunos simios superiores, con cerebros similares al humano, poseen consciencia. Ahora, remplácese "cerebros" por "dientes" y el argumento seguirá igual de victorioso. Claro, suena más lógico hablar de "cerebros" y no de "dientes" porque los cerebros intervienen más en la consciencia que los dientes; no obstante, el punto es que si buscamos una característica particular de los seres humanos y a dicha característica le atribuimos la generación de semántica, incluso a partir de la aplicación de reglas, entonces no habrá forma de refutar el argumento.

En la medida en que se establezca una relación de simetría entre lo que acontece en el campo de los símbolos -sintaxis- y lo que acontece en el campo de los significado -semántica-, mediante el establecimiento de los mapeos correspondientes, el cerebro y el computador pueden interpretarse como *motores* que posibilitan y preservan las propiedades semánticas de un lenguaje -y, por lo tanto, de las representaciones- en virtud de los códigos sintácticos que permiten relaciones causales entre símbolos y que permiten relaciones causales entre representaciones. Esto no es contrario a la idea de Fodor pues, como se señaló líneas atrás, el contenido semántico de una actitud proposicional resulta del contenido semántico de los símbolos mentales que, a su vez, están estrechamente relacionados con la física del cerebro. De esta manera, la internalización permitirá el paso de la sintaxis a la semántica, lo cual requiere un concepto fuerte de implementación y computación.

Concepto fuerte de implementación y computación

Según Chalmers (1994), el concepto de computación es claro en el ámbito de las matemáticas, donde se interpreta desde una perspectiva abstracta, tal como Searle lo reconoce. Sin embargo, este no es el único concepto de computación posible y, por lo tanto, es necesario hallar un vínculo estable entre la computación matemática y la computación relacionada con la ciencia cognitiva. Según Chalmers (1994), una teoría de este tipo consiste en una correcta teoría de la *implementación*. Esta *implementación* consiste en establecer una relación entre la concepción abstracta matemática de computación y un sistema físico; esto es lo que generalmente se conoce como la *realización* en un sistema físico.

Searle asegura que cualquier sistema físico puede interpretarse como un sistema con posibilidades computacionales; sin embargo, para Chalmers (1994), no es correcto sostener una postura tan permisiva. Generalmente, se habla de computaciones relacionadas con algún formalismo; según Chalmers (1994), las máquinas de Turing, los programas de Pascal, los autómatas celulares y las redes neuronales son algunos ejemplos de formalismos. Todos estos, a su vez, se pueden interpretar como el tipo de formalismos que consisten en

Autómatas de Estados Combinatorios o "CSAs" por sus siglas en inglés[13]; sin embargo, también es posible encontrar Autómatas de Estados Simples Finitos, o "FSAs" también por sus siglas en inglés[14]. Por su parte, los FSAs están especificados en términos de un conjunto de *inputs* $I_1,...,I_k$, un conjunto de estados internos $S_1,...,S_K$ y un conjunto de *outputs* $O_1,...,O_K$; estos conjuntos se relacionan mediante estados de transición que consisten en$(S,I)^{\circledast}(S',O')$.[15]

Ahora bien, la principal diferencia entre el FSA y el CSA consiste en la especificación de su estado interno, pues en el CSA dicho estado consiste en un vector $[S^1, S^2, S^3,\ldots,S^n]$ y no en una etiqueta S simple. Según Chalmers (1994), los elementos que componen el vector pueden interpretarse como las células en un autómata celular:

> Hay un número finito de posibles valores S^i, donde S^i_j es el j-ésimo posible valor para el i-ésimo elemento. Estos valores pueden interpretarse como subestados […]. Las reglas de los estados de transición se determinan especificando, para cada elemento del vector estado, una función por la cual el nuevo estado depende del antiguo total del vector estado y el vector input, y lo mismo para cada elemento del vector output (Chalmers 1994, p. 5).

Así, la principal distinción entre un CSA y un FSA consiste en que el sistema físico que implementa el CSA cuenta con unos estados internos especificados en un vector, de manera que cada elemento del vector pueda corresponderse con cada elemento independiente del sistema físico; de esta manera, el sistema físico puede dividirse en varias regiones funcionales, con lo que cada estado interno se relaciona con cada una de las regiones.

13 Combinatorial-State Automata.

14 Finite -State Automata.

15 Un sistema físico P implementa un FSA M si hay un mapeo de estados internos de P en estados internos de M, inputs de P en estados de entrada de M, y outputs de P en estados de salida de M, de de la siguiente forma:

> "[…] para cada relación de estado de transición $(S,I)^{\circledast}(S',O')$ de M, la siguiente condición se mantiene: si P es un estado interno s y recibe input i donde f(s)=S y f(i)=I, esta relación permite la entrada al estado interno s´ y produce un output o´ tal que f(s´)=S´y f(o´)=O´. (Chalmers 1994, p. 4).

La conclusión de Chalmers (1994), con respecto al análisis de las FSAs y las CSAs es que la relación entre el sistema físico en que se implementa la computación, y la computación como tal, es un isomorfismo de una relación entre la estructura formal de las computaciones y la estructura causal del sistema; esto quiere decir que la computación "es simplemente una especificación abstracta de la organización causal" (Chalmers 1994, p. 7); es decir, una especificación abstracta que guardará cierta relación con las relaciones causales de los sistemas físicos que puedan implementarla. Esto no quiere decir que cualquier sistema físico pueda instanciar cualquier tipo de computación. Por ejemplo, las condiciones necesarias para implementar una computación compleja como la de un CSA con un vector compuesto de 1000 elementos, 10 posibilidades para cada elemento de entrada y de salida y complejas reglas de transición de estados, serán lo suficientemente robustas como para que cualquier sistema físico, como una pared, no pueda ejecutarlo. Esto sucede porque en el caso de las CSAs no sólo se requiere un mapeo entre los estados del sistema en estados del CSA, sino que se requiere que los mapeos de los estados puedan satisfacer las reglas de estados de transición. No hay razones para suponer que cualquier sistema físico puede implementar este tipo de computación; de hecho, "la posibilidad de que un conjunto de estados arbitrarios satisfaga estas limitaciones es algunas veces menor que uno en $(10^{1000})^{10^{1000}}$ "(Chalmers 1994, p. 7). De igual manera, Bechtel (1985), Churchland y Sejnowski (1990) y Churchland (1999) han llamado la atención acerca de la proporcionalidad requerida entre el soporte físico y el soporte lógico de un sistema. Esto quiere decir que el soporte físico debe ser lo suficientemente complejo en términos causales para permitir el almacenamiento y ejecución de un determinado programa o porción de información.

Ahora bien, con respecto a la posibilidad de que un sistema físico determinado pueda implementar más de una computación, la respuesta de Chalmers (1994) es positiva, pues un sistema físico que esté implementando una computación compleja, a la vez estará implementando múltiples computaciones simples; así, se puede asegurar que al interior de cada sistema físico hay múltiples computaciones implementadas (Chalmers 1994, p. 7). Pero, como queda claro hasta ahora, del hecho de que al interior de un sistema físico se pueda implementar múltiples computaciones, no se sigue que se pueda

implementar *cualquier* computación; esto quiere decir que, si bien un sistema físico puede implementar múltiples computaciones, la mayor parte de las computaciones pueden implementarse en una cantidad muy limitada de sistemas físicos (Chalmers 1994, p. 8). La conclusión es que no cualquier sistema físico puede implementar cualquier computación; así, cuando Searle asegura que incluso el fenómeno de la digestión podría interpretarse como un sistema físico que implementa computaciones cognitivas, en realidad está confundiendo el hecho de que solamente ciertas computaciones complejas -por las definiciones de los estados de transición entre las relaciones causales del sistema y por los estados de entradas y salidas- pueden interpretarse como cognitivas y por lo tanto, implementadas en ciertos sistemas físicos. Algunas computaciones no necesariamente tienen que interpretarse como cognitivas porque no todas las computaciones deben guardar similitud con las relaciones causales del sistema físico del cerebro ni con los formalismos de los estados mentales. No cualquier programa, por el hecho de ejecutarse en virtud de *inputs*, *outputs* y *estados internos*, es cognitivo. Esto quiere decir que solamente ciertas computaciones complejas que determinan transformaciones estructurales del sistema nervioso -bien sea sistema nervioso de silicio o de carne- pueden entenderse como cognitivas y, a su vez, estas computaciones solamente podrán implementarse en ciertos sistemas físicos.

Tal como lo concibe Rapaport, la constitución de representaciones es un resultado de la implementación de la computación; no es necesario sostener que hay un elemento propiamente semántico en la definición de un programa y su computación: "es así como debería ser: las computaciones son específicamente sintácticas, no semánticas" (Chalmers 1994, p. 9); sin embargo, el contenido semántico resultará de la ejecución de la computación aunque no sea necesario buscarlo en la definición de la computación. Cuando los diseñadores de computadores se percatan de que hay una correspondencia entre la máquina y el programa que dicha máquina puede implementar, no se preocupan acerca del contenido semántico del programa pues este es una consecuencia de la implementación, solamente se preocupan de que el mecanismo tenga las correctas relaciones causales (Chalmers 1994, p. 9).

El hecho de que no cualquier computación puede implementarse en cualquier sistema físico se refleja, de manera intuitiva, en la

relación que ha guardado el desarrollo de los computadores comerciales con el desarrollo de los programas. El tipo de *software* que se podía implementar en los primeros computadores es muy distinto al tipo de *software* que se puede implementar en los computadores actuales pues, a medida que aumenta la velocidad y capacidad de procesamiento, así como la velocidad y capacidad de almacenamiento, hay mayores posibilidades de procesamientos más complejos, esto es, computaciones más complejas, *ergo*, el desarrollo del *software* va acompañado de la proporcionalidad correspondiente con el desarrollo del *hardware*.

Lo anterior refuerza la interpretación de la semántica como un espejo que guarda simetría con la sintaxis. En la medida en que la implementación de computaciones consiste en una serie de relaciones causales que guardan proporción con los estados de relación específicos y con las relaciones causales del sistema en que se implementa, no se puede asegurar que la computación es una simple abstracción matemática en el ámbito de la ciencia cognitiva. Un diseño computacional correcto compartirá las relaciones causales del sistema que se modela y, por lo tanto, la implementación de computaciones cognitivas no consistirá solamente en un modelo de cogniciones reales, sino que consistirá en una replicación de dichas cogniciones. De esta manera, Chalmers (1994) concluye que Searle está confundiendo *programas* con *implementaciones* de los programas:

> Mientras los programas por sí mismos son objetos sintácticos, las implementaciones no lo son: estos son sistemas físicos reales con organización causal compleja, con causación física real dentro de ellos. En un computador electrónico, por ejemplo, los circuitos y los voltajes chocan entre sí de manera análoga a como las neuronas y las activaciones chocan entre sí. Es precisamente en virtud de esta causación que las implementaciones pueden tener propiedades cognitivas y, por lo tanto, semánticas (Chalmers 1994, p. 16).

Esto no quiere decir que la sintaxis y la semántica sean idénticas, o que haya una confusión entre el nivel del *software* y el nivel del *hardware*; esto sólo quiere decir que mientras los programas son sintácticos, sus implementaciones son semánticas. Sin embargo, para Chalmers (1994) es claro que continúa presente la cuestión de cómo el computador pasa del procesamiento sintáctico al contenido semántico (Chalmers 1994, p. 16); pero esta cuestión es idéntica a la pregunta acerca de cómo los cerebros constituyen este mismo

tipo de contenido. Así, no basta con asegurar que los cerebros poseen semántica y los computadores no, sino que es necesario observar los tipos de implementaciones que se dan en cada uno de estos sistemas físicos; "la sintaxis puede no ser suficiente para la semántica, pero el tipo correcto de causación sí lo es"(Chalmers 1994, p. 16).

¿La persona dentro de la habitación china, verdaderamente debe entender chino?

Hasta aquí se ha mostrado la posibilidad de que la persona dentro de la habitación china entienda chino gracias a la costumbre y a la compilación de sus procesamientos. Esto plantea una pregunta interesante: ¿Es verdaderamente necesario que la persona dentro de la habitación china entienda los símbolos chinos, para que se pueda concluir que la instanciación de un programa puede constituir entendimiento? Searle describe un sistema que en conjunto es una máquina de cómputos y concluye que la persona dentro de la máquina no entiende; frente a esto parece válido preguntar: ¿Qué artefacto, dentro de un computador, cumple el papel de la persona? ¿Es la memoria virtual, la memoria de procesamiento, la memoria fija o el cableado de datos lo que debe entender los ceros y unos de los programas? Esta pregunta, a su vez, se puede complejizar si se cuestiona qué es lo que hay detrás de la piel de una persona, que cumpla el papel de la persona dentro de la habitación china: ¿Es el cerebro, las neuronas o las conexiones sinápticas? En el siguiente capítulo se examinarán los niveles de la analogía que se pueden establecer entre la habitación china y un computador y, por lo tanto, entre la habitación china y la mente humana.

Sumario

a). El entendimiento *semántico* consiste en una correspondencia entre un dominio A y un dominio B. El agente cognitivo entiende uno de los dominios, en este caso el dominio A, en términos del otro dominio, el dominio B. Esto implica que el dominio A es desconocido y se interpreta como puramente sintáctico -este es el caso de los símbolos que procesa la persona dentro de la habitación- y el dominio B es conocido

y se interpreta como semántico -este es el caso de la lengua nativa de la persona que se encuentra dentro de la habitación-.

b). El mapeo entre el dominio sintáctico y el dominio semántico se puede construir en una red de presentación semántica.

c). Entender X significa que:

(c.1.) X se entiende *relativamente* a algo. Este es el caso de la semántica por correspondencia, en la que se establecen correspondencias entre un dominio desconocido y un dominio conocido. El mapeo que se da en este tipo de entendimiento semántico por correspondencia, consiste en asociaciones sintácticas.

(c.2.) Uno se acostumbra a usar X. En este caso no hay remisión entre dos dominios, sino que un sólo dominio se entiende en términos de sus partes. Este es el caso del entendimiento puramente sintáctico.

d). Tanto en el entendimiento semántico por correspondencia, como en el entendimiento sintáctico, la constante aplicación de reglas de manipulación permite la constitución de representaciones.

e). El entendimiento sintáctico permite el entendimiento del primer dominio, en términos de sus propias partes. Esto representa un tipo de autorreferencialidad que es coherente con el funcionamiento del sistema nervioso. La circularidad del entendimiento sintáctico puede interpretarse como perteneciendo a un sistema autopoietico y, por lo tanto, consecuente con la naturaleza de la mente humana.

f). En el sistema nervioso autopoietico interviene un programa que permite la generación de *inputs*. El ADN es el programa que en el caso de los seres vivos, especifica las reglas de transformación internas de la estructura, las cuales, a su vez, permiten la constitución del *input*. Esto quiere decir que el *input* también es un resultado de la autorreferencialidad del sistema nervioso y, por lo tanto, la postulación de un programa es indispensable para explicar la consciencia humana.

g). En el experimento mental de la habitación china hay un caso de entendimiento semántico por correspondencia porque la persona que se encuentra dentro de la habitación -acerca de la cual se hace la

exigencia de entendimiento-, cuenta con dos dominios: un dominio sintáctico, que son los símbolos chinos, y un dominio semántico, que es su lengua nativa.

h). Si bien el programa que ejecuta la persona dentro de la habitación es puramente sintáctico, la implementación de dicho programa permite la construcción de contenido semántico. No es necesario buscar un contenido semántico en el nivel de definición del programa, sino en el nivel de implementación.

i). La constante aplicación de las reglas de procesamiento, permite que la persona que se encuentra dentro de la habitación establezca una relación de correspondencia entre los símbolos chinos y su lengua nativa. Cuando esto sucede, de manera rápida e inconsciente, se puede decir que la persona dentro de la habitación ha *compilado* las reglas. Esto mismo sucede cuando, al aprender un segundo idioma, establecemos traducciones de manera lenta e inconsciente y luego, cuando las reglas de traducción se han compilado, "pensamos" en el nuevo idioma, es decir, lo procesamos rápida e inconscientemente.

j). El mapeo de dominios, en tanto implementación ejecutada en un sistema físico adecuado, permite la formación de representaciones y contenido semántico.

Capítulo 4:

Cerebro e intencionalidad

En el capítulo anterior se mostró la posibilidad de que, en algún momento, la persona encerrada en la habitación china entienda o represente algo acerca de los símbolos chinos; después de que la persona dentro de la habitación manipule constantemente los símbolos, compilará las reglas y formará alguna representación acerca de los símbolos chinos. Esto no quiere decir que la persona dentro de la habitación entenderá chino en el mismo sentido en que cualquier persona lo entiende, sin embargo, del hecho de que no entienda lo mismo que entiende cualquier nativo chino, no se sigue que en estricto sentido no entienda nada. Básicamente, la persona dentro de la habitación china experimentará algo similar a lo que cualquier persona experimenta cuando está aprendiendo un nuevo idioma; al principio, el idioma se manipula de manera lenta y prestando especial atención a las reglas de traducción; luego, cuando la persona se acostumbra a las reglas, el nuevo idioma se manipula de manera rápida e inconscientemente. La diferencia en lo que entiende la persona dentro de la habitación y una persona común que aprende un nuevo idioma, radica en la riqueza del contexto en que se encuentra.

No es nula la posibilidad de que la persona dentro de la habitación entienda el chino tal como lo entiende cualquier hablante nativo del chino. Si la persona dentro de la habitación accede a los objetos del mundo, esto es, a aromas, formas, imágenes, sabores, sonidos y texturas, que pueda relacionar con los símbolos, es muy probable que forme representaciones acerca de los símbolos; de hecho, representaciones bastante similares a las representaciones que formaría cualquier hablante del chino. La persona simplemente comenzaría a relacionar cada símbolo o serie de símbolos con los estímulos y objetos correspondientes. Esta posibilidad está relacionada con la réplica del robot expuesta en el segundo capítulo, porque da a la persona que se encuentra dentro de la habitación, la posibilidad de acceder al mundo externo. A esta réplica, Searle responde que si la persona que se encuentra dentro de la habitación se colocara dentro del robot, y los fajos que recibiera correspondieran a los estímulos provenientes de las cámaras de televisión, entonces la persona continuaría sin entender alguno de los símbolos que procesa. Esto quiere decir que si la persona dentro de la habitación se coloca dentro del robot, entonces la persona jugaría el papel del cerebro o de algún elemento cerebral. Ahora bien, si la persona dentro de la habitación es el cerebro, la neurona o la sinapsis, entonces ¿Es correcto exigirle que entienda chino como lo hace cualquier persona que se encuentra por fuera de la habitación? ¿Es correcto afirmar que nuestros cerebros entienden chino, inglés, o cualquier otro idioma, de manera idéntica a como lo entiende cualquier hablante nativo del chino o del inglés? Si los cerebros no pueden entender un idioma, tal como lo entiende cualquier persona, entonces ¿Qué tan relevante es el hecho de que la persona que se encuentra dentro de la habitación entienda o no entienda, se represente o no se represente algo acerca de los símbolos?

Para Searle, la persona dentro de la habitación no entiende chino y esto implica que un computador o un robot carecen de entendimiento, *ergo*, si la persona dentro de la habitación entendiera chino, un computador o un robot podrían también entender algo. De esto resulta la exigencia de que una de las partes del robot, y no el robot en conjunto, tenga que entender lo que se está procesando; exigencia que se examinará en el desarrollo del presente capítulo.

Cuando en IA se crea un robot que imita el entendimiento humano, lo común es que se exija que ese robot *entienda*, tal como

cualquier persona entiende; no se exige que una parte específica del robot entienda. No se exige que la memoria de almacenamiento o los dispositivos de procesamiento entiendan, cada una, lo que está procesando; se exige que el sistema en conjunto, lo haga. Igualmente, cuando hacemos una exigencia de entendimiento a un ser humano, la exigencia se hace acerca de toda la persona y no de su cerebro o una de sus neuronas; si abrimos la cabeza de una persona y sacamos su cerebro, no sería correcto verificar si allí podemos encontrar el mismo tipo de entendimiento que encontramos en la persona que es portadora del cerebro. En el planteamiento original del juego de la imitación, Turing pretende que el agente de IA -lo que él llama *máquina-* cumpla el papel de una persona y no de una de una de las partes de la persona; parece prudente examinar la exigencia de entendimiento que hace Searle, sobre una de las partes del agente y no sobre el agente completo.

Aunque Turing deja claro que en el juego de la imitación debe participar un agente cognitivo de IA y no una de las partes del agente, si se acepta que la habitación china en conjunto es una analogía de la máquina computadora, entonces Searle estaría exigiendo entendimiento a una de las partes del agente y no a todo el agente. Esto se percibe en su respuesta a la réplica del robot: Searle no acepta la posibilidad de que el robot en conjunto entienda chino, sino que continúa exigiendo que una de las partes computacionales del robot entienda chino. Esta exigencia es similar a exigir que uno de los órganos del entramado cerebral humano entienda chino, pero ¿Qué parte, de todo el entramado mental humano, se supone que debe comprender lo que leemos o escuchamos?

Para Searle, en el planteamiento original del experimento mental, la *persona/computadora*, es decir, la persona que computa dentro de la habitación, es un agente cognitivo de IA y, por esto, se exige que entienda un idioma. Sin embargo, en la respuesta a la réplica del robot, Searle interpreta a la *persona/computadora* como una parte del agente cognitivo. A continuación se examinan las consecuencias de interpretar a la persona dentro de la habitación como agente cognitivo o como parte del agente cognitivo, para verificar en qué caso la exigencia de entendimiento es correcta o incorrecta.

Desentramando la analogía: El micro-nivel y el macro-nivel de la habitación china

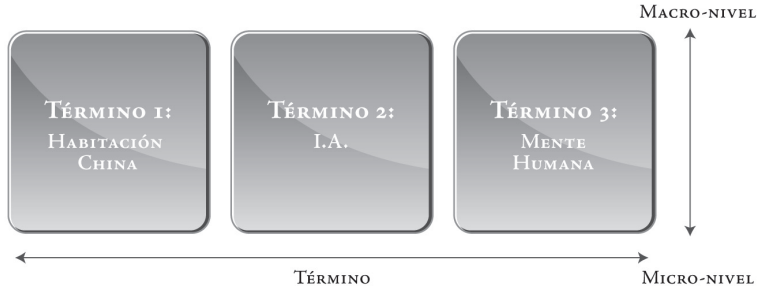
En el planteamiento del experimento mental de la habitación china, Searle asegura que la persona dentro de la habitación se llama *computador* y que ese computador está instanciando un programa. ¿Qué quiere decir Searle con computador? ¿Está hablando de un agente de IA o de una de las partes computadoras que puede constituir a un agente de IA? En ambos casos se podría hablar de un sistema o de un dispositivo que ejecuta computaciones. Para responder a esta pregunta apelaré a la distinción que Searle hace entre rasgos de macro-nivel y rasgos de micro-nivel de la mente humana.

Según Searle, en los sistemas físicos se pueden distinguir micro y macro-propiedades. Con la distinción entre micro-nivel y macro-nivel se puede diferenciar la interpretación de la persona que se encuentra dentro de la habitación como participante de un rasgo de micro-nivel o como participante de un rasgo de macro-nivel. Si la persona que se encuentra dentro de la habitación forma parte de un rasgo de micro-nivel, entonces el experimento mental de la habitación china sería una analogía del cerebro o de una de las partes que ejecutarían computaciones dentro de un agente cognitivo de IA; por otra parte, si la persona que se encuentra dentro de la habitación participa de un rasgo de macro-nivel, entonces, el experimento mental de la habitación china sería una analogía de una persona o de un agente cognitivo de IA, es decir, un sistema completo que ejecuta computaciones.

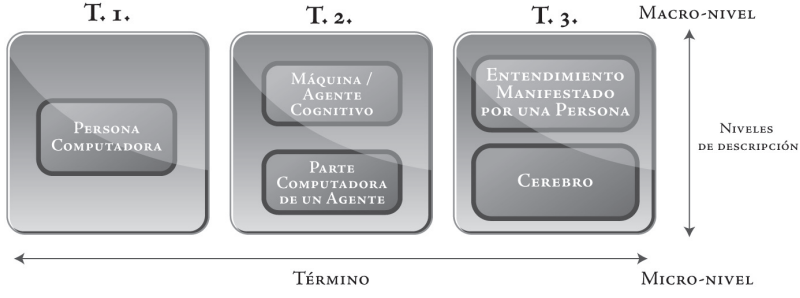
Si bien la analogía puede interpretarse en dos niveles, ésta no consta de dos términos solamente. La habitación china puede interpretarse como la analogía de una máquina y, como en la IA una máquina puede interpretarse como la analogía de la mente humana, entonces lo que se diga acerca de una máquina, vía pronunciamientos acerca de la habitación china, puede extenderse a la naturaleza de la mente humana. Esto quiere decir que, en realidad, la analogía consta de tres términos: a). habitación china, b). IA y c). mente humana; las observaciones acerca de la habitación china son observaciones acerca de la IA y estas, a su vez, pueden ser observaciones acerca de la mente humana. Nótese que algunos de estos términos cambian en la medida en que la analogía se inter-

prete desde los rasgos de micro-nivel o desde los rasgos de macro-nivel. A continuación presentamos un gráfico que permite distinguir los términos y los niveles de la analogía. Este gráfico cambiará en la medida en que se opte por interpretar la analogía desde los rasgos de micro o macro-nivel. En la gráfica 2 aparecen las posibilidades que ofrece interpretar la analogía desde el micro o el macro-nivel.

Gráfica 1. Términos y niveles de la analogía



Gráfica 2. Términos y niveles de la analogía desde el micro o el macro-nivel



Analogía de macro-nivel: Si Searle entiende por computador un agente cognitivo de IA

Un agente cognitivo de IA puede ser un robot o un programa de computador. En general, un agente cognitivo de IA es aquel sistema al

que los teóricos de la IAF suelen adscribir intencionalidad, consciencia o entendimiento. Este sistema físico suele denominarse *máquina*. Ahora bien, en el test de Turing, se propone que una máquina imite a una persona común, por lo tanto la máquina se encuentra, en la analogía, al mismo nivel de términos del entendimiento manifestado por una persona corriente:

> Preguntemos ahora: "¿Qué sucedería si una máquina tomara el papel de A en este juego?" ¿Se equivocaría el examinador con la misma frecuencia si los participantes fueran un hombre y una mujer? Estas preguntas reemplazarán nuestra pregunta original: ¿puede pensar una máquina? (Turing, 1950/1990).

La propuesta de Turing (1950) permite pensar que en la IA se adscribe verdadero pensamiento y entendimiento a un agente cognitivo que pueda desempeñar tareas de pensamiento que, por lo general, consideramos como exclusivas de la mente humana. Turing (1950) no pretende que un chip o una de las partes computadoras de la máquina tome el lugar de la mujer en el juego de la imitación, sino que una máquina completa, esto es, un agente cognitivo completo, tome el papel de A, que es una mujer. Con esto, tenemos que para una interpretación de macro-nivel, la gráfica 1 se puede alterar de la siguiente manera:

Gráfica 3. Términos de la analogía en el macro-nivel

Esto, a su vez, permite suponer que el argumento de Searle está dirigido a refutar la idea de que una máquina, como la que Turing (1950) propone en el juego de la imitación, puede poseer y manifestar verdadero entendimiento; de hecho, se asegura que el experimento mental de la habitación china muestra que una máquina puede superar el test de Turing y, no obstante, carecer de verdadero entendimiento. Así, en esta interpretación del experimento mental de la habitación china, como argumento que refuta la posibilidad de pensamiento genuino

en las máquinas, nos encontramos al nivel de agentes cognitivos: un agente cognitivo de IA y un agente cognitivo de pensamiento humano. Si se interpreta de esta manera el propósito del experimento mental de la habitación china, entonces es necesario entender a la *persona/computadora* que se encuentra encerrada en la habitación, como un agente cognitivo de IA. En este orden de ideas, la conclusión que resulta del experimento mental de la habitación china será: en la medida en que la persona encerrada en la habitación no entienda el chino, entonces ninguna máquina -en tanto agente cognitivo de IA- puede poseer y manifestar verdadero entendimiento.

Los términos de comparación

Establezcamos los términos de esta analogía desde rasgos de macro-nivel. Por una parte, tenemos que la persona dentro de la habitación se interpreta como el agente de IA. Esta sería la parte A de la analogía (AA). Por otra parte, tenemos que en el juego de la imitación se pretende que el agente de IA -la máquina- esté al mismo nivel de una persona común y corriente; de hecho, el agente de IA puede reemplazar a una mujer. Esta sería la parte B de la analogía (AB). Tanto AA como AB se encuentran al mismo nivel de descripción porque no estamos hablando de partes del agente, natural o artificial, sino de los agentes completos: la persona dentro de la habitación, el agente cognitivo de IA y la persona normal afuera de la habitación, se encuentran en el nivel de descripción de macro-nivel. Al unir la parte AA y la parte AB de la analogía, el experimento mental de la habitación china puede interpretarse como una analogía entre la persona dentro de la habitación *-persona/computadora-*, un agente cognitivo de IA y una persona común; esto explica por qué Searle aplica la misma exigencia tanto a la *persona/computadora* como a la persona nativa.

Ahora bien, como la *persona/computadora* y la persona común se encuentran en el mismo nivel de descripción, entonces las exigencias que se hagan sobre una deben ser procedentes para la otra. Esto quiere decir que la exigencia que se hace sobre la *persona/computadora* debe ser una exigencia correcta para la persona común. Si la cuestión importante es que la persona dentro de la habitación pueda entender chino, entonces sería necesario que ella contara con las herramientas necesarias para entender chino, al menos, las herramientas que necesitaría un hablante nativo del chino, para aprender inglés. La *persona/computadora* dentro de la habitación tiene una lengua nativa que es el

inglés, y la persona común, afuera de la habitación, tiene otra lengua nativa que es el chino. Cada una de estas personas posee un idioma nativo y se encuentra ante la exigencia de aprender un segundo idioma; por este motivo, se puede preguntar: ¿Bajo qué condiciones se le exigiría a la persona común aprender y entender un segundo idioma? Si esto mismo se puede exigir a la persona común bajo las mismas condiciones en las que se le hace dicha exigencia a la persona dentro de la habitación, entonces la exigencia de entendimiento que hace Searle es correcta; si por el contrario, las condiciones bajo las cuales se le podría exigir a la persona común que dominara un segundo idioma son distintas a las condiciones bajo las cuales se le exige a la persona dentro de la habitación que entienda chino, entonces la exigencia de entendimiento que hace Searle para ser incorrecta.

Aprendiendo un segundo idioma

La *persona/computadora* dentro de la habitación y la persona común, afuera de la habitación, se encuentran en el mismo nivel descriptivo; no obstante, la exigencia de entender un segundo idioma sólo se hace sobre la persona computadora. ¿Qué sucedería si encerráramos a la persona china, dentro de una habitación con una ranura de entrada y una ranura de salida, con unos fajos de instrucciones escritos en chino que le indican cómo aparear unos símbolos que, sin que ella sepa, pertenecen a la lengua inglesa? ¿Qué sucede si unas personas inglesas, afuera de la habitación, le entregan un fajo con caracteres ingleses para ejecutar las manipulaciones correspondientes que están escritas en chino, y luego las personas inglesas reciben un fajo con el resultado de las manipulaciones que, sin que el chino lo sepa, son respuestas a preguntas sobre relatos? ¿En este caso sería correcto exigirle a aquella persona china nativa que, tan pronto fue encerrada y manipuló símbolos, entienda inglés como cualquier inglés nativo? Creemos que no.

Es probable que aquel chino, encerrado y aislado del mundo, en una situación idéntica a la que Searle plantea para la *persona/computadora*, no entienda inglés. Esto no quiere decir que aquel chino nunca llegará a entender inglés; sin embargo, es muy poco probable que lo haga en estas condiciones. Se puede asegurar que, al principio, aquel chino nativo formará algún tipo de representaciones acerca de estos símbolos ingleses que para él son desconocidos; esto es muy probable porque al menos cuenta con un dominio conocido para él, a saber, su lengua chi-

na nativa y "los símbolos de una lengua natural que entendemos ponen en marcha diversos tipos de actividad mental" (Boden 1990, p. 114); una de estas actividades, es la formación de representaciones.

Asimetría en la exigencia de entendimiento

Para que le pudiéramos exigir a la persona nativa china que entienda los símbolos de la lengua inglesa, lo menos correcto sería encerrarlo en una habitación, aislarlo del mundo y proporcionarle material escrito en inglés sin, al menos, decirle que esos símbolos pertenecen a una lengua denominada "inglesa"; por lo que se sabe, solamente se puede entender una segunda lengua con procedimientos distintos a los descritos por Searle.

Lo anterior no quiere decir que la sintaxis sí es suficiente para la semántica, tan sólo quiere decir que la exigencia de que la *persona/computadora* entienda chino tal como lo hace cualquier hablante nativo del chino, es una exigencia asimétrica en la analogía y, en cierta manera, incorrecta. Una exigencia simétrica sería exigir que la *persona/computadora* entienda chino, en condiciones similares a las que serían necesarias para que la persona común pudiera entender una segunda lengua. Es muy probable que encerradas, ninguna de las dos personas entienda un nuevo idioma y esto, por supuesto, no obedece a que la sintaxis no sea suficiente para la semántica, sino a que a ninguna de ellas se les ha dado las herramientas necesarias para aprender, entender y usar un segundo idioma.

En este orden de ideas, la situación planteada por la habitación china, en realidad, no es un fiel reflejo de cómo funciona la mente humana pues el hecho de que la *persona/computadora* no entienda chino no quiere decir que el hablante nativo del chino es poseedor de una semántica de la que carece el hablante nativo del inglés. Además, en estricto sentido, hay algo que la persona inicialmente encerrada en la habitación china sí entiende como entiende el chino nativo: la lengua nativa. Searle propone que lo imaginemos, en tanto hablante nativo del inglés, encerrado en la habitación. Así, Searle tiene que entender, al menos, las palabras en las que están escritas las instrucciones. Estas palabras no tienen que ser *toda* la lengua inglesa; aunque el mismo Searle concibe la posibilidad de que él entienda una lengua nativa, puede que sólo sea un pequeño grupo de palabras; las necesarias para entender las reglas de manipulación:

> En suma, "Searle en la habitación" necesita entender sólo ese subconjunto del inglés de Searle que equivale al lenguaje de programación que entiende una computadora para generar el mismo comportamiento de entrada y salida de preguntas y respuestas por la ventana (Boden 1990, 115).

Lo anterior tampoco es un argumento que asegure que la sintaxis es suficiente para la semántica, pero el hecho de que la *persona/computadora* dentro de la habitación no entienda chino, tampoco es un argumento irrefutable de que la sintaxis no es suficiente para la semántica; tan solo es un argumento que muestra que la persona encerrada en la habitación sí entiende un idioma tal como lo entiende la persona que se encuentra afuera de la habitación; de hecho, ¡es un argumento que muestra cómo no se debe enseñar un segundo idioma!

Analogía de micro-nivel: Si Searle entiende por computador una parte del agente cognitivo de IA

Podría pensarse que Searle solamente ataca la posibilidad de que un agente cognitivo de IA manifieste y posea entendimiento idéntico al de un ser humano, pues en el test de Turing queda claro que una máquina completa -esto es, un agente cognitivo completo-, toma el papel de una persona. No obstante, según la respuesta de Searle a *la réplica del robot*, también puede interpretarse a la persona dentro de la habitación como una parte computacional de un agente cognitivo, como un microchip; en últimas, tanto un chip como la persona dentro de la habitación ejecutan cómputos. Si la persona dentro de la habitación es una parte computacional, entonces la analogía ya no se interpretará desde los rasgos de macro-nivel, sino desde los rasgos de micro-nivel, como se propone en la siguiente gráfica:

Gráfica 4. Términos de la analogía en el micro-nivel

Los términos de comparación

En la réplica del robot se indaga la posibilidad de colocar un computador dentro de un robot que cuente con capacidades motoras y de procesamiento de estímulos externos; de esta manera, se supone que el robot podría establecer una relación más rica con su mundo -incluso causal- y así constituiría verdadero entendimiento. A esta réplica Searle responde:

> Supóngase, aunque yo lo ignoro, que algunos de los símbolos que recibo provienen de una cámara de televisión integrada al robot y que los otros símbolos que remito al exterior sirven para echar a andar los motores dentro del robot que mueven las piernas o los brazos del robot (Searle 1990, p. 92).

En la réplica se propone colocar un computador dentro del robot, y en la respuesta Searle asegura que la persona dentro de la habitación cumple el papel del computador dentro del robot; en esta medida, la persona dentro de la habitación será un artefacto computador que se encuentra dentro de un agente cognitivo de IA como un robot. Esta comparación justifica la interpretación de la analogía presentada en la gráfica 4: la persona dentro de la habitación cumple el papel, no de un agente cognitivo de IA, sino de una parte computacional del agente, de un artefacto computador. Boden (1990 p. 111) reconoce que Searle "discute este punto imaginando un robot que, en lugar de contar con un programa de computadora que lo haga funcionar, tiene un Searle miniaturizado en su interior, en su cráneo tal vez."

De esta manera, el tercer término de la analogía no consistirá en la naturaleza del entendimiento de una persona común, sino en la naturaleza de una de las partes que, al interior de una persona, procesa la información de su entorno; esto, porque se ha renunciado a la interpretación de la analogía desde rasgos de macro-nivel (agente cognitivo de IA) para hacer una interpretación de un rasgo de micro-nivel (una parte que se encuentra dentro del agente cognitivo de IA). A su vez, esto quiere decir que todo aquello que se predique acerca de la persona dentro de la habitación será extensivo a una parte del agente cognitivo de IA y, de esta manera, a algún órgano dentro de una persona común. Es necesario determinar qué parte humana de procesamiento representa la persona dentro de la habitación: ¿Cerebro, neurona o sinapsis? Dado que en la respuesta a la réplica Searle habla de una sola persona y no de

muchas personas que procesan información, entonces el tercer término de la analogía puede interpretarse como un órgano de procesamiento centralizado, esto es, como un cerebro. La persona que ahora se encuentra dentro del robot, "ordena los papeles, recibe y entrega *güiris y guaras* casi de igual manera como antes lo hizo *Searle dentro de la habitación*" (Boden 1990, p. 111). De esta manera, el experimento mental de la habitación china puede interpretarse como una analogía entre a). la persona dentro de la habitación *-persona/computadora-*, b). una parte de un agente cognitivo de IA y c). una parte de una persona común.

La respuesta de Searle a la réplica del robot finaliza con la idea de que, como la persona dentro del robot no entiende ningún símbolo que está procesando, entonces el robot no puede constituir ni poseer verdadero conocimiento. Dado que Searle dentro del robot no está entendiendo la información que está procesando, entonces el robot tampoco puede entender nada; dado que la persona dentro del robot carece de entendimiento, entonces el robot carece de estados intencionales (Searle 1990, p. 92). Como se puede observar, la respuesta de Searle consiste en estas dos premisas:

- Dado que la persona dentro del robot no entiende chino tal como lo entiende un chino nativo que esté por fuera del robot,

- Entonces, el robot carece de estados intencionales.

Esto quiere decir que para aceptar que el robot manifiesta y posee verdadero entendimiento, es necesario que la persona dentro del robot entienda la información que está procesando; entienda chino tal como lo entiende cualquier persona china ¿Es correcto exigir que la persona dentro de la habitación, en tanto parte computadora, entienda chino en el mismo sentido en que lo entiende una persona china nativa? Ya se verificó la exigencia de Searle a la luz de los rasgos de macro-nivel, ahora es necesario verificar esta exigencia a la luz de los rasgos de micro-nivel.

Lo que el cerebro entiende

La interpretación de la analogía, propuesta en la gráfica 4, exige comparar en el mismo nivel a la persona dentro de la habitación y al cerebro. Poner en el tercer término de la analogía a una persona común implicaría una alteración en la concordancia de los niveles de descrip-

ción de la analogía y, por lo tanto, se estaría comparando un chip carente de posibles funciones motrices, con una persona dotada de estas funciones. Una vez establecido el nivel de comparación, es necesario cuestionar si las mismas exigencias que se hacen sobre una persona, pueden hacerse sobre un cerebro; en otras palabras, ¿Un cerebro entiende, en el mismo sentido en que entiende una persona? Para Boden, esta pregunta tiene una respuesta negativa:

> Searle supone que "Searle dentro del robot" ejecuta las funciones que realiza el cerebro humano (de acuerdo con teorías de computación). Sin embargo, la mayoría de los estudiosos de la computación no atribuyen intencionalidad al cerebro […], Searle caracteriza a "Searle dentro del robot" como a alguien que goza plenamente de intencionalidad genuina, tal como él mismo. La psicología computacional no atribuye al cerebro capacidad de ver germen de soya o de comprender inglés; estados intencionales como estos son propiedades de las personas, no de los cerebros (Boden 1990, p. 112).

Lo anterior quiere decir que, por lo general, la intencionalidad se le atribuye a las personas como sistemas completos y no a alguna parte del procesamiento cerebral; no predicamos intencionalidad de las neuronas, los neurotransmisores o del cerebro, sino de las personas completas; tampoco predicamos intencionalidad de un riñón, un hígado o un dedo. Se acepta que el cerebro es un órgano causal de la intencionalidad y que, por lo tanto, se puede predicar de él la capacidad de causar intencionalidad; sin embargo, no se predica de él, ser portador de intencionalidad genuina en el mismo sentido en que una persona es portadora de intencionalidad.

El hipotálamo, el hipocampo, las emociones y la intencionalidad

La ciencia y la tecnología han permitido identificar las áreas del cerebro que regulan tareas específicas. Se ha determinado que una gran cantidad de emociones se procesan en el hipotálamo. Algunas funciones endocrinas importantes, la sed, el hambre, y algunas emociones, se originan en el hipotálamo. De igual manera, enfermedades mentales que implican una carga emocional están relacionadas con desórdenes que por lo general implican al hipotálamo, la tiroides y las glándulas pancreáticas. La estimulación artificial de distintas áreas del hipotálamo genera respuestas de agresión. Por ejemplo, se ha observado que una estimulación en la región lateral del hipotálamo de un gato, produce respuestas motoras

que caracterizan la agresión, como arqueo del lomo y aumento en la presión sanguínea. Sin embargo, algunos investigadores han asegurado que estas respuestas motoras son "vacías", en la medida en que carecen del contenido consciente de la emoción de agresión:

> Dicha expresión emocional es sólo la manifestación externa de lo que mentalmente asociamos con algunos estados emocionales internos, pero cuyo contexto no existe en ese momento [...]. El animal puede producir los sonidos y los gestos asociados con lo que reconocemos como miedo, dolor o agresión, tales como sisear y mostrar los dientes, pero se trata de una condición simulada. Estos sonidos y gestos se emiten en ausencia del estado emocional que normalmente sirve de contexto para generar tales manifestaciones (Llinás 2002, p. 190).

Diversas investigaciones tienden a darle importancia al flujo de información cortical proveniente del hipocampo y a la formación sensorial proveniente del tálamo ventral. Así, se interpreta al hipocampo como el foco de la experiencia emocional consciente, mientras que el hipotálamo se ha interpretado como el foco de la expresión emocional (Felloús, 1990). Podría asegurarse que en el hipotálamo y el hipocampo se encuentra el fundamento causal de algunos estados intencionales, no obstante ¿Esto quiere decir que el hipotálamo, en sí mismo, siente miedo, rabia, sed o hambre o experimenta alguna sensación o emoción? Esta cuestión apunta a algunas confusiones en las categorías del lenguaje que usamos para hablar acerca de nuestra vida mental.

Cuando el cerebro "siente sed"

Aunque conozcamos el funcionamiento y la contrapartida biológica de *la sensación de sed* y, por lo tanto, podamos determinar cuándo un cerebro se encuentra en el estado propio de *la sed*, no podemos asegurar, en estricto sentido, que el cerebro *tiene sed*. Siguiendo a Searle, tener sed es un estado intencional si se observa desde un rasgo de macro-nivel; es decir, cuando observamos a una persona que se comporta como si tuviera sed o cuando escuchamos a alguien decir que tiene sed, podemos asegurar que estamos ante una actitud proposicional: la persona con sed está deseando un objeto líquido especifico, una gaseosa, una limonada o un vaso de agua; además, cuenta con algunas creencias específicas acerca de estos objetos y la manera de acceder a ellos. Sin embargo, en el micro-nivel, el cerebro tan sólo está respondiendo a ciertos estímulos, con la liberación o inhibición de los neurotransmisores correspondien-

tes y está ejecutando un complejo entramado de reacciones químicas y eléctricas. En el macro-nivel observamos a una persona con sed y en el micro-nivel tan sólo observamos el comportamiento del mecanismo biológico que permite la actitud proposicional de tener sed. Asegurar que el cerebro *tiene sed* en el mismo sentido en que cualquier persona tiene sed, implica la posibilidad de predicar intencionalidad acerca de los órganos del cuerpo; implica un isomorfismo radical entre el cerebro y la persona portadora de su cerebro.

En un sentido metafórico, podemos decir que un cerebro tiene sed cuando se encuentra en el estado biológico en que su portador manifiesta la sensación de sed; sin embargo, en sentido estricto, el cerebro no tiene sed porque no está deseando un objeto específico que se encuentre en el mundo en el que vive su portador. Cuando el portador del cerebro tiene hambre, desea una comida específica; el portador del cerebro podrá desear un trozo de carne, un trozo de pan o una barra de chocolate y ninguno de estos objetos existen en el micro-nivel de descripción del funcionamiento del hipotálamo y el hipocampo. En el macro-nivel, el portador del cerebro ingiere carne, mientras que en el micro-nivel el cerebro recibe proteínas, grasas y lípidos. Es posible que a un nivel inconsciente, nosotros también deseamos las grasas, las proteínas y los lípidos. No obstante, la actitud proposicional y el objeto intencional de nuestro deseo, de manera consciente, no son las grasas, los lípidos y las proteínas, sino el trozo de carne. Nuestras creencias y nuestros deseos recaen sobre objetos que se encuentran en el macro-nivel y no en el micro-nivel, si por micro-nivel entendemos el nivel de funcionamiento químico y eléctrico. La disparidad entre el nivel descriptivo del deseo, en el macro-nivel, y el funcionamiento bioquímico del cerebro, en el micro-nivel, se manifiesta en la carga subjetiva del lenguaje y en nuestra manera de satisfacer nuestras necesidades químicas.

Cuando el cerebro "está enojado"

Puede asegurarse que sentir sed o sentir hambre no son rasgos emocionales sino simples respuestas biológicas; no obstante, el hipotálamo también regula conductas emocionales como la agresión y el enojo. Esto quiere decir que observando directamente el cerebro, podemos determinar cuándo se encuentra en un estado en el que su portador siente enojo y deseos de agredir algo o alguien. En esta medida, podemos formular la misma pregunta: ¿Cuando el cerebro se encuentra

sobrecargado de epinefrina, dopamina y serotonina, podemos asegurar que se encuentra enojado y está agrediendo, en el mismo sentido en que aseguramos que una persona está enojada y está agrediendo? Cuando una persona siente una fuerte ira hacia otra, sus posibles deseos consistirán en golpear, gritar y agredir objetos que solamente existen en el macro-nivel, mientras que el cerebro solamente estará segregando neurotransmisores y activando áreas.

Lenguaje, objetos intencionales y la satisfacción de nuestros deseos

Para predicar lo mismo acerca de los cerebros y las personas, sería necesario que los mismos objetos intencionales que encontramos en los rasgos de macro-nivel se encontraran en los rasgos de micro-nivel. Cuando los objetos de deseo no se encuentran en el mismo nivel descriptivo, se debe ser cuidadoso con el uso del lenguaje; por ejemplo, no es lo mismo *desear* la paz mundial, que desear un dulce que tenemos en el bolsillo. En ambos casos hablamos de *un deseo*, sin embargo, el objeto intencional de dicho deseo es distinto y, por lo tanto, el sentido en que deseamos es también distinto. De igual manera, desear tener el sol en la mano es distinto a desear almorzar dentro de una hora; en el primer caso el criterio físico de satisfacción del deseo es prácticamente imposible, mientras que en el segundo caso sí estamos ante un deseo que puede satisfacerse.

Cuando decimos que el cerebro *desea*, lo más probable es que el sentido de éste deseo sea distinto al sentido en que deseamos comer un trozo de carne. Si hubiese un total isomorfismo entre el macro-nivel y el micro-nivel de descripción, es probable que las actitudes proposicionales no contaran únicamente con objetos intencionales como los trozos de carne y las limonadas, sino que habláramos en términos de la necesidad que tenemos de ingerir iones que permitan el correcto funcionamiento de las células. Por lo general, los objetos intencionales del macro-nivel son distintos a los objetos que encontramos en el macro-nivel. Lo más probable es que, cuando estemos hablando de los deseos y las necesidades del cerebro, no nos estemos refiriendo al mismo tipo de deseos y necesidades de las personas comunes, en el macro-nivel.

Actividad perceptivo-motora e intencionalidad

El movimiento parece ser un elemento importante en la presencia de intencionalidad en un sistema físico. Cuando Husserl (1939/1980) muestra la dinámica en que el *yo* orienta su interés hacia algún objeto especifico, parece claro que esta dinámica sólo es posible en la medida en que el sistema acerca del cual se está predicando intencionalidad, esté dotado de capacidades perceptivas y motoras. La intencionalidad implica una orientación consciente del *yo* hacia algún objeto o evento específico y esto implica la posibilidad de manipulación motriz sobre los aparatos de percepción. Si bien la orientación consciente no solamente se refiere a dirigir la mirada o mover la cabeza para que el campo de percepción aumente en función de un objeto determinado, sí es claro que gran parte de las orientaciones conscientes del interés se dan en función de la motricidad de las capacidades perceptivas.

Aquellos objetos que al convertirse en objetos de interés comienzan a ser objetos intencionales, y que están dados en función de capacidades motrices y perceptivas, pertenecen a lo que Husserl (1939/1980) denomina esfera de lo sensible.[16] Puede haber objetos intencionales que se encuentran por fuera de esta esfera, pero tiene que haber, en condiciones normales,[17] algunos objetos intencionales que pertenezcan a ella. En general, gracias al contraste por homogeneidad o heterogeneidad, en el campo de pre-datos, ciertos objetos captarán y lograrán la orientación del yo. Así, el objeto que se impone al yo puede darse con diferentes grados según la intensidad con que el yo se entregue al estímulo. Todo este proceso consiste en la vigilia del yo, en la que hay una orientación hacia un objeto. A partir de esta actividad, se producen actividades más pronunciadas en las que aparece el movimiento manifestado en cinestesias y en movimientos más conscientes que permiten la exploración de aquel objeto que en un primer momento se ha destacado.

16 La esfera de lo sensible se contrapone a la esfera de las ideas (Husserl, 1939/1980).

17 Es posible que en condiciones de permanente imposibilidad motriz, como cuando una persona se encuentra en estado comatoso, también haya objetos intencionales en la vida mental de la persona. Esto es algo similar a lo que sucede cuando, en un sueño, tenemos deseos, creencias o angustias.

Lo anterior no tiene el propósito de argumentar que, como el robot de la réplica cuenta con aparatos perceptivos, entonces se puede concluir que está dotado de intencionalidad. Lo anterior solamente evidencia la importancia de la percepción motriz en la dinámica intencional, no como determinante causal, sino como constituyente de la misma. Husserl (1939/1980) siempre tiene presente la posibilidad perceptiva en aquello acerca de lo cual se predica intencionalidad y, de esta manera, se puede mostrar que esta predicación parece incorrecta para el caso de órganos aislados como el cerebro.

El cerebro entiende, pero no como entiende una persona

Parece incorrecto adscribir al cerebro aislado, la misma intencionalidad que se le adscribe a las personas que portan cerebros. El cerebro procesa estados emocionales, sensaciones y necesidades y, en esta medida, se puede establecer la contrapartida fisiológica de algunos estados intencionales; sin embargo, esto no quiere decir que el cerebro, conscientemente, experimente los mismos estados intencionales que experimenta su portador. Dos han sido las razones para argumentar a favor de la idea de que el cerebro no puede interpretarse como un órgano poseedor de intencionalidad en el mismo sentido en que las personas poseen intencionalidad. En primer lugar, decir que un cerebro *siente sed* o *tiene hambre*, no quiere decir lo mismo que "esa persona tiene sed" o "esa persona tiene hambre". Si el cerebro se encuentra en el estado correspondiente a la sed, es probable que digamos que "ese cerebro tiene sed", pero la disparidad entre el lenguaje descriptivo para el macro y el micro-nivel indicará que el cerebro, en sí mismo, no experimenta la misma sensación de sed que experimenta su portador; el cerebro ejecuta reacciones químicas que, cuando se dan en el contexto del cuerpo humano, se convierten en conductas. En segundo lugar, el cerebro no es un sistema que por sí mismo posee funciones motrices y, por lo tanto, perceptivas; el cerebro es un órgano y no un organismo y, por este motivo, no puede interpretarse como un sistema que posee intencionalidad.

Ahora se debe proceder a examinar la exigencia de entendimiento que hace Searle. Si se acepta que la intencionalidad no es algo que se predica acerca de los cerebros, entonces puede reconocerse que el entendimiento tampoco es algo que se predica o se exige acerca de los

cerebros. Los mismos dos motivos que sirven para refutar la idea de que el cerebro es portador de intencionalidad, sirven para refutar la idea de que, en estricto sentido, los cerebros entienden lo que procesan. Como señalamos líneas atrás, podemos decir que en un momento determinado un cerebro *tiene sed o está enojado* para referirnos al hecho de que podemos identificar el estado cerebral de tener sed o estar enojado. De igual manera, cuando el cerebro se encuentra en el estado en que su portador está leyendo o hablando, podemos decir que ese cerebro está leyendo o está hablando; sin embargo, el sentido en que usamos los verbos cambia si hablamos del cerebro o del portador del cerebro.

Lo mismo que sucede con los verbos hablar, leer o escuchar, sucede con el verbo entender; cuando decimos que un cerebro entiende, podemos usar el término de dos maneras: (a) para asegurar que en el micro-nivel las neuronas cuentan con un lenguaje constituido por índices de disparo y que, en esa medida, hay cierto entendimiento en el cerebro, o (b) para señalar el estado en que el cerebro se encuentra cuando el portador del cerebro está entendiendo algo. Tanto (a) como (b) son usos distintos al sentido en que decimos que una persona "entiende". Así, decir que un cerebro "entiende" es distinto a decir que una persona "entiende"; exigirle a un cerebro que entienda, en el mismo sentido en que una persona entiende, parece ser una exigencia incorrecta, entre otros motivos, porque no hay criterio de satisfacción para la exigencia del cerebro o, al menos, el criterio de satisfacción de ambas exigencias son completamente distintos. Todos podemos verificar con cierto nivel de veracidad, si una persona entiende o no entiende un lenguaje, pero no podemos hacer lo mismo para el cerebro; salvo en el que caso en que queramos comprobar si un cerebro determinado se encuentra en el estado en que se encuentra todo cerebro cuando su portador está entendiendo un lenguaje -por ejemplo, hablando de un lenguaje-, pero en ese caso ya estaríamos usando en un sentido distinto el verbo "entender".

En conclusión, no parece correcto exigir a un cerebro que entienda un lenguaje; la exigencia de entendimiento es algo que se le hace a las personas y no a los cerebros. De igual manera, no sería correcto exigir que una parte de un agente cognitivo de IA sea portador de entendimiento; no sería correcto exigir que la memoria de procesamiento o la memoria de almacenamiento, por sí mismas y por separado, entiendan algo; esa exigencia se le hace al agente cognitivo completo y

no a una de sus partes. Para Boden (1990), exigir que la *persona/computadora* que está encerrada en la habitación entienda chino, resulta de un error de categorías:

> En suma, la descripción de Searle de que el pseudocerebro del robot (es decir, "Searle dentro del robot") comprende inglés, entraña un error de categoría comparable a considerar al cerebro como portador de inteligencia y no como su fundamento causal (Boden 1990, p. 112).

La exigencia de que la persona dentro de la habitación entienda chino

El examen de la analogía desarrollado en este capítulo muestra que es incorrecto exigir a la persona dentro de la habitación que entienda chino; si la persona dentro de la habitación se interpreta desde un rasgo de macro-nivel o desde un rasgo de micro-nivel, parece incorrecto exigirle que entienda chino.

Si la persona dentro de la habitación se interpreta como un agente cognitivo, esto es, desde un rasgo de macro-nivel, entonces surge cierta asimetría en las condiciones que le permiten a una persona entender un nuevo idioma: la persona común, que está por fuera de la habitación, se encuentra en unas condiciones distintas a las de la persona que se encuentra dentro de la habitación; estas condiciones son suficientes para justificar el hecho de que la persona común entiende chino y la otra no, sin necesidad de apelar a semántica o sintaxis. En este caso, la exigencia es asimétrica porque para exigirle a una persona que entienda un idioma completamente nuevo, por lo general, no se le encierra en una habitación. En esta medida, la conclusión de que un robot o un computador no pueden entender nada en virtud de que la persona dentro de la habitación tampoco entiende, al parecer, no es correcta porque resulta de una asimetría entre los términos de la analogía interpretada desde el macro-nivel.

Por otra parte, si la persona dentro de la habitación se interpreta como una parte de un agente cognitivo de IA, esto es, desde un rasgo de micro-nivel, entonces se incurre en un error de categorías. Interpretar a la persona dentro de la habitación como un artefacto computacional que realiza una tarea de procesamiento centralizado

dentro de un agente cognitivo de IA -un robot, en este caso-, implica que el tercer término de la analogía lo ocupa el cerebro humano; en esta medida, todo lo que se diga acerca de la persona dentro de la habitación, se dirá acerca del artefacto de procesamiento centralizado dentro del robot y, también, se dirá acerca del cerebro. Así, el resultado es que la exigencia de entendimiento también recae sobre el cerebro humano; es decir, exigir que la persona dentro de la habitación entienda inglés, equivale a exigirle al cerebro de una persona que entienda inglés. Según Boden (1990), esto conlleva un error de categorías en el que el cerebro se interpreta como portador de intencionalidad e inteligencia, y no como lo que es, a saber, el fundamento causal de la intencionalidad e inteligencia.

Si la exigencia de que la persona dentro de la habitación entienda chino es incorrecta, parece haber una falla en la conclusión que resulta de dicha exigencia: que una máquina no puede poseer mente en virtud de instanciar un programa. Por supuesto, invalidar las exigencias de entendimiento que hace Searle no implica argumentar a favor de que una máquina puede poseer y manifestar verdadero entendimiento, consciencia e intencionalidad. Lo señalado en el presente capítulo tampoco quiere decir que la sintaxis es suficiente para la semántica y que instanciar un programa implica constituir una mente genuina, idéntica a cualquier mente humana. Lo señalado en este capítulo sólo quiere decir que las exigencias parecen ser incorrectas y que, por lo tanto, la conclusión de que una máquina nunca podrá constituir una mente genuina, no se sigue del hecho de que la persona dentro de la habitación no entiende chino; de esta manera, lo que Searle desde un principio considera como un axioma -que la sintaxis no es suficiente para la semántica-, continua siendo un axioma, una verdad indudable, y no una conclusión que pueda resultar de la situación propuesta en el experimento mental de la habitación china.

Sumario

a). El experimento mental de la habitación china puede interpretarse desde rasgos de macro-nivel y rasgos de micro-nivel. En cada interpretación, la analogía cuenta con tres términos: el primer término corresponde a la habitación china, el segundo término corresponde a su contrapartida en la IA y el tercer término corresponde a su contrapartida en la mente humana.

b). Si el experimento mental de la habitación china se interpreta desde los rasgos de macro-nivel, entonces la persona dentro de la habitación debe interpretarse como un agente cognitivo de IA en el segundo término de la analogía, y como una persona común, en el tercer término.

c). En la interpretación de macro-nivel, parece incorrecto exigirle a la persona dentro de la habitación que entienda chino, pues las condiciones no obedecen a las necesarias para que cualquier persona -tercer término- pueda entender un nuevo idioma.

d). Si el experimento mental de la habitación china se interpreta desde los rasgos de micro-nivel, entonces la persona que se encuentra dentro de la habitación debe interpretarse como una parte de procesamiento centralizado de un agente cognitivo de IA en el segundo término de la analogía, y como un cerebro humano en el tercer término.

e). En la interpretación de micro-nivel parece incorrecto exigirle a la persona dentro de la habitación que entienda chino, porque esto equivaldría a exigirle al cerebro que entienda chino. Esto, según Boden (1990), no es más que un error categorial que resulta de interpretar al cerebro como portador de inteligencia y no como su fundamento causal.

f). Ni (c) ni (e) son argumentos a favor de la IAF, ni demuestran que la sintaxis sí es suficiente para la semántica. La conclusión (c) y la conclusión (e) tan sólo demuestran que la conclusión que según Searle resulta del experimento mental de la habitación china, no queda completamente demostrada en virtud de la situación propuesta en el experimento.

Capítulo 5:

Robots, actos de habla e imposibilidad de verificación de intencionalidad[18]

En el presente capítulo se muestra que la adscripción intencional es un factor importante en la teoría de los actos de habla; esto, porque no siempre se puede verificar la calidad intencional del emisor de un acto, razón por la que procedemos a adscribir intencionalidad antes de verificarla. Consideramos indispensable adelantar el examen de la teoría de los actos de habla, porque la capacidad de que un robot genere efectos psicológicos es determinante de la IA. La idea de que buena parte de la generación de los actos de habla depende de la adscripción que asignen los receptores, tiene consecuencias directas en la manera como interpretamos los sistemas físicos que llamamos *robots*; además, tiene consecuencias en la posibilidad de que los interpretemos como sistemas intencionales.

El capítulo consta de cuatro partes. En la primera parte se muestra la relación entre la intencionalidad y los actos de habla. En

18 Agradecemos a Raúl Meléndez por sus comentarios y aportes a la primera versión del presente capítulo, que apareció inicialmente como el borrador de Método número 22. Muchas ideas aquí propuestas no habrían sido posibles sin su colaboración.

la segunda parte se propone un experimento mental para mostrar que no siempre podemos verificar si una emisión determinada proviene de un ser intencional. También mostramos que la ausencia de intencionalidad intrínseca no es motivo para que un receptor no interprete un sonido o una marca como un acto de habla; esto tendrá algunas consecuencias en la teoría de los actos de habla de Searle. En la tercera parte se muestra que sistemas físicos tradicionalmente interpretados como carentes de Intencionalidad, a saber, los robots, también pueden producir actos de habla porque sus emisiones pueden tener efectos psicológicos idénticos a los producidos por los seres humanos. Se concluye que la pregunta por la Intencionalidad de los robots no es importante, porque las adscripciones de Intencionalidad son suficientes para que sus emisiones sean interpretadas como actos de habla.

Intencionalidad y actos de habla

Según Searle (1992), puede pensarse que la cuestión del significado se puede reducir a formas primitivas de Intencionalidad; esto no es trivial porque "definimos el significado del hablante en términos de formas de Intencionalidad que no son intrínsecamente lingüísticas" (Searle 1992, p. 98). Por este motivo, para Searle (1991, 1992) la filosofía del lenguaje es una rama de la filosofía de la mente, pues nociones semánticas como el significado, pueden definirse a partir de nociones básicas psicológicas, como creencias, deseos e intenciones.[19] Para mostrar cómo la Intencionalidad interviene en el proceso de significación, a continuación reconstruiremos, a manera de ejemplo, el análisis del acto de habla de prometer, tal como lo desarrolla Searle (1991).

Intencionalidad, intenciones y promesas

Searle (1991) determina las condiciones necesarias y suficientes para que el acto ilocucionario de la promesa se realice correctamente. Cada condición es una condición necesaria y el conjunto de condiciones es la condición suficiente para dicha realización. Si se tiene que un hablante H emite una oración O frente a un oyente A, "entonces en la emisión de O, H promete sincera (y no defectivamente) que p a A si y sólo si": (Searle 1991, p. 441)

19 Searle reconoce que tras este enfoque se encuentra la idea de Grice (1969/1991), de que la significación está determinada por las intenciones.

1). Hay condiciones normales de *input* y *output*: entiéndase por *output* las condiciones para hablar inteligiblemente y por *input* las condiciones para comprender el lenguaje. Dentro de estas condiciones "normales" de comunicación se tienen las siguientes: saber hablar el lenguaje, consciencia en H y A, que H no esté coaccionado u obligado, que no haya impedimentos físicos como sordera, y que H y A no estén actuando.

2). H expresa que p en la emisión O: según Searle, esta condición aísla el contenido proposicional.

3). Al expresarse que p, H predica un acto futuro X de sí mismo: en la promesa se predica un acto que solamente es posible en el futuro y en primera persona. Esto quiere decir que es incorrecto prometer algo que ya ha sucedido o prometer que alguien más hará algo. Si bien se puede prometer cuidar de que alguien más haga algo, no se puede prometer, en nombre de esa otra persona, que esa persona hará algo.

4). A prefiere que H haga X -a que no lo haga- y H cree que A prefiere que H haga X -a que no lo haga-: esta condición permite distinguir entre una verdadera y legítima promesa, y una amenaza. En la promesa, la acción X se ejecuta para beneficiar, mientras que en la amenaza dicha acción se hace para afectar; de esta manera, en el primer caso se prefiere que suceda X mientras que en el segundo se prefiere que no suceda.

5). No es obvio para H ni A que H hará X: sólo tiene sentido prometer la ocurrencia de X si no es obvio que X ocurrirá, si así es, entonces no es necesario prometer. Lo mismo sucede para prometer la ocurrencia de X.

6). "H tiene la intención de hacer X" (Searle 1991, p. 444): esta condición permite diferenciar entre promesas sinceras y promesas no sinceras. En el primer caso, H tiene la intención de ejecutar el acto que se promete, mientras que en el segundo caso esta intención no está presente. Searle (1991) no considera necesario enunciar una condición adicional que indique que H debe creer poseer la capacidad de ejecutar o no ejecutar X, porque si la intención está presente, es de suponer que H cree tener la posibilidad de cumplir el acto prometido.

7). "H tiene la intención de que la emisión de O le coloque a él bajo la obligación de hacer A" (Searle 1991, p. 444): a diferencia de otros actos ilocucionarios, la emisión de la promesa implica obligación por parte de H para cumplir X. Searle (1991) denomina esta condición como la condición esencial.

8). "H tiene la intención de que la emisión de O produzca en A la creencia de que las condiciones (6) y (7) se dan por medio del reconocimiento de la intención de producir esa creencia, y él tiene la intención de que este reconocimiento se logre por medio del reconocimiento de que la oración se usa convencionalmente para producir tales creencias" (Searle 1991, p. 445): esta condición liga la intención de producir un determinado efecto ilocucionario con el efecto ilocucionario convencionalmente asociado a una determinada emisión. De esta manera, H tiene la intención de producir un efecto en A logrando que A reconozca su intención de producir dicho efecto. Sin embargo, H también tiene la intención de que este reconocimiento se logre mediante el "hecho de que el carácter léxico y sintáctico del ítem que emite se asocia convencionalmente con la producción de ese efecto" (Searle 1991, p. 445). Según Searle (1991), puede pensarse que la iteración de todas estas intenciones no es necesaria, de manera que basta con que H conozca el significado -¿convencional?- de la oración.

9). "Las reglas semánticas del dialecto hablado por H y A son tales que O se emite correcta y sinceramente si y sólo si se dan las condiciones (1) - (8)" (Searle 1991, p. 445): esta condición indica que la oración emitida se usa con el propósito de hacer una promesa únicamente de acuerdo a las reglas semánticas del lenguaje.

Si bien Searle (1991) indica las condiciones de realización de una promesa sincera, reconoce que las promesas no sinceras son también un tipo de promesas, por lo cual es necesario determinar las variaciones que se deben hacer a las condiciones ya expuestas, para hacer una promesa no sincera. La característica más importante de una promesa no sincera es que, si bien en este caso el hablante no posee todas las intenciones que están presentes en la promesa sincera, sí tiene la intención de hacer parecer como si las tuviera. De esta manera, el hablante puede decir que promete hacer X y, sin embargo, no tiene la intención de hacer X. Cuando el hablante dice que promete hacer X, no se implica que tiene la intención de hacer X, sino que solamente está asumiendo la responsabilidad de tener la intención de hacer X,

de manera que basta revisar la condición (6) para mostrar "no que el hablante tiene la intención de hacer X, sino que él asume la responsabilidad de tener la intención de hacer X" (Searle 1991, p. 446). Una reformulación de la regla (6) sería la siguiente: "(6*) H tiene la intención de que la emisión de O le hará a él responsable de tener la intención de hacer X" (Searle 1991, p. 446).

Intencionalidad intrínseca, intencionalidad derivada y los actos de habla

Si se pregunta qué hace que un simple acto, como la emisión de un sonido o la escritura de una marca sobre un papel, sea un acto de habla, la respuesta es que dicho acto ha sido producido por un ser que posee genuina intencionalidad e intenciones especificas de significación. Según Searle, el hecho de que un sonido o una marca sobre un papel sea producido por un ser con intenciones de significación y comunicación, hace que estas emisiones posean un significado y posibiliten comunicación. Así, en la intencionalidad se muestra cómo pasamos de lo físico -sonido o marca sobre papel- a la semántica; en la medida en que los humanos poseemos la biología necesaria para generar intencionalidad, entonces somos una especie de seres afortunados porque podemos convertir objetos físicos -como piedras y papeles- en objetos significantes; poseemos la capacidad de pasar de la sintaxis a la semántica:

> Hago un ruido que recorre mi boca o hago algunas marcas sobre papel. ¿Cuál es la naturaleza de la intención en la acción compleja que hace de la producción de esas marcas o sonidos algo más que solamente la producción de marcas o sonidos? La respuesta breve es que intento que su producción sea la realización de un acto de habla. La respuesta más larga consiste en caracterizar la estructura de intención (Searle 1992, p. 171).

Las respuestas de Searle (1992), tanto la breve como la larga, muestran que en el proceso de emisión están las características que hacen que el acto físico sea un acto de habla significativo. Las características de producción del acto físico permiten determinar si éste puede ser interpretado como un acto de habla. El emisor posee ciertos estados intencionales y dichos estados imprimen capacidad lingüística comunicativa al proceso físico del sonido o la marca sobre el papel. Así, para Searle (1992), la presencia de intencionalidad y de la intención específica de significar son elementos constitutivos del

acto de habla. En la medida en que Searle presenta dicha intencionalidad como regla constitutiva del acto de habla de prometer, puede suponerse que la ausencia de intencionalidad implica, necesariamente, una falla en el acto de prometer; incluso, puede pensarse que la ausencia de intencionalidad impide que un acto de habla sea un acto de habla, pues la ausencia de intencionalidad eliminaría las capacidades significantes de un sonido o una marca sobre un papel y harían que ese sonido y esa marca fueran tan solo un sonido y una marca sobre un papel. Esta idea es bastante intuitiva pues cuando nuestro automóvil hace un ruido no pensamos que está hablando como nos hablan las personas que nos rodean. En sentido estricto, el sonido producido por las hojas de un árbol al moverse por causa del viento o una marca sobre un papel, producida por el paso del tiempo, son actos físicos carentes de un ser intencional en su producción y, por lo tanto, no pueden considerarse actos de habla.

Piénsese en la distinción entre Intencionalidad derivada (I.d,), o intencionalidad *como si*, e intencionalidad legitima e intrínseca (I.i.). Es de suponerse que solamente la I.i. permite la capacidad de significación que posee una emisión física que constituye un acto de habla; solamente la I.i. presente en humanos y algunos animales, puede hacer que un sonido o una marca sobre un papel sean actos de habla. Esto quiere decir que la I.d. no puede generar actos de habla, pues esta intencionalidad no es más que un uso metafórico de la I.i.; la I.d. en realidad no es un tipo de intencionalidad. Así, podemos adscribir intencionalidad a sistemas físicos que no la tienen, con lo que estaríamos hablando de I.d.; sin embargo, estos sistemas no pueden generar actos de habla. En este orden de ideas, cuando decimos que el césped está sediento no queremos decir que el césped está emitiendo un acto de habla; el césped, un robot o un programa de computador son sistemas físicos carentes de intencionalidad intrínseca y, por tal motivo, no pueden emitir actos de habla. Cuando nuestro computador nos formula una pregunta, si nosotros suponemos que el computador nos está preguntando algo, en realidad estamos haciendo una adscripción metafórica pues éste no está, en sentido estricto, preguntando ni teniendo la intención de preguntar nada, simplemente porque las máquinas carecen completamente de I.i.; "esa clase de artefactos […] no tienen ninguna intencionalidad intrínseca u original, sólo una intencionalidad derivada" (Dennett 1991 citando a Fodor).

Restricciones espacio-temporales en la verificación de I.i.

Searle señala que una condición indispensable para que un acto de habla sea interpretado como tal, es que el emisor sea un ser intencional, esto es, un ser con la intención de significar y, por lo tanto, con intencionalidad genuina, original e intrínseca. Gracias a la presunción de dicha intencionalidad, un jeroglífico es interpretado como un acto de habla y no como una secuencia de mamarrachos producidos al azar y carentes de significado alguno. Ahora bien, si se afirma que un sonido o una marca es un acto de habla por la presencia de un ser intencional en la emisión, entonces es de suponerse que siempre que interpretamos un sonido o una marca como un acto de habla, lo hacemos porque estamos seguros de que dicho sonido o dicha marca han sido producidos por un ser intencional.[20] Esto sucede cuando estamos conversando frente a alguien o cuando estamos hablando por teléfono; nos parece evidente que en estos casos estamos hablando con seres intencionales y que, por lo tanto, sus sonidos pueden poseer significado y ser actos de habla. Si podemos verificar la fuente del sonido o la marca y nos percatamos de que han sido producidos por un ser intencional, entonces procedemos a interpretar dichos sonidos o dichas marcas como actos de habla. Aquí surge la pregunta: ¿Cómo podemos estar seguros o verificar que un ser es intencional? La respuesta a esta pregunta puede desembocar en un escepticismo radical con respecto a la posibilidad de verificar la calidad de intencional de un ser. En la medida en que un sistema físico es susceptible de antropomorfización, entonces lo consideramos como un ser intencional. Investigaciones neuropsicológicas recientes han permitido identificar los mecanismos biológicos que permiten reconocer a otros sistemas físicos como un miembro de nuestra especie e identificar sus emociones (Adolphs y Tranel, 2000;

20 Esto no quiere decir que todo sonido o toda marca producida por un ser intencional es un acto de habla, pues una persona puede producir mamarrachos carentes de cualquier significado; sin embargo, esto sí quiere decir que todo acto de habla debe interpretarse como producido por un ser Intencional. La presencia de un emisor intencional podría interpretarse como una condición necesaria pero no suficiente de la constitución de un acto de habla. Ahora bien, como se verá más adelante, esta condición tampoco es necesaria pues incluso en ausencia de emisores intencionales se puede configurar un acto de habla.

Adolphs, Baron-Cohen y Tranel, 2002; Adolphs, Tranel, Damasio y Damasio, 1994; Adolphs, Tranel y Damasio, 2003). El mecanismo neurológico denominado neuronas espejo (Gallese, Fadiga, Fogassi y Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese y Fogassi, 1996; Gallese, Keysers y Rizzolatti, 2004) y el mecanismo psicológico denominado Teoría de las otras Mentes (Baron-Cohen, 2000), se consideran relevantes para dicho reconocimiento (Salcedo-Albarán, Zuleta, León-Beltrán y Rubio, 2007).[21]

Si nos percatamos de que los sonidos o las marcas no han sido producidos por un ser Intencional, sino por el movimiento de unos árboles o el paso de los años, por ejemplo, entonces no los interpretamos como actos de habla sino como simples sonidos y marcas carentes de significado alguno.

Obsérvese que la dinámica en que decidimos qué entender como acto de habla y qué no, supone la constante posibilidad de verificación; al respecto, lo importante es reconocer que en la vida diaria, la posibilidad de verificación no siempre es posible; no siempre podemos determinar la naturaleza del emisor y, en algunos casos, ni siquiera podemos determinar la presencia de un emisor. Entonces: ¿Qué sucede cuando no podemos estar seguros de que un sonido o una marca han sido producidos por un ser intencional? ¿Esta incertidumbre nos obliga a abandonar la interpretación de dichos sonidos o dichas marcas como actos de habla? Al parecer, la respuesta es negativa. Cuando no estamos seguros de que el sonido o la marca son el resultado de un ser Intencional, no siempre desistimos de interpretarlos como actos de habla; en estos casos, nuestro aparato

21 Este escepticismo implicaría que nunca podemos verificar la Intencionalidad de los demás sistemas físicos que nos rodean y que, por lo tanto, siempre la estamos suponiendo. Ahora bien, inferir Intencionalidad a partir de la similitud que el sistema físico tiene con nosotros, también puede interpretarse como verificación. No aseguramos que el único criterio de verificación de Intencionalidad es la antropomorfización del sistema, tan sólo creemos que es uno de los criterios que utilizamos. Siguiendo esta última interpretación, supondrémos que en algunos casos, más no siempre, se verifica la Intencionalidad de un sistema físico. Esto quiere decir que, si interpretamos la inferencia por similitud como un criterio que utilizamos para verificar Intencionalidad, entonces en algunos casos ni siquiera esta verificación es posible.

lingüístico no se detiene sino que procedemos a adscribir Intencionalidad. Son muy comunes los casos en que el emisor del acto de habla se encuentra alejado del receptor, ya sea temporal o espacialmente, de manera que el receptor no puede estar seguro de que el emisor tiene la intención de significar lo que el receptor entiende; en algunos casos, el receptor ni siquiera puede estar seguro de si la causa es verdaderamente un sistema intencional.

Símbolos cavernícolas y la verificación de la I.I.

Supongamos que unos arqueólogos encuentran una secuencia de símbolos dentro de una caverna. Hay indicios pero no certeza de que hace miles de años una tribu habitó esta caverna, por lo que parece interesante descubrir si hay posibles mensajes implícitos en la secuencia de símbolos. Si encontramos algún tipo de coherencia, esto es, que al menos estén dispuestos según algún patrón, será fácil suponer que en todos esos mamarrachos hay algún tipo de acto de habla. En la identificación del patrón lingüístico de los mamarrachos se pueden dar las siguientes posibilidades:

a). Después de mucho trabajar en la traducción de los símbolos, nos percatamos de que carecen de significado y no son más que mamarrachos dispuestos uno detrás del otro. No encontramos indicios del dialecto que compone el mensaje o el sistema críptico es completamente desconocido. No podemos preguntarle a quien dibujó estas líneas si él tenía la intención de significar algo, de manera que concluimos que dicha intención no estaba presente en la producción de los mamarrachos; así, suponemos que éstos son el resultado de la erosión o cualquier otra causa natural. No obstante, esta conclusión resulta de que fue completamente imposible descifrar el patrón críptico de esa extraña escritura pero, en realidad, las marcas sí fueron producidas por seres intencionales: la tribu que habitó esa caverna hace miles de años. Esto quiere decir que suponemos erróneamente que no hay mensaje en las marcas gracias a la imposibilidad de traducción y de verificación de intencionalidad. Como hay indicios, mas no certezas, de la existencia de la tribu, no hay razones para suponer que las marcas son el resultado de la acción de algún ser intencional; pero esta falta de certeza sólo se debe a que no se han encontrado las tumbas que están a unos metros de las cavernas y en las que están las pruebas definitivas de la existencia pasada de la tribus. Cuando se encuentren las tumbas, no habrá duda

acerca de la presencia de la tribu, de que las marcas fueron producidas por miembros de dicha tribu y que, por lo tanto, tienen significado.

Lo anterior quiere decir que aun si H tiene la intención I de significar algo con la emisión E, si A no puede percatarse o reconocer I, entonces A concluirá que E no es un acto de habla (c.a.). Recuérdese que es imposible que A se percate o reconozca I, porque no puede verificar -en este caso por restricciones temporales- la calidad de la producción de E.

Dado (c.a.) podemos concluir que esas marcas fueron "escritas" por la naturaleza. Con este enunciado adscribimos I.d. a la naturaleza, diciendo en sentido metafórico que la naturaleza escribió, cuando en realidad la naturaleza no tuvo la intención de escribir algo en alguna parte. Puede que las marcas sean el resultado de I.i. y nosotros, por la imposibilidad de percatarnos y reconocer dicha I.i., concluimos que las marcas son el resultado de I.d y que, por lo tanto, no hay acto de habla alguno. (c.a.1). Ahora bien, si I.i. es el tipo de intencionalidad necesaria para constituir un acto de habla, entonces ¿Por qué aunque la I.i. está presente en la producción de las marcas, nosotros no pudimos identificar un acto de habla en ellas?

b). Después de mucho trabajar en la traducción de los símbolos, finalmente nos percatamos de que hay un mensaje: "El mundo se acabará en diciembre del próximo año". Este mensaje se difunde en los medios de comunicación y por el pánico, la sociedad colapsa. Años después del colapso social, nos encontramos con un descendiente directo de la tribu que escribió la profecía y él nos comenta que ese es un típico ejercicio de escritura realizado por los niños de la tribu cuando cumplían 2 años de edad. Los niños realizaban este ejercicio cuando aprendían sus primeras palabras: los meses del año, los números y las palabras *mundo* y *acabar*. Aunque en la producción de las marcas intervinieron seres intencionales, ellos no tenían la intención de significar lo que nosotros supusimos que querían significar, a saber, la profecía del fin del mundo. Esto quiere decir que incluso si H no tiene la intención i de significar algo con la emisión E, si A adscribe i, aun cuando i no hubiera estado presente en la producción original de E, entonces A concluirá que E es un acto de habla que *significa* en función de i. (c.b.). También se puede asegurar que en este caso hay una separación radical entre el signifi-

cado adscrito y el significado que en función de la intención se quiso expresar; en otras palabras, aún si H no tiene la intención i de significación, si A adscribe i a H entonces A concluirá que el significado S está dado en función de la i adscrita. (c.b.).

A partir de (c.b.), se puede concluir que el hecho de que las marcas sean el resultado de un ser con Intencionalidad, no se sigue que la intención de significación sea la que nosotros le atribuimos a ese ser. En este caso, aunque los niños son sistemas "I"ntencionales, ellos no tenían la intención de significar una profecía, pues sólo pretendían hacer un ejercicio de escritura. Con ésto se evidencia que cuando estamos alejados de la causa de E, importa más la adscripción de la intención que el receptor haga, que la intención que interviene en la emisión de E, aunque ésta sea "I"ntencionalidad.

c). Después de esforzarnos en la traducción, encontramos el mismo mensaje planteado en (b), a saber, la profecía del fin del mundo, con las mismas consecuencias nefastas para la sociedad. Supongamos que esta interpretación de E se da en t. La única diferencia con lo sucedido en (b) es que el descendiente de la tribu milenaria nos explica que esas marcas son el resultado de una reacción natural y que, aunque la tribu existió y habitó esa caverna, ninguno de ellos escribió marca alguna. Esta aclaración por parte del descendiente se da en t+1. Él también explica que en realidad fue "la madre naturaleza" la que escribió esas marcas. Así, solamente en t+1 sabemos que la única posible intencionalidad que intervino en la producción de las marcas fue I.d.; es decir, en la producción de las marcas no intervino I.i. y nosotros, no obstante, encontramos un acto de habla, a saber, una amenaza. Esto quiere decir que aún si H no tiene la intención I de significar algo con la emisión E, si A adscribe I a H, aun cuando I no hubiera estado presente en la producción original de E, entonces A concluirá que E es un acto de habla (c.c.).

A partir de (c.c.), en t supusimos que las marcas eran un acto de habla porque habían sido producidas por seres con Intencionalidad y con la intención de significar esa profecía. Esto quiere decir que en t nosotros adscribimos I.i. a un ser que no intervino en la producción de las marcas. Aunque sólo se pueda hablar de I.d. en la creación de estas marcas, por nuestras adscripciones, nosotros concluimos que dichas marcas son el resultado de I.i. y que, por lo tanto, sí son actos de habla. (c.c.1). Ahora bien, si I.d. no es un tipo de intencionalidad y por lo

tanto no puede constituir y generar un acto de habla, entonces ¿Por qué aunque no había I.i. en la producción de E, nosotros identificamos un acto de habla en las marcas de las cavernas? Esto sucedió porque adscribimos I.i. a un sistema físico que, a lo sumo, desde la distinción de Searle, podía ser poseedor de I.d. o intencionalidad *como si*.

Ausencia de un emisor intencional y adscripción de I.i.

Searle (1991) asegura que la Intencionalidad hace que una señal sobre un papel, o que un sonido, adquieran capacidades comunicativas; si no fuera por esta intención, la señal y el sonido serían simples manifestaciones de actos físicos sin efectos psicológicos en sus receptores. Ahora bien, según lo señalado en (c.c.) y en (c.c.1), una señal o un sonido que son el resultado del azar o de una fuerza natural, sí pueden interpretarse como actos de habla. Dada la incertidumbre acerca de la causa, el receptor puede suponer la existencia de un ser intencional que haya intervenido en la emisión.[22] Un ejemplo de lo planteado en (c.c.) y en (c.c.1) es el caso de las señales que han aparecido en algunos campos de trigo. Muchas de estas marcas han sido la obra de bromistas, pero otras son interpretadas como el resultado de la acción de algún ser sobrenatural o misterioso. Lo siguiente fue publicado en internet en

22 Es posible que aunque se esté seguro de la inexistencia de un ser intencional que haya intervenido en la producción de la emisión, se siga interpretando el sonido o la marca como actos de habla. Por esta razón, las personas creyentes hallan un ser intencional en la lluvia, en un árbol o en cualquier fenómeno de la naturaleza; tal vez por este motivo las personas devotas encuentran intención en cada fenómeno natural. Si un devoto le pide a Dios que llueva y llueve, lo más probable es que se crea que dicha lluvia supone la presencia de un ser intencional que se comunica y responde a las plegarias. Si estamos en una sesión de espiritismo y cuando se formula una pregunta entra una corriente de viento que produce un sonido similar a "nooouuuuuuu", lo más probable es que los participantes de la sesión adscriban intencionalidad a la producción de dicho sonido. Si bien esto puede interpretarse como simple uso metafórico del lenguaje, la frontera entre el uso metafórico y la verdadera suposición de intencionalidad es muy difusa. En algunos casos, las personas usan metafóricamente un verbo intencional pero son conscientes de dicha metáfora y no suponen verdadera intencionalidad en el sistema físico; en otros casos puede, aparentemente, observarse un simple uso metafórico pero las personas suponen y adscriben verdadera intencionalidad. No parece correcto afirmar que cuando las personas le rezan a un ícono religioso, tan sólo están haciendo un uso metafórico y no están suponiendo verdadera intencionalidad.

el año 2004, con respecto a una marca que apareció el 15 de agosto de 2003 en la Crabwood Farm House, Winchester, New Hampshire:

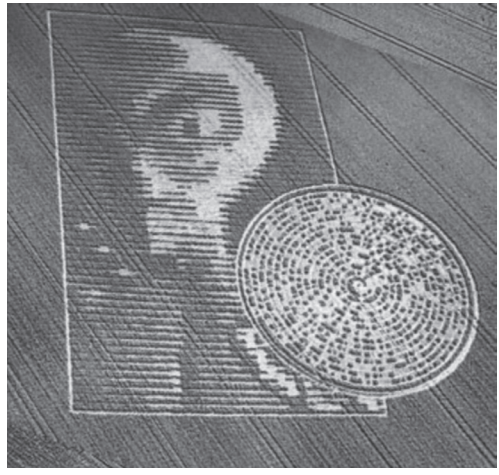


Imagen de los círculos de Crabwood Farm House.
Tomada de http://www.mmmgroup.altervista.org/e-crab.html.

> El objetivo primario de la formación de Crabwood es mostrarnos cómo establecer un diálogo de dos vías con una raza de extraterrestres […]. El posicionamiento de la formación, al lado de una estación de microondas, es una muestra clara del hecho, pues la parte superior de la formación apunta en dirección a ese sitio. Así, ésta formación claramente nos muestra que las microondas son el medio de comunicación [...].[23]

Aunque no es posible corroborar que el emisor de este mensaje tiene la intención de significar lo que se asegura que significa, algunas personas experimentan un efecto psicológico idéntico al que experimentarían si estuvieran seguras de que el mensaje dice lo que ellas suponen que quiere decir. Lo curioso es que esta seguridad con respecto al significado de la marca y al ser que intervino en la producción de la marca, se dan aunque no haya certeza ni acerca del significado de la

23 Ver (http://www.yowusa.com/Archive/August2002/crabwood1/crabwood1.html). Vale la pena señalar que lo escrito en la página de internet no tenía un sentido de broma, sino que era una interpretación seria, acerca de cómo establecer comunicación con extraterrestres.

marca, ni acerca del ser que intervino en la producción; estas certezas son el resultado de suposiciones y adscripciones. Esto no quiere decir que dichas señales no hayan sido producidas por un ser intencional, sólo quiere decir que no es claro que la intención de dicho ser haya sido significar lo que se asegura que la marca significa. Estas señales pueden haber sido producidas por un ser intencional, persona o extraterrestre y, sin embargo, no tener la intención de significar el significado que se le atribuye. Si las marcas han sido producidas por un ser intencional con intenciones distintas a las adscritas, estaríamos frente a una situación similar a la expuesta en (b); si las marcas son el resultado de un fenómeno natural, entonces estaríamos frente a una situación similar a la expuesta en (c). La marca pudo haber sido causada por un fenómeno natural y, en este caso, la ausencia de intencionalidad es radical, entonces, ¿Por qué las personas experimentan el efecto psicológico que se manifiesta en el anterior fragmento de interpretación extraído de una página de internet, el cual, parece similar al efecto psicológico que se experimentaría si se tuviera certeza de que las señales fueron producidas por un ser intencional?

Cuando es imposible "estar seguros"

Con respecto a los jeroglíficos, que nos introducen en una situación similar a la de las marcas de los campos de trigo, Searle (1991) señala que:

> Una presuposición lógica de los intentos corrientes de descifrar los jeroglíficos mayas consiste en que al menos avanzamos en la hipótesis de que las marcas que vemos sobre las piedras fueron producidas por seres más o menos parecidos a nosotros mismos y producidos con ciertos géneros de intenciones. Si estuviéramos seguros de que las marcas eran una consecuencia de, digamos, erosión producida por el agua, entonces la cuestión de descifrarlas o incluso de denominarlas jeroglíficos no podría plantearse. Interpretarlas bajo la categoría de comunicación lingüística incluye necesariamente interpretar su producción como actos de habla (Searle 1991, p. 432).

El problema con esta explicación radica en "estar seguros". Si estuviéramos seguros de que tales símbolos son el resultado de la erosión, entonces no los interpretaríamos como queriendo significar algo, excepto si somos creyentes y religiosos. De igual manera, si estuviéramos

seguros de que los símbolos en los campos de trigo fueron producidos por un fenómeno natural, entonces no los interpretaríamos como queriendo decir algo, o por lo menos no se aseguraría que quieren decir lo que en la actualidad algunos aseguran que quieren decir. El problema radica en que no siempre podemos "estar seguros", por limitaciones espaciales o temporales. Cuando no sabemos qué o quién produjo la emisión, es básicamente imposible asegurar la presencia de intenciones y, por lo tanto, es imposible asegurar que una determinada emisión de sonido o una determinada marca son un acto de habla. Sin embargo, a pesar de dicha imposibilidad para corroborar, nuestro aparato lingüístico continúa funcionando y no nos detenemos a esclarecer las intenciones del emisor de cada emisión que interpretamos como acto de habla. En realidad, la intencionalidad que tiene el emisor importa menos en el lenguaje, que la Intencionalidad y las intenciones que el receptor de la emisión adscribe a la emisión y al emisor; incluso, importa menos que la suposición de presencia de emisor.[24]

Dada la imposibilidad de verificar la existencia y calidad del emisor, cuando hay buenas razones, decidimos suponer y adscribir intencionalidad, antes de verificarla hasta el cansancio. Esta adscripción de intencionalidad parece ser constante; aún si esta adscripción se hace en términos metafóricos, es suficiente para constituirse en una importante dinámica psicológica:

> Heider y Simmel (1944) fueron los primeros en caracterizar la naturaleza básica perceptual de la atribución intencional. A los sujetos de sus experimentos, se les mostraban películas de simples objetos geométricos moviéndose contra un trasfondo estático, y se les pedía describir el contenido de las películas. En elocuciones espontáneas, prácticamente todos los sujetos usaban palabras como "quiere", "teme" o "necesita", para describir los movimientos de las formas geométricas. La antropomorfización de estas formas era completamente automática; los sujetos encontraban extremadamente difícil describir la escena en términos puramente geométricos incluso cuando se les pedía hacerlo […]. Los humanos tienden a atribuir estados intencionales, incluso a las más mínimas percepciones […] (Scassellati 2001, p. 37).

24 Esta suposición de emisor, hecha por el receptor, incluso puede redundar en invención de emisor. Esto deja de manifiesto que la raíz de significación del acto de habla se encuentra en las adscripciones y suposiciones que hace el receptor y no en la verdadera Intencionalidad presente en la emisión.

El hecho de que los humanos atribuyamos intencionalidad casi automáticamente, no implica que los objetos de atribución en realidad poseen intencionalidad. El experimento de Heider y Simmel (1944) evidencia que la adscripción de intencionalidad es una dinámica psicológica relevante en nuestra vida mental, ya que aparece desde los primeros años de edad. Aunque nadie infiere presencia de verdadera Intencionalidad en las figuras geométricas de su experimento, sí se evidencia una adscripción casi automática.

Los efectos psicológicos de los actos de habla

La más clara manifestación de un acto de habla es el efecto psicológico: el pánico ante una amenaza o la expectativa ante una promesa son algunos ejemplos de efectos psicológicos causados por actos de habla. Así, puede afirmarse que aquellos sonidos y aquellas marcas que no interpretamos como actos de habla, no generan el efecto psicológico propio de los actos de habla:[25] ¿Qué es un acto de habla que no ejerce algún efecto psicológico sobre el receptor? ¿Qué es una promesa que no induce a la expectativa del cumplimiento, una amenaza que no induce al pánico o una pregunta que no induce al interés por responder? Ciertamente diríamos que estas no son promesas, amenazas o preguntas pues un acto de habla que no genera un efecto psicológico sobre el receptor, en realidad, no es un acto de habla; al menos, puede asegurarse que una amenaza que no genera pánico no puede interpretarse como un acto de habla genuino de amenaza. Puede asegurarse que un acto de habla que no ejerce efecto psicológico sobre un receptor, en realidad, no es un acto de habla.

Un acto de habla es un acto de habla en tanto posea capacidad de inducir a un efecto psicológico; a su vez, ésta capacidad de inducir está mediada por intencionalidad; sin embargo, el *quid* del asunto radica en si esta intencionalidad debe ser ubicada en la adscripción que de ella haga el receptor o en la que está presente en el emisor

25 Los efectos psicológicos causados por sonidos y marcas que no son interpretados como actos de habla son distintos a los causados por promesas, preguntas, amenazas, saludos, despedidas, aseveraciones, imperativos y represiones, entre otros. Si el ruido de un vidrio rompiéndose puede tener algún efecto psicológico en nosotros, dicho efecto será distinto al causado por un genuino acto de habla.

al momento de la emisión. Se puede afirmar que la mayor parte de efectos psicológicos causados por los actos de habla tienen su origen en nuestras decisiones de adscripción y suposición, antes que tener su origen en los sonidos que salen por la boca o las marcas realizadas por las manos de otra persona. Incluso los efectos psicológicos resultantes de actos de habla producidos por alucinaciones se originan en nuestras decisiones, de adscripción. Una persona X que sufre de esquizofrenia responderá las preguntas hechas por su alucinación únicamente si decide adscribir intencionalidad.

Robots y actos de habla: eliminación de la distinción entre I.i. e I.d.

Searle, asegura que una máquina carece completamente de algún tipo de intencionalidad; las máquinas manipulan símbolos, pero nunca pueden pasar de la sintaxis a la semántica. Piénsese en el caso de un robot RH que emite una promesa frente al auditorio A, y momentos más tarde cumple su promesa.[26] Supongamos que RH "emite un sonido que es interpretado por el auditorio A como una promesa" y que, momentos más tarde, RH cumple lo que se señalaba en dicha emisión ¿Esto quiere decir que pueden producirse promesas sinceras sin tener la intención de cumplirlas? ¿Pueden emitirse promesas incluso en ausencia de cualquier tipo de Intencionalidad? Si se responde afirmativamente a la última pregunta, entonces las reglas para formular promesas sinceras y legítimas serían bastante similares a las reglas para producir promesas no sinceras e ilegítimas, porque producir una promesa sincera no requeriría tener la intención de cumplir lo que se promete y, en general, no requeriría algún tipo de Intencionalidad. Por otra parte, si se responde negativamente a la pregunta, de manera que toda promesa sincera requiera tener la intención de cumplimiento, entonces sería necesario aceptar que RH posee intencionalidad o que RH no produce actos de habla, sino emisiones carentes de algún efecto en auditorios. Como Searle no acepta la presencia de Intencionalidad en las máquinas, entonces la respuesta sería que la emisión de RH no es un acto de habla. Esto nos dejaría con una sola posibilidad

26 Piénsese, además, que RH no está programado para siempre cumplir sus promesas, de manera que incumple en algunas ocasiones. Esto quiere decir que RH también puede emitir promesas no sinceras.

para explicar la conducta de A: la I.d. es suficiente para crear efectos psicológicos idénticos a los efectos psicológicos causados por emisiones con presencia de I.i. En este caso, no habría distinción entre I.i. e I.d. con respecto a los efectos psicológicos que los actos de habla generan en las personas.

Si el auditorio A experimenta un efecto psicológico por el sonido emitido por RH, similar al efecto psicológico producido por los sonidos emitidos por personas, entonces es plausible que el auditorio A esté dispuesto a adscribirle a RH algún tipo de Intencionalidad, en lugar de asegurar que la emisión de RH no es una promesa. Esto, porque si se asegura que RH carece de Intencionalidad y que no produce actos de habla, entonces sería necesario que frente a cualquier sonido emitido por RH, A no actuara como si estuviera interpretando un acto de habla. Para que esto fuera así, si RH se para frente a una persona B, le apunta con un arma y le dice "en cinco segundos te dispararé", la persona tendría que estar dispuesta a no experimentar algún tipo de ansiedad, pánico o temor por su vida. Para B, la emisión de RH debería ser idéntica a un sonido causado por hojas moviéndose por el viento o al sonido causado por la lluvia al caer sobre el pavimento. Sin embargo, es más probable que B decida huir porque interpreta la emisión de RH como un acto de habla -como una amenaza- y, por lo tanto, le adscribe intencionalidad a RH; sobre todo, si RH le apunta con un arma real y, sobre todo, si B ha observado que en el pasado, cuando RH ha emitido estas frases, las personas que no han huido han muerto.

Lo anterior quiere decir que si RH emite lo que conocemos como promesa o amenaza, entonces importa más la adscripción de la intencionalidad que la intencionalidad presente en el emisor. No importa si RH carece completamente de intencionalidad, A estará dispuesta a adscribirle la intención de cumplir la promesa, sobre todo, si ha visto que en el pasado RH ha cumplido sus promesas y sus amenazas. Para corroborar esto, bastaría observar el comportamiento de A frente a la emisión de RH. Esto quiere decir que la pregunta verdaderamente importante de un acto de habla no es si el emisor es un sistema intencional, sino si el receptor está dispuesto a adscribir y suponer intencionalidad en la producción de la emisión. Así, con respecto a los robots, la pregunta importante no es si poseen o no poseen I.i., pues incluso si sólo poseen I.d., ésta es suficiente para que ellos causen efectos psicológicos y variaciones en las conductas de

los humanos; incluso, efectos psicológicos y variaciones conductuales idénticas a las que causan sobre nosotros otras personas de carne y hueso. El día que un robot se siente a la mesa a comer con nuestra familia y actúe como una persona más, nos importará muy poco si tiene o no tiene Intencionalidad. Lo más probable es que en esta situación hipotética responderemos a sus preguntas, a sus peticiones e interactuaremos de manera normal. La pregunta importante no es sobre la "I"ntencionalidad presente en los robots:

> En lugar de preguntar cómo cualidades mentales pueden reducirse a cualidades físicas, nosotros preguntamos cómo adscribir cualidades mentales a sistemas físicos (McCarthy 1979, p. 4).

Sumario

a). Según la teoría de los actos de habla (Searle, 1991), la presencia de I.i. es indispensable para que un acto físico constituya un acto de habla.

b). Para Searle (1990, 1995, 1994, 1997a, 1997b) los robots son sistemas físicos carentes de I.i. En este orden de ideas, un robot no puede emitir un verdadero acto de habla.

c). Los actos de habla deben generar efectos psicológicos en el receptor del acto. Un acto físico (como un sonido o una marca sobre el papel) que no genera efectos psicológicos no es un acto de habla.

d). Un acto físico (como un sonido o una marca sobre el papel) puede generar un efecto psicológico incluso si no fue producido por un sistema intencional, como un robot. Esto sucede porque el receptor del acto está dispuesto a adscribir I.i. al acto físico del robot antes de verificar cuidadosamente si dicho robot posee verdadera y genuina intencionalidad.

e). Teniendo en cuenta (d), la constitución de un acto de habla y su correspondiente efecto psicológico en el receptor, depende más de la I.i. que el receptor esté dispuesto a adscribir, que de la verdadera I.i. que posea el emisor. Esto sucede porque el receptor no siempre puede verificar la presencia o calidad de un emisor intencional y esto no es impedimento para nuestro sistema lingüístico. Cuando no podemos verificar la presencia y calidad del emisor, preferimos adscribir intencionalidad antes que detener nuestro sistema comunicativo.

f). Lo verdaderamente importante con respecto a los robots no es si son sistemas físicos poseedores de I.i., sino bajo cuáles condiciones, los humanos estamos dispuestos a adscribirles intencionalidad. Dicha adscripción posibilitará la convivencia y comunicación armónica entre los sistemas físicos orgánicos -humanos- y los robots.

Capítulo 6:

Información infecciosa y funcionalismo[27]

Como se señaló en el primer capítulo, según el funcionalismo, la mente se puede instanciar en un soporte físico distinto al cerebro humano, siempre que dicho soporte sea lo suficientemente complejo en términos causales para soportar todas las transformaciones físicas establecidas en el programa/mente. Ahora bien, la idea de que la información puede diferenciarse e incluso separarse del soporte físico, ha sido el centro de un gran debate alimentado, en parte, por Searle (1994, 1997) quien asegura que el funcionalismo es un marco conceptual errado porque la información y las computaciones no son rasgos intrínsecos de la naturaleza. En el presente capítulo mostraremos que en la naturaleza sí hay al menos un sistema físico en el que es evidente la distinción y separación del soporte físico y la información.

El presente capítulo tiene tres partes. En la primera, mostramos que una característica definitoria de la molécula de ADN y de los genes es la información. En la segunda parte exponemos el proceso de replicación de los virus. En dicho proceso, conocido como *infección*, se hace

27 Una primera versión de este capítulo, que apareció como Borrador de Método número 29 en el año 2004, se realizó con la colaboración de David A. Sastre.

evidente la distinción y separación entre el soporte físico y el soporte lógico o informacional. Todo lo anterior permite afirmar que, para negar la existencia de sintaxis, información y computación en la naturaleza, es necesario negar algunos hechos registrados y reconocidos por la biología molecular. A continuación exponemos algunos de éstos hechos.

Searle (1994, 1997) asegura que cualquier sistema físico puede interpretarse como ejecutando computaciones e información pues, como la sintaxis no es intrínseca a la naturaleza, entonces es relativa al observador y surge de la atribución arbitraria que dicho observador haga; de hecho, cualquier observador puede interpretar cualquier patrón como computación. La sintaxis, la información y las computaciones no son intrínsecas a la naturaleza porque la sintaxis es relativa a los observadores y todas las computaciones están definidas en términos de sintaxis (Moural, 2003):

> "El verdadero problema es que la sintaxis es esencialmente una noción relativa al observador. […] La caracterización de un proceso como computaciones es una caracterización de un sistema físico desde afuera; y la identificación del proceso como computacional no identifica rasgos intrínsecos del sistema. […] Por la propia definición de computación y cognición, no hay manera de que la ciencia cognitiva computacional sea una ciencia natural, porque la computación no es una característica intrínseca de la naturaleza. Ésta es asignada de manera relativa a los observadores (Searle 1994, p. 212).

Genes e información

Aunque el concepto "gen" ha sido el centro de un profundo debate (Lederman, 1987; Kitcher, 1982), en la biología contemporánea se acepta el aspecto informacional de los genes. De hecho, la mayoría de nosotros está familiarizado con la noción de *información genética* y con el hecho de que dicha información está almacenada en la molécula de ADN (Waters, 1994 p. 174). Estas nociones informacionales de los genes y del ADN aparecen en cualquier texto actual de Biología: "El ADN es una molécula puramente informacional. La información en el ADN está codificada en la secuencia de bases" (Purves, Sadava, Orians y Heller, 2001). No obstante, debe reconocerse el debate acerca de qué tipo de información puede considerarse como un *gen*. Se llama la atención con respecto a la idea de que los genes tienen

poco significado informacional por sí mismos, de manera que dicho significado sólo surge "en contexto con otros genes y en un ambiente que es celular, extracelular y extra-organísmico" (Schaffner, 1998 p. 233), Oyama (1985) y Lewontin (1993) también han llamado la atención sobre la importancia del contexto al momento de interpretar una porción genómica como información. A pesar de éste debate, es aceptado casi de manera unánime en la biología contemporánea que los genes son instrucciones acerca de cómo las células se convierten en un ser humano (Dawkins 1976). Dichas instrucciones están almacenadas en la molécula de ADN:

> Hay un consenso en la biología, en que una versión simplificada del "dogma central de la síntesis de proteínas" es correcta. Ésta forma del "dogma central" sostiene que el flujo de información va del ADN (o ARN) a la proteína, de manera que el ADN (o el ARN) tiene una prioridad informacional especial (Schaffner, 1998 p. 234).

El aspecto informacional de los genes es relevante para el dogma central de la biología porque la expresión de dichos genes explica la conexión entre el fenotipo y el genotipo (Waters, 1994 p. 175). Es tan reconocida la capacidad informacional de la molécula de ADN que actualmente se consideran sus capacidades computacionales (Bhalla et al. 2003) en distintos campos de investigación en los que se interpreta a los patrones formados por los pares de bases como expresiones algorítmicas (Dirks et al. 2004). De igual manera, se ha llamado la atención acerca de las capacidades electrónicas y de almacenamiento de la molécula de ADN; así, se ha considerado que la computación mediante ADN permitirá superar las limitaciones físicas impuestas por la computación cuántica (Bhalla et al. 2003 p. 442). En este mismo orden de ideas, Adleman (1998) ha señalado la similaridad entre la polimerasa y una máquina de Turing. La definición informacional de los genes también ha fundamentado experimentos que muestran cómo los procesos biomoleculares pueden programarse para ser portadores de algoritmos lógicos (Winfree y Bekvolato, 2003). En general, las nociones algorítmicas e informacionales son relevantes para la biología actual:

> La información y los algoritmos aparecen como centrales a la organización y a los procesos biológicos, desde el almacenamiento y reproducción de información genética, pasando por el control de procesos de desarrollo, hasta las

> sofisticadas computaciones realizadas por el sistema nervioso. Así como la tecnología humana usa microprocesadores electrónicos para controlar los dispositivos electro-mecánicos, los organismos biológicos usan circuitos bioquímicos para controlar eventos moleculares y químicos. (Winfree, 2003a).

Ahora bien, si se acepta que los genes son porciones de información, entonces también debe aceptarse que es irrelevante el soporte físico en el que se almacenan y se ejecutan:

> El gen es un paquete de información, no un objeto. El patrón de los pares de bases en la molécula de ADN especifica el gen. Pero la molécula de ADN es el medio; no es el mensaje. Es absolutamente indispensable mantener la distinción entre el soporte físico -medio- y el mensaje para lograr la claridad de pensamiento acerca de la evolución [...]. En biología, cuando se está hablando de cosas como los genes y de genotipos [...], se está hablando de información, no de la realidad física objetiva. Los genes son patrones. (Williams 1996:42).

El virus: información flotante

Un virus es estructural y morfológicamente simple (Leigh-Brown y Holmes, 1994), compuesto por una cápside de proteínas que protege el núcleo, en el que se encuentra la información viral genética (Hutchinson, 2001). La información genética ubicada en el núcleo del virus puede estar compuesta por una cadena sencilla o doble de ADN o ARN (Nahmias y Reanney, 1977). En dicha información genética están consignadas las instrucciones para realizar el proceso de auto-replicación que permite la existencia de las especies virales (Hutchinson, 2001). Por ejemplo, el material genético que se encuentra dentro del virus del VIH, está compuesto por dos cadenas de ARN con cerca de 9200 bases nucleicas (Greene, 1993; Stine, 2000; Fauci, 1988). El material genético viral también permite la evolución del virus, al generar copias mutantes que mejoran su capacidad de supervivencia (Hutchinson, 2001). Esto quiere decir que la modificación de la información genética puede interpretarse como un proceso evolutivo (Bull y Wichman, 2001).

Por lo menos tres genes intervienen de manera directa en la síntesis de las proteínas necesarias para completar el proceso de auto-replicación del virus: *gag, pol* y *env* (Dimmock y Primrose,

1987; Stine, 2000; Greene, 1993; Fauci, 1988; Leigh-Brown y Holmes, 1994). Las proteínas que se encuentran al rededor de la cápside le permiten al virus adherirse a la célula con el propósito de liberar la información genética viral al interior de la célula. La cantidad de información que porta un virión simple es de aproximadamente 240 bits, que es "cerca de 10 millones de veces menos que la información contenida en el genoma humano (que es de tres trillones de bits). Estos 240 bits están dispuestos en un cromosoma circular (el equivalente al medio de almacenamiento de un computador) y contiene un conjunto de información que le permite a la molécula duplicarse a sí misma" (Levine, 1991 p. 2).

Pero el hecho de que el virus porte las instrucciones necesarias para auto-replicarse, no quiere decir que pueda hacerlo por sí mismo. El tamaño de un virión es de aproximadamente 100 nm. Esto implica que, por lo general, un virus es más pequeño incluso que una mitocondría, que es un organelo celular. Esto, a su vez, quiere decir que un virus es acelular, carece de organelos y, por lo tanto, no tiene sistema respiratorio, circulatorio o reproductivo. Dos implicaciones resultan de la simplicidad morfológica del virus. En primer lugar, un virus solamente porta información auto-replicativa y, por lo tanto, no es más que una pieza de información auto-replicativa protegida por una cápside (Nahmias y Reanney, 1977). En segundo lugar, teniendo en cuenta que ninguna información puede ejecutarse por sí misma, en virtud de sí misma, y teniendo en cuenta que un virión no cuenta con el *hardware* necesario para ejecutar dicha información, entonces los algoritmos auto-replicativos portados en el virus necesitan un soporte físico específico para ejecutarse y auto-replicarse. Esto último, porque la ejecución y eterna auto-replicación es una característica definitoria de los genes y de la molécula de ADN (Dawkins, 1976 p. 23). La reproducción de los virus implica la replicación de su material genético (Kay, 1986) y, para que dicha replicación pueda darse, se requiere algún soporte físico: la maquinaria celular.

Liberación de información

La principal implicación de la simplicidad morfológica de un virión es la necesidad de contar con la maquinaria necesaria para ejecutar la información auto-replicativa que porta. El virus, por sí mismo, cuenta con un soporte físico para almacenar y proteger la información; no

obstante, dicho soporte no es lo suficientemente complejo en términos causales, como para ejecutar esa misma información. Por este motivo, el virus necesita una maquinaria más compleja en términos causales y en términos de las transformaciones físicas que puede realizar. Cuando el virus entra en contacto con una célula, debe liberar la información genética viral para que ésta sea ejecutada en el soporte físico adecuado para realizar las transformaciones del sistema.

Analizando el caso del virus de VIH, se encuentra que una vez dentro de la célula, el gen viral "*Nef* puede modificar la célula para que ésta pueda manufacturar viriones de VIH". (Hutchinson, 2001 p. 89). Para el efecto, se debe integrar la información genética viral a la información genética de la célula (Nahmias y Reanney, 1977). De esta manera, las instrucciones originalmente portadas en la cápside del virus, se combinan con las instrucciones inicialmente almacenadas y ejecutadas en la célula que ahora servirá para ejecutar la información viral. La información genética almacenada y ejecutada en el núcleo de la célula indica las funciones que la célula debe cumplir; por lo tanto, estas funciones se alteran cuando la información genética viral comienza a ejecutarse luego de combinarse con la información genética celular. Cuando la información inicialmente ejecutada en la célula se transforma, el funcionamiento de la célula cambia como resultado de que" […] La doble cadena retroviral de ADN del VIH se mueve al núcleo, donde se inserta dentro del ADN de la célula […]. La infección de la célula es, entonces, permanente" (Hutchinson, 2001 p. 88).

Cuando las instrucciones originalmente almacenadas en el núcleo de la célula son alteradas, la nueva versión de información -la que contiene también la información genética viral- comienza a ejecutarse para que toda la maquinaria trabaje en función de producir y empacar nuevas copias de la información originalmente almacenada en el virus (Leigh-Brown y Holmes, 1994). Cada paquete de información será un nuevo virión: "La copia de ADN del virus es entonces integrada al genoma de la célula hospedera, donde reside como un provirus y es replicado" (Leigh-Brown y Holmes, 1994 p. 129).

Cuando la información genética viral es liberada dentro de la célula, la maquinaria celular es usada para dos eventos principales: (i) la síntesis de nuevas proteínas que se expresarán secuencialmente y (ii) la

replicación o duplicación de sus ácidos nucleicos con el propósito de producir un nuevo linaje. Luego, cada nuevo virión abandona la célula infectada y continúa con el mismo proceso en otra célula saludable. Este proceso es lo que se conoce como infección.

Una vez combinada con la información genética originalmente almacenada y ejecutada en el núcleo de la célula, la información genética viral puede ejecutarse inmediatamente o permanecer dormante. En el segundo caso, se apagan los genes virales, excepto aquellos asociados con la transcripción a la latencia: "el provirus puede permanecer dormante por un extenso periodo de tiempo. Sus genes no pueden expresarse hasta que las copias de RNA sean hechas por la maquinaria de transcripción de la célula hospedera" (Hutchinson, 2001 p.88). En este caso, se observan tres fases en el ciclo latencia-reactivación: (i) establecimiento, (ii) mantenimiento y (iii) reactivación. El estado dormante puede permanecer, incluso, durante toda la vida de la célula hospedera. La reactivación está asociada con cambios en las proteínas u hormonas relacionadas con la regulación del sistema inmune, de manera que cuando se presenta un cuadro de inmunosupresión o inmunodepresión, el virus puede reactivar su ciclo de auto-replicación. (Dimmock y Primrose, 1987; Fauci, 1988; Greene, 1992).

Infectando la distinción entre soporte lógico y soporte físico

Cuando el virus libera la información genética para ser ejecutada en la maquinaria celular, es evidente la distinción funcionalista básica entre el soporte físico y el soporte lógico: el virus desecha el soporte físico en el que su información genética viral estuvo inicialmente almacenada, y luego libera y ejecuta la información genética dentro de la maquinaria de la célula hospedera con el fin de completar el proceso auto-replicativo. Esto quiere decir que en la naturaleza puede encontrarse al menos un fenómeno en el que la información es separada de su medio inicial de almacenamiento y luego ejecutada en un nuevo soporte físico.

Las siguientes ideas sustentan la existencia de un sistema natural en el que el soporte físico y el soporte lógico pueden distinguirse: (i) el ADN es una molécula informacional, (ii) los genes son

información y algoritmos -y no objetos físicos- almacenados en la molécula de ADN, (iii) los virus solamente portan genes, *ergo*, (iv) un virus solamente porta información auto-replicativa. Estas ideas están sustentadas por experimentos y trabajos teóricos en la biología molecular y el análisis computacional. De hecho, las ideas (i) y (ii) constituyen parte del "dogma fundamental" de la biología molecular. En éstas áreas de investigación el orden algorítmico en los pares de bases de los ácidos nucleicos es considerado como patrones de información; alterar dichos patrones de información modifica las transformaciones físicas del sistema.

Searle (1994) no reconoce la relevancia de los patrones de información -genes- en las transformaciones físicas y biológicas que permiten la existencia de las especies. Si Searle, entre otros, quiere sostener que las computaciones, la información y la sintaxis no existen en la naturaleza, sino que son totalmente relativos al observador, entonces debe negar que los genes son información e instrucciones para replicar la vida. De hecho, Searle tendría que negar que cuando los genes se ejecutan, un proceso de auto-replicación se completa y que alterar la sintaxis de la información genética modifica el tipo y orden de las transformaciones físicas del sistema. Searle tendría que asegurar que los genes son únicamente objetos físicos y que cuando se habla de genes no se hace referencia a información que puede distinguirse y separarse del soporte físico en el que se almacena o ejecuta. No parece una tarea sencilla mostrar que la información no existe en la naturaleza, porque diversas observaciones de la biología contemporánea prueban que una vez dentro de la célula, los genes son instrucciones que se ejecutan para producir nuevas copias del virión inicial. Si Searle quiere negar la relevancia y existencia de la información, las computaciones, las instrucciones y la sintaxis, entonces debería discutir directamente con la actual biología teórica y experimental. Específicamente, Searle tendría que argumentar en contra de las siguientes ideas: en los virus hay un soporte físico que es desechado cuando éste libera el soporte lógico - ADN o ARN- a la célula, para que sea ejecutado en una nueva maquinaria.

La existencia de un fenómeno natural en el que la distinción *software/hardware* es evidente no implica ni es argumento a favor de la noción computacional de la mente. No obstante, el proceso de infección de los virus sí evidencia la posibilidad de separar información

natural de su soporte físico inicial, para que dicha información luego sea almacenada y ejecutada en otro soporte físico. Esta noción es coherente con la posibilidad de, por ejemplo, almacenar y ejecutar información genética en soportes físicos distintos a su soporte original. De hecho, esta noción es totalmente aceptada y reconocida por los genetistas y la mayoría de biólogos moleculares quienes, por ejemplo, pasan información genética de una célula A a una célula B, para alterar las transformaciones de B. Así lo demuestra la creación de cromosomas artificiales o la simple alteración de la sintaxis de un genoma para modificar los cambios fenotípicos de una especie o de un organismo particular. En este orden de ideas, la ingeniería genética no es más que el paso de leer el genoma de un organismo, a reescribirlo con el propósito de, incluso, crear nuevas especies como el *Mycoplasma laboratorium*. A finales de 2007, Craig Venter declaró haber creado una bacteria natural/artificial gracias a la elaboración de un cromosoma totalmente artificial. Dado que un cromosoma es la contracción de una larga cadena de material genético, se tiene que el *Mycoplasma laboratorium* es el resultado de modificar y reorganizar la sintaxis genética, para luego insertarla y ejecutarla en una nueva maquinaria distinta a la que inicialmente la almacenó: "el equipo de científicos ha logrado transplantar exitosamente el genoma de un tipo de bacteria en la célula de otro, cambiando efectivamente la célula de la especie" (The Guardian, 2007). Según Venter, la estabilidad metabólica y autoreplicativa de *Mycoplasma laboratorium* dependerá de la aceptación del genoma artificial que se inyectó en la maquinaria celular.

No sólo la naturaleza, sino las posibilidades tecnológicas también parecen cada vez más coherentes con el principio teórico del funcionalismo. De esta manera, la modificación genética y la ejecución de genomas alterados será, muy seguramente, una constante en las próximas décadas. Con el aumento progresivo y exponencial de la capacidad de procesamiento y almacenamiento de información, es probable que llegue el día en que, como lo señala Kurzwail (1999), todo la información de nuestra mente y que define lo que somos, pueda descargarse, almacenarse y ejecutarse en un soporte físico distinto al cerebro. Cuando esto suceda, tampoco será descabellado que nuestra mente, la información de lo que somos, simplemente viaje por redes sin necesidad de ejecutarse en un único soporte físico individual. De hecho, si se tiene en cuenta que lo que nos define en términos naturales no es más que información genética, entonces se concibe la

posibilidad, por ahora teórica, de que dicha información se almacene y ejecute en maquinarias inorgánicas, lo cual llevará a la aparición de *Robo sapiens*, como el siguiente paso evolutivo de *Homo sapiens* (Menzel y D´aluiso, 2003). Luego, probablemente estableceremos relaciones sociales indiferenciadas con los miembros de esta especie inorgánica cercana a *Homo sapiens* (Levy, 2007). Incluso, sin hablar de *Robo sapiens*, muchos teóricos contemplan la posibilidad de que en la materia inorgánica robótica se encuentre la evolución de Homo sapiens (Stock, 2002; Brooks, 2002; Moravec, 1999; Minsky, 1994; Salcedo-Albarán, 2003). Por otra parte, las investigaciones para construir nano-tubos que permiten establecer conexión directa entre nanomateriales y neuronas (Mazzatenta et al, 2007; Llinás et al, 2005), entre sistemas inorgánicos de computación y almacenamiento, y sistemas orgánicos de almacenamiento y procesamiento como nuestro cerebro, parecen acercarnos vertiginosamente al nacimiento de mentes híbridas e inorgánicas que resultarán de la oscilación entre el límite, ya difuso, de lo natural y lo artificial.

Sumario

a). Los virus son morfológicamente simples pues están compuestos de una cápside que encierra un núcleo con información genética.

b). Los genes son algoritmos -informáticos- y no son objetos físicos.

c). La información que se denomina *gen* se caracteriza por ser autoreplicativa.

d). Los genes están almacenados en la molécula de ADN, que en biología se define como una molécula informacional.

e). Aunque los virus cuentan con el soporte necesario para almacenar la información autoreplicativa, carecen de la maquinaria necesaria para ejecutar dicha información.

f). Los virus necesitan de la maquinaria celular para ejecutar su información genética, esto es, para auto-replicarse.

g). Con el fin de ejecutar el proceso de auto-replicación, el virus desecha su cápside y libera la información genética que almacenaba.

h). Cuando el virus desecha el soporte físico de almacenamiento, su información genética pasa al interior de la célula y se combina con la información genética que estaba almacenada y ejecutándose en el núcleo de la célula hospedera.

i). La información genética viral y la información genética de la célula hospedera se combinan. Cuando esto sucede, la maquinaria celular comienza a ejecutar las instrucciones que estaban contenidas en el material genético viral. Esto quiere decir que la maquinaria celular deja de cumplir sus funciones normales, y comienza a trabajar en la producción de copias de información viral idénticas a la información almacenada en el virus. Después de producir copias de información genética artificial, dichas copias son empacadas y liberadas de la célula para que repliquen el mismo procedimiento en otra célula saludable. Este proceso se conoce como infección.

j). A partir de todo lo anterior, se puede concluir que en la naturaleza sí hay, al menos, un fenómeno en el que se observa la existencia de información que puede cambiar de medio de almacenamiento y de medio de ejecución. De esta manera, en la naturaleza sí se pueden encontrar sistemas en los que se hace evidente la distinción funcionalista entre soporte físico y soporte lógico.

Bibliografía

Adleman, Leonard. (1998). "Computing with DNA". En *Scientific American*, agosto.

Adolphs, Ralph y Tranel, Daniel. (2000). "Emotion recognition and the Human Amygdala". *En The amygdala. A functional analysis*, J. P. Aggleton (editor). New York: Oxford University Press, pp. 587 - 630.

Adolphs, Ralph; Baron-Cohen, Simon y Tranel, Daniel (2002). "Impaired Recognition of Social Emotions Following Amygdala Damage". En *Journal of Cognitive Neuroscience*, 14 (8), pp. 1 - 11.

Adolphs, Ralph; Tranel, D.; Damasio, H. y Damasio, A. (1994). "Impaired Recognition of Emotion in Facial Expressions Following Bilateral Damage to the human Amygdala". En *Nature*, 372, pp. 669 - 672.

Adolphs, Ralph; Tranel, Daniel y Damasio, Antonio. (2003). "Dissociable neural systems for recognizing emotions". En *Brain and Cognition*, 52, pp. 61 - 69.

Aydede, Murat (2004). "The Language of Thought Hypothesis". En *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)*, Edward N. Zalta, editor.[consultado julio 25 de 2008] *URL* = <http://plato.stanford.edu/archives/spr2004/entries/language-thought/>.

Baron-Cohen, Simon (2000). "Autism and Theory of Mind". En *The Applied Psychologist*, Hartley, J y Braithwaite, A. (editores). Philadelphia: Open University Press.

Bechtel, William (1985). "What Happens to Accounts of Mind-Brain Relations If We Forego an Architecture of Rules and Representations?". In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1986, Volume One: Contributed Papers, pp. 159 - 171.

Bechtel, William (1991). *Filosofía de la mente*. Madrid: Tecnos.

Bhalla, V; Bajpai R. y Bharadwaj, L. (2003) "DNA electronics ". En *EMBO reports*, Vol. 4, No. 5, pp. 442 - 445.

Block, Carolyn (2003). "Extending Moore´s Law with Nanotechnology". Intel, documento de investigacion número: IR-TR-2003-15.

Block, Ned (1980). "Are Absent Qualia Impossible?". En *The Philosophical Review*, Vol. 89 (2-Apr.), pp. 257 - 274.

Block, Ned (1981). "Psychologism and Behaviorism". En *The Philosophical Review*, Vol. 90(1), pp. 5 - 43.

Block, Ned (1995). "The Mind as the Software of the Brain". En *An Invitation to Cognitive Science*. D. Osherson, L. Gleitman, S. Kosslyn, E. Smith y S. Sternberg, editores. Cambridge: MIT Press.

Block, Ned (1995). "Las dificultades del funcionalismo". En *Filosofía de la mente y ciencia Cognitiva*. Eduardo Rabossi, compilador. Barcelona: Paidós, pp. 105 - 142.

Block, Ned (1996). "What is Functionalism". En *The Encyclopedia of Philosophy Supplement*. Donald M. Borchert, editor. New York: MacMillan.

Boden, Margaret (1990). "Escape de la habitacion china". En *Filosofía de la inteligencia Artificial*. Margaret Boden, compiladora. México: Fondo de Cultura Económica, pp. 105 - 121.

Breazeal, Cynthia y Brian Scassellati (2000). "Infant-like Social Interactions Between a Robot and a Human Caretaker". En *Adaptive Behavior,* Vol. 8, No. 1.

Brooks, Rodney (2002). *Flesh and Machines*. New York: Panteón Books.

Brooks, Rodney; Breazeal, Cynthia; Marjanovic, Matthew; Scassellati, Brian y Williamson, Matthew (1998). "The Cog Project: Building a Humanoid Robot". En *Computation for Metaphors, Analogy and Agents*, Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*. New York: Springer-Verlag.

Bull, J. y Wichman, H. (2001). "Applied Evolution". En *Annual Review of Ecology and Systematics*, Vol. 32, pp. 83 - 217.

Chalmers, David (1994). *A Computational Foundation for the Study of Cognition*. [consultado mayo 25 2004]. URL = <http://jamaica.u.arizona.edu/~chalmers/papers/computation.html>.

Churchland, Patricia (1999). "Densmore and Dennett on Virtul Machines and Consciousness". En *Philosophy and Phenomenological Research*, Vol. 59, No. 3. (Sep.), pp. 763 - 767.

Churchland, Patricia S. y Sejnowski, Sejnowski J. (1990). "Neural Representation and Neural Computation". En *Philosophical Perspectives*, Vol. 4, *Action Theory and Philosophy of Mind*, pp. 343 - 382.

Churchland, Paul M. y Churchland, Patricia S. (1990). "Could a Machine Think?". En *Scientific American*, Vol. 262, Issue 1 (Ene.), pp. 26 - 31.

Dawkins, Richard (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, Daniel (1991). *La actitud intencional*. Barcelona: Gedisa.

Dennett, Daniel (1999). *La peligrosa idea de Darwin*. Barcelona: Galaxia Gutenberg.

Descartes, René (1641/1967). *Meditaciones Metafísicas*. Buenos Aires: Aguilar.

Descartes, René (1644/1995). *Los principios de la filosofía*. Madrid: Alianza.

Diccionario enciclopédico Salvat (1994). Frances Navarro, dirección editorial. Barcelona: Salvat Editores.

Dimmock, N. y Primrose, S. (1987). *Introduction to Modern Virology*. Oxford: Blackwell Sci.

Dirks, R., Lin, M., Winfree, E. y Pierce, N. (2004). "Paradigms for computational nucleic acid design". En *Nucleic Acids Research*, Vol. 32(4), pp. 1392 - 1403.

Fauci, Anthony S. (1988). "The human immunodeficiency virus: infectivity and mechanisms of pathogenesis". En *Science*, 239 (3840), pp. 617 - 22.

Fellous, J-M. (1990). "The Neuromodulatory Basis of Emotion". En *The Neuroscientist*, No. 5, Vol. 5 (1999), pp. 283 - 294.

Fodor, Jerry (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Fodor, Jerry (1981). "El Problema cuerpo-mente". En *Investigación y Ciencia*, Número 57, pp. 62 - 75.

Fodor, Jerry (1991). *La explicacion psicologica*. Madrid: Cátedra.

Fodor, Jerry (1994). *Psicosemántica*. Madrid: Tecnos.

Frege, G. (1991), "Sobre sentido y referencia". En *La búsqueda del significado*. Luis Valdes, compilador. Madrid, Tecnos.

Gallese, Vittorio; Fadiga, Luciano; Fogassi, Leonardo y Rizzolatti, Giacomo (1996). "Action recognition in the premotor cortex". En *Brain*, 119, pp. 593 - 609.

Gallese, Vittorio; Keysers, Christian y Rizzolatti, Giacomo (2004). "A unifying view of the basis of social cognition". En *Trends in Cognitive Sciences*, Vol. 8, No. 9, pp. 396 - 403.

Garner, Michael (2003). "Nano-materials & Silicon Nano-technology". Intel, documento de investigacion número: IR-TR-2003-16.

Greene W. (1993). "AIDS and the immune system". En *Scientific American*, 269 (3), pp. 98 - 105.

Grice, Paul (1991). "Las intenciones y el significado del hablante". En *La búsqueda del significado*. Luis Valdes, compilador. Madrid, Tecnos, pp. 481 - 510.

Guardian, The (2007). "I am creating artificial life, declares US gene pioneer". En *The Guardian*, octubre 6.

Heider, Fritz y Simmel, Marianne (1944). "An Experimental Study of Apparent Behavior". En *The American Journal of Psychology*, Vol. 57, No. 2. (Apr.), pp. 243 - 259.

Horgan, Terence (1984). "Functionalism, Qualia, and the Inverted Spectrum". En *Philosophy and Phenomenological Research*, Vol. 44, No. 4. (Jun.), pp. 453 - 469.

Husserl, E. (1980). *Experiencia y juicio*. México: Folio.

Hutchinson, J. (2001). "The Biology and Evolution of HIV". En *Annual Review of Anthropology*, Vol. 30, pp. 85 - 108.

Kay, L. E. (1986). "W. M. Stanley's Crystallization of the Tobacco Mosaic Virus, 1930-1940". En *Isis*, Vol. 77 (3), pp. 450 - 472.

Kim, Jaegwon (1995). "Mental Causation: What? Me Worry?". En *Philosophical Issues*, Vol. 6, Content, pp. 123 - 151.

Kitcher, Philip. (1982). "Genes". En *The British Journal for the Philosophy of Science*, Vol. 33 (4), pp. 337 - 359.

Lederman, M. (1987). "'Genes' Amplified". En *The British Journal for the Philosophy of Science*, Vol. 38 (4), pp. 561 - 566.

Leigh-Brown, A. y Holmes, E. (1994). "Evolutionary Biology of Human Immunodeficiency Virus". En *Annual Review of Ecology and Systematics*, Vol. 25, pp. 127 - 165.

Levine, A. (1991). *Viruses*. New York: Scientific American Library.

Lewontin, R. (1993). *Biology as Ideology: The Doctrine of DNA*. New York: Harper Collins.

Llinás RR, Walton KD, Nakao M, Hunter I y Anquetil PA (2005). "Neurovascular central nervous recording/stimulating system: using nanotechnology probes". En *J. Nanoparticle Res*, 7, pp. 111 - 127.

Llinás, Rodolfo (2002). *El cerebro y el mito del yo*. Bogotá: Editorial Norma.

Lycan, William G. (1987). *Consciousness*. Cambridge: MIT Press.

Maturana, Humberto y Varela, Francisco (1990). *El árbol del conocimiento*. Madrid: Debate.

Mazzatenta, Andrea; Giugliano, Michele; Campidelli, Stephane; Gambazzi, Luca; Businaro, Luca; Markram, Henry; Prato, Mauricio y Ballerini, Laura (2007). "Interfacing Neurons with Carbon Nanotubes: Electrical Signal Transfer and Synaptic Stimulation in Cultured Brain Circuits". En *The Journal of Neuroscience*, June 27, 27 (26), pp. 6931 - 6936.

McCarthy, John (1979). *Ascribing Mental Qualities to Machines*. Stanford Artificial Intelligence Lab, *Computer Science Report*, No. STAN-CS-79-725, AIM-326.

McCorduck, Pamela (1991). *Máquinas que piensan*. Madrid: Tecnos.

Menzel, Peter y D´aluiso, Faith (2003). *Robo Sapiens*. MIT Press.

Miller, G.A. y Gildea, P.M. (1987). "Cómo aprenden las palabras los niños". En *Investigación y Ciencia*, No. 134, pp. 80-85.

Minsky, Marvin L. (1985). "Why People Think Computers Can't Think?". En *The Computer Culture*. Donnelly, editor. Cranbury NJ: Associated Univ. Press.

Minsky, Marvin L. (1994). *Will Robots Inherit the Earth?*. [consultado 25 de julio de 2008]. URL = <http://web.media.mit.edu/~minsky/papers/sciam.inherit.html>.

Moore, Gordon (1965). "Cramming More Components Onto Integrated Circuits". En *Electronics*, No. 38, abril 19.

Moravec, Hans (1988). "Mind in Motion". En *Mind Children*. London: Harvard University Press, pp. 6 - 12.

Moravec, Hans (1994). *Robots, Re-Evolving Mind*. Octubre. [consultado julio 25 de 2008].URL = <http://www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/2000/Cerebrum.html>.

Moravec, Hans. "Rise of the Robots". En *Scientific American*, (Dic.), pp.124-135.

Moural, Josef (2003). "The Chinese Room Argument". En *John Searle*. Barry Smith, editor. New York: Cambridge University Press, pp. 214 - 260.

Nahmias, A.J. y Reanney, D. C. (1977). "The Evolution of Viruses". En *Annual Review of Ecology and Systematics*, Vol. 8, pp. 29 - 49.

Nelson, R. J. (1976). "Mechanism, Functionalism, and the Identity Theory". En *The Journal of Philosophy*, Vol. 73, No. 13, *The Mind and the Self*. (Jul. 15), pp. 365 - 385.

Oyama, S. (1985). "The Ontogeny of Information: Developmental Systems and Evolution". Cambridge: Cambridge University Press.

Parnavelas, John (1998). "The human Brain: 100 Billion Connected Cells". En *From Brain to Consciousness?* Steven Rose, editor. New Jersey: Princeton University Press, pp. 18 - 32.

Penrose, Roger (1991). *La nuvea mente del emperador.* Barcelona: Grijalbo Mondadori.

Pescador H., José (1989). *Principios de filosofía del lenguaje*. Madrid: Alianza.

Pinker, Steven (2007a). "A brief history of violence". Video filmado en marzo de 2007 y publicado en septiembre de 2007 [Consultado en diciembre de 2007]. Disponible en <http://www.ted.com/index.php/talks/steven_pinker_on_the_myth_of_violence.html>

Pinker, Steven (2007b). "A history of Violence", Edge, The Third Culture. [Consultado en diciembre de 2007]. Disponible en <http://www.edge.org/3rd_culture/pinker07/pinker07_index.html>

Purves, W.; Sadava, D.; Orians, G. y Heller, H. (2001) *Life*: *The Science of Biology*. Massachusetts: Sinauer.

Radosavljevic, Marko y Chau, Robert (2004). "Nanotechnology for High-Performance Computing". Intel, documento de investigación número: IR-TR-2004-3.

Rapaport, William (1995a). "Searle´s Experiments with Thought". En *Philosophical of Science*, Vol. 53, Issue 2, (Jun), pp. 271 - 279.

Rapaport, William (1995b). "Understanding Understanding: Syntactic Semantics and Computational Cognition". En *Philosophical Perspectives*, Vol 9, Issue AI, *Connectionism and Philosophical Psychology*, pp. 49 - 88.

Rapaport, William J. y Kibby, Michael W. (2002) "Contextual Vocabulary Acquisition: A Computational Theory and Edutational Curriculum". En *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics*. Vol. II: *Concepts and Applications of Systemics, Cybernetics, and Informatics*. Nagib Callaos, Ana Breda y Ma. Yolanda Fernandez J., editores. Orlando: International Institute of Informatics and Systemics, pp. 261-266.

Rizzolatti, G.; Fadiga, L.; Gallese, V. y Fogassi, L. (1996). "Premotor cortex and the recognition of motor actions". En *Cognitive Brain Research*, 3 (2), pp. 131 - 141.

Robbins, Trevor (1998). "The Pharmacology of thought and Emotions". En *From Brain to Consciousness?* Steven Rose, editor. New Jersey: Princeton University Press, pp. 33-52.

Ryle, Gilbert (1967). *El concepto de lo mental*. Buenos Aires: Paidós.

Salcedo-Albarán, Eduardo (2004). "La Constitución epistemológica del Universo". En *Borradores de Método*, No. 19, Bogotá, Colombia.

Salcedo-Albarán, Eduardo (2004). "Robots, la Gran Cadena del ser y evolución". En *Borradores de Método*, No. 17, Bogotá, Colombia.

Salcedo-Albarán, Eduardo; Zuleta, María-Margarita; León-Beltrán, Isaac y Rubio, Mauricio. (2007). *Corrupción, cerebro y Sentimientos. Una indagación neuropsicológica en torno a la corrupción*. Bogotá: Editorial Método.

Salcedo, Eduardo (2007). "Decodificando a *Matriz*, la película". En *Los Límites de la Estética de la Representación*, Chaparro, Adoldo (editor). Bogotá: Editorial Universidad del Rosario, pp. 288 - 324.

Scassellati, Brian (2001). "Foundations for a Theory of Mind for a Humanoid Robot". Tesis doctoral. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, Junio.

Schaffner, K. F. (1998). "Genes, Behavior, and Developmental Emergentism: One Process, Indivisible?". En *Philosophy of Science*, Vol. 65 (2), pp. 209-252.

Schank, Roger y Abelson, Robert (1987). *Guiones, planes, metas y entendimiento. Un estudio de las estructuras del conocimiento humano*. Barcelona: Paidós.

Searle, John (1990). "Mentes, Cerebros y Programas". En *Filosofía de la inteligencia Artificial*. Margaret Boden, compiladora. México: Fondo de Cultura Económica, pp. 82 - 104.

Searle, John (1991). "Qué es un acto de habla". En *La búsqueda del significado*. Luis Valdes Villanueva, editor. Madrid: Tecnos, pp. 24-45.

Searle, John (1992). *Intencionalidad*. Madrid: Técnos.

Searle, John (1994). *The Rediscovery of the Mind*. Cambridge: MIT Press.

Searle, John (1995). "Mentes y cerebros sin programas". En *Filosofía de la mente y ciencia Cognitiva*. Eduardo Rabossi, compilador. Barcelona: Paidós, pp. 413 - 443.

Searle, John (1997a). *La construccion de la realidad social*. Barcelona: Paidós.

Searle, John (1997b). *The Mistery of Consciousness*. New York: Nyrb.

Shapiro, Stuart C. y Rapaport, William J. (1987). "SNePS Considered as a Fully Intensional Propositional Semantic Network". En *The Knowledge Frontier: Essays in the Representation of Knowledge*. Nick Cercone y Gordon McCalla, editores. New York: Springer-Verlag, pp 262-315.

Smart, J.J. (1962) "Sensations and brain process". En *The Philosophy of Mind*. V. Chapell, editor. New York: Prentice Hall.

Stephen, Reed y Lenat, Douglas (2002). "Mapping Ontologies into Cyc". En *AAAI 2002 Conference Workshop on Ontologies For The Semantic Web*, Edmonton, Canada, Julio.

Stine G.J. (2000). *AIDS Update 2000: An Annual Overview of Acquired Immune Deficiency Syndrome*. Englewood Cliffs, NJ: Prentice Hall.

Stock, Gregory (2002). *Redesigning Humans. Our inevitable Genetic Future*. New York: Houghton Mifflin Company.

Turing, Alan (1950). "Computing machinery and intelligence". En *Mind*, Vol. 59, No. 236, pp. 433 - 460.

Turing, Alan (1950/1990). "La maquinaria de computación y la inteligencia". En *Filosofía de la inteligencia Artificial*. Margaret Boden, compiladora. México: Fondo de Cultura Económica, 1990, pp. 53 - 80.

Waters, C. K. (1994). "Genes Made Molecular". En In *Philosophy of Science*, Vol. 61 (2), pp. 163-185.

Williams, George C. (1996) . "A Package of Information". En *The Third Culture*. John Brockman, editor. New York: Touchstone Books. pp. 38-50.

Winfree, E. (2003a). "DNA Computing by Self-Assembly". En *The Bridge*, Vol. 33 (4), pp. 31 - 38.

Winfree, Erik (2003). "DNA Computing by Self-Assembly". En *The Bridge*, Vol. 33, No. 4, pp. 31 - 38.

Winfree, Erik y Bekvolato, Renat (2003). "Proofreading Tile sets: Error Correction for Algorithmic Self-A-ssembly". En *Ninth International Meeting on DNA Based Computers*, Junio.

Este libro se terminó de imprimir
en los talleres de Imprenet y Contextos Gráficos
en el mes de febrero de 2009.