

Universidad de Chile
Facultad de Filosofía y Humanidades
Departamento de Filosofía
Filosofía de la Mente y de la Inteligencia Artificial
Profesor: Rodrigo González
Alumno: Demian Schopf



La Máquina de Coser:

Inteligencia Artificial y Lenguaje Natural

Abstract

El presente trabajo se concentra en la pregunta acerca de la relación entre Inteligencia Artificial y Lenguaje Natural, y particularmente en la relación de este problema en atención al ‘modelo computacional de la mente’. Dos ensayos han sido referencia principal: *Minds, Brains, and Programs*, de John Searle y *The Computer Model of the Mind* de Ned Block. Se propone, además, un experimento mental que descansa en una aplicación real en la que ha trabajado quién escribe y suscribe esta introducción (y que puede ser visitada en la Web, en las direcciones indicadas en la correspondiente nota al pie de página). Las agudas distinciones de Searle entre ‘computación’ y la dimensión propiamente ‘simbólica’ de los signos proporciona un paradigma insospechadamente fructífero para leer críticamente las intrincadas descripciones de Block. Ello queda ejemplarmente bien expuesto en el experimento mental de ‘La Pieza China’ (*The Chinese Room*). Por otra parte, el experimento mental propuesto en este texto ha permitido, creo, expresar de un modo aún más claro como es que la dimensión semántica podría ‘emerger’ (estadísticamente) de los ámbitos de la sintaxis (y cómo podría ‘no emerger’ del mismo modo que en una mente humana). Imprescindible señalar que este trabajo no pretende agotar la riqueza del argumento de Searle sino aspectos muy particulares del mismo (así, no se examinan, por ejemplo, las seis réplicas a ‘La Pieza China’ en su totalidad, sino las nociones más relevantes para este trabajo: las consideraciones sobre la causalidad y los contextos causales). Igualmente necesaria ha sido la aguda lectura de Searle en *The Chinese Room: are programmed computers minds?* de Rodrigo González.

* * *

Inteligencia Artificial y Lenguaje Natural.

¿Es posible que toda la dimensión semántica del lenguaje natural no sea más que una propiedad de su computación sintáctica? (tal como propone el funcionalismo) ¿Se agota la semántica lingüística mediante el análisis funcional de la sintaxis? Dentro de la Ciencia Cognitiva estas dos hipótesis, tan insoslayablemente entrelazadas, tienen numerosos apologetas. Uno de ellos —uno moderado— es Ned Block.

Según la lectura que Block hace de los rendimientos críticos del funcionalismo¹ respecto a este problema —el de la semántica como una propiedad emergente de la sintaxis— no cabría distinguir ‘inteligencia de máquinas’ de la inteligencia humana (de lo cual se desprende que la hipótesis de la mente humana como una máquina sintáctica que maneja una máquina semántica² no sólo se aplica a la Inteligencia Artificial, sino que puede comprenderse como una teoría con pretensiones antropológicas, que explicaría la genética de la semántica en el ámbito de los lenguajes naturales). Y si así fuera, habría que asumir que ambas ‘inteligencias’ (que ‘a la larga’ serían —y serán— la misma) tienen, efectivamente, comprensión de los lenguajes naturales. Más aún: la única inteligencia de la que se sabe que existe a ciencia cierta es la humana. De esto se desprende que si la Inteligencia Artificial ha de lograr su antropométrico objetivo, será capaz de producir un programa que tenga exactamente la misma comprensión de lenguaje que un ser humano.

El objeto de este trabajo es discutir algunas nociones que se han instrumentalizado tanto a favor como en contra de la hipótesis funcionalista y la manera como ésta entra en relación con la semántica de las expresiones lingüísticas. La primera de estas nociones es la de intencionalidad. En este punto es necesario señalar que para Block es posible que la inteligencia surja —literalmente— de un pantano sin que exista intencionalidad; más aún, el análisis funcional ciertamente no explica la intencionalidad³.

Del otro lado, John Searle le niega cualquier estado mental a cualquier artificio humano que acontezca fuera del cerebro — precisamente por que carece de intencionalidad: (“*Whatever it is that the brain does to produce intentionality, it cannot*

¹ Block, Ned, *The Computer Model of the Mind*, En Brook, Andrew y Stainton, Robert. **Knowledge and Mind. A Philosophical Introduction**, 2000, MIT Press, Cambridge, Mass, USA

² *Ibid.* Págs. 264 – 267.

³ Véase el *Gedankenexperiment* del *Swampman*. *Ibid.* Pág. 263.

*consist in instantiating a program since no program, by itself, is sufficient for intentionality*⁴). Para Searle la intencionalidad es una condición necesaria para que haya comprensión de lenguaje (y como adentro de un programa no hay intencionalidad difícilmente podrá encontrarse ahí inteligencia lingüística).

Es el mismo Block quién nos dice que existen ciertos estados mentales que carecen de contenido intencional⁵. Buen ejemplo son el dolor o el hambre, de lo cual se desprende implícitamente que para Block estos estados mentales no forman parte de la cognición (ni humana, ni mucho menos de máquinas programables). Aunque jamás se pronuncia explícitamente a favor de este argumento, tampoco se preocupa de la relación entre la cognición y los estados mentales sin contenido intencional. Deseamos dejar abierta esta pregunta que, si bien no podemos evitar mencionar, no obstante, no alcanzaremos a desarrollar aquí: ¿tienen relevancia cognitiva estados mentales como el dolor o el hambre, o se trata tan sólo de epifenómenos absolutamente irrelevantes?

Para John Searle una Máquina de Turing que lograra pasar el Test de Turing no sólo carece de estados mentales sin contenido intencional, sino lisa y llanamente de intencionalidad.

Si bien el objeto ejemplar de esta crítica, de que no hay intencionalidad alguna en un programa que pretenda comprensión de lenguaje, es el trabajo de Roger Schank y Robert P. Abelson —(SAM, Schank y Abelson, 1977)— ésta se comprende como extensiva a otros experimentos como *SHRDLU* (Winograd, 1973) y *ELIZA* (Weizenbaum, 1966). Aquí se hace preciso señalar que la dimensión experimental de estos programas es la demostración, material y materialista, de que una máquina puede tener comprensión de lenguaje (y resulta difícil desoír que justo en este ámbito el mismo Weizenbaum concibe su *ELIZA* como una parodia a la Inteligencia Artificial⁶).

Searle se aboca frontalmente a esta hipótesis mediante un no menos paródico experimento mental conocido como ‘La Pieza China’ (*The Chinese Room*). Como se establece desde la primera línea, éste se propone demostrar que las competencias cognitivas de la mente humana no pueden ser implementadas en ninguna otra máquina que no sea el cerebro (en contra de las pretensiones de la Inteligencia Artificial dura).

⁴ Searle, John R., *Minds, Brains, and Programs*, *The Behavioral and Brain Sciences*, vol. 3, 1980, Cambridge, University Press, USA

⁵ *Ibid.* Pág. 262.

⁶ A este respecto véase Weizenbaum, Joseph, *The Computer and Natural Language* en *Computer Power and Human Reason*, London, Penguin Books Ltd., 1984 (Primera edición: W.H. Freeman and Company, 1976, UK).

Para Searle es imposible que un programa como SAM — o cualquier otra implementación basada en una Máquina Universal de Turing— tenga intencionalidad, estados mentales o comprensión de lenguaje (no sólo de los lenguajes naturales sino de cualquier sistema lingüístico, incluyendo los argumentos matemáticos).

Cómo la Pieza China es una respuesta explícita al programa de Schank Y Abelson, nos es insoslayable dar una pequeña caracterización de éste. El programa simula la capacidad humana de ‘comprender’ historias y de ‘adivinar’ ciertos desenlaces no explícitamente expuestos a partir de datos que el relato le entrega (el programa es capaz de completar verosímilmente un ‘script’ igualmente verosímil). Esta aplicación, cuyo nombre completo es *Script Applier Mechanism*, tiene su base epistemológica en una teoría que supone que la comprensión humana descansa en un saber general que le permite anticipar situaciones (en este caso, y a modo ejemplar, las historias y los scripts que les dan espesura). Este saber general es, para sus autores, un verdadero ‘saber común’ —quizás una cuestión de ‘sentido común’— y se manifiesta en una combinación entre la condición compartida de necesidades y métodos estandarizados con saberes más específicos (aquellos requeridos para las tareas y situaciones específicas). Esta combinación estaría ‘internalizada’ a modo de ‘scripts’ en la mayoría de los individuos que comparten una cultura y por lo tanto es posible de ser implementada en un sistema como SAM. Por ejemplo (el ejemplo es de Searle): dada una historia donde un hombre entra a un restaurante, pide una hamburguesa y se la traen carbonizada, el programa es capaz de ‘inferir’ que el comensal no se comerá la hamburguesa (ni felicitará al cocinero, ni dejará propina). Por el contrario, si le traen la hamburguesa bien cocinada, el programa inferirá que el hombre se la comerá, mandará a felicitar al cocinero y le dará una buena propina al garzón (y no se necesita ser programador para reconocer aquí a un antepasado primitivo de los actuales sistemas expertos, en donde toda la información se encuentra codificada en la base de datos). No se trata de más que de un gigantesco organigrama de relaciones —de causas y efectos programados— similar a un mapa de un enorme bosque de árboles categoriales, de contenidos rígidamente asignados y, más importante, construido a partir de bifurcaciones binarias del tipo si Q y B conducen a P y P es diferente de Y, entonces no se llegará a Y ni desde Q ni desde B [pero si desde Z y desde A]). Lo binario no está en que A y Z sean dos opciones (podrían ser 22 o 33 0 666...) sino en que o se llega o no se llega. SAM es capaz de ‘adivinar’ el desenlace de un script determinado —de responder algunas preguntas acerca del desenlace— por el simple hecho de que adentro

de la base de datos la ruta que los vincula ya se encuentra preestablecida: su ‘navegación’ no es más que una ‘instanciación’ de algo que la preexiste virtualmente en potencia. En jerga informática se podría agregar que los datos están hipervinculados. Más sorprendente, por cierto, resultan las pretensiones de Schank y Abelson, de que el programa literalmente ‘comprende’ la historia y de que ‘explica’ la capacidad humana de comprenderlas (lo controversial no radica en lo que se espera del programa sino en lo que se espera que éste refleje de la cognición humana: que es una aplicación de ‘scripts’, representaciones y saberes —generales y específicos— más o menos estandarizados e internalizados en el individuo cognoscente. El sentido común sería un paquete de miles de scripts y el operador humano no más que un ‘procesador’ que sólo entiende literalmente ‘lo que hay que entender’⁷).

Searle replica mediante un *Gedankenexperiment* —modo de argumentación ciertamente no exento de riesgos para la epistemología⁸— que es el siguiente: John Searle se imagina al interior de una pieza cuya única comunicación con el mundo son dos ranuras. Por la primera, recibe *inputs* escritos en chino; por la segunda, emite *outputs* que constituyen una precisa respuesta en esta misma lengua. Afuera se encuentra un hablante nativo de chino quién —de ranura en ranura— sostiene una alegre conversación escrita con la pieza. Sin embargo, la parodia al trabajo de Schank es más detallada: el primer *input* es un conjunto de ideogramas chinos cuyo significado Searle ignora (sólo reconoce formas y contornos). El segundo *input* es otra secuencia de símbolos chinos, pero además un conjunto de reglas sintácticas escritas en inglés, mediante las cuales se está en condiciones de responder verosímilmente al interlocutor chino. El tercer *input* es otro conjunto de ideogramas, y otro conjunto de reglas escritas en inglés, que constituyen instrucciones precisas para cotejar los dos primeros *inputs*, —mediante una lectura comparada de las sombras chinescas— para emitir otro *output*. Sin que Searle lo sepa, las personas que le han proporcionado estos tres *inputs* se refieren al primero como ‘script’, al segundo como ‘historia’ y al tercero como ‘preguntas’. Esto significa que el organigrama relacional que vincula ‘scripts’, que permite adivinar el desenlace de ‘historias’ y que constituye las respuestas a las ‘preguntas’, no es más que una máquina lógica (por compleja ésta que sea). No es más que un conjunto de

⁷ A este respecto véase: González, Rodrigo, en *The Chinese Room Revisited. Artificial Intelligence and the Nature of Mind*, 2007, Katholieke Universiteit Leuven, Institute of Philosophy, Centre for Logical and Analytic Philosophy, Leuven, Bélgica, Págs. 73-76.

⁸ *Ibid.* . 3. 1 *The nature and use of thought experiments*. Págs. 82 – 88.

relaciones formales que preexisten a Searle en potencia, pero que van desplegándose en actos sucesivos. Es esto lo que permite que John Searle opere esa máquina sin entrar en relación con la dimensión semántica de los símbolos —y sin entrar en una relación intencional con este ámbito de lo simbólico (que para él define la totalidad y propiedad de lo simbólico). Simplemente se limita a responder con P ahí donde ve Q. Sigue el algoritmo específico para cada caso. A este conjunto los hablantes chinos le llaman ‘programa’. Para complejizar un tanto más el experimento, Searle imagina ahora que sus interlocutores le proporcionan relatos en inglés, que sí entiende semánticamente (por lo tanto está en condiciones de dar cuenta no sólo de la forma de cada historia sino también de su contenido). Si Searle y los programadores —en este caso Schank y Abelson— se ponen de acuerdo en guardar el secreto, podrán engañar a un interlocutor chino que los someta al Test de Turing (y por supuesto también a un interlocutor inglés que ignore que Searle está dentro de la pieza china — aunque para esto último no se requiere de la existencia de Schank y Abelson). En el caso chino Searle se limita a manipular signos cuyo ámbito semántico desconoce (salvo las relaciones sintácticas que hacen de ciertas combinaciones combinaciones válidas y de otras combinaciones combinaciones inválidas). En el primer caso Searle responde con esquemas sintácticos preexistentes, en el caso inglés responde al nivel de la semántica, que además puede modular en diversos esquemas sintácticos.

Pero bastaría la sola existencia de un elemento que no esté contenido en el organigrama —y que tampoco esté relacionado con él— para que el entendimiento del sistema colapse. Hemos considerado, a modo de ejemplo, variables como P, Q, B, Y, A y Z. Consideremos ahora la variable D, que no estaría contenida en ninguna de las posiciones del esquema de Schank y Abelson. Lo más probable es que el sistema de Schank y Abelson no sepa nada de D (y Searle tampoco). Alguien como Block quería creer que atendiendo a la sintaxis el sistema podrá ‘inferir’ que D podría ser ‘similar’ a ‘A’ y que en otro sistema es su equivalente (por ejemplo: $D = A$). ¡Pero D no existe! Lo que ni Schank, ni Abelson ni Block dicen es cómo este proceso tiene lugar (y lo que nosotros sabemos es que mientras la base de datos sea una base de datos cerrada donde D no existe, no es posible imaginar cómo es que tal sistema podrá ‘inferir’ lo que es D —o que D podría ser algo parecido a A). Ambos podrían replicar que un operador humano enfrentado a un dilema similar se encontraría en idénticas dificultades. Imaginemos a Searle de viaje por Egipto. De pronto ve una frase en árabe y quiere saber que significa. Le pregunta a un egipcio que habla inglés y éste le explica el significado

de la frase. Una instancia tan básica cómo ésta —la variable del ‘script’ D— no está contenida en el programa de Schank y Abelson, que sólo puede preguntar cosas relativas a sí mismo. Ahora bien: imaginemos que el programa de Schank y Abelson cuenta con una aplicación que va a Internet en busca del significado de D (llamaremos a esta aplicación ‘Máquina de Coser’⁹). Nótese que la Máquina de Coser tiene el algoritmo que tiene porque su Leitmotiv es la máxima funcionalista de que la dimensión semántica de la lengua es una propiedad emergente de su dimensión sintáctica; o para decirlo en palabras del análisis funcional: una propiedad derivada de su función. El robot busca D en Google y encuentra 25.440.000.000 resultados en sólo 23 segundos. Cabe imaginar que un sistema que cuente con una base de datos lo suficientemente extensa podría concluir el significado de esa variable (digamos, con la totalidad de combinaciones de caracteres alfanuméricos que se usan en la lengua castellana, exceptuando la variable D). Puede cotejarla con una definición (consultando el diccionario de la RAE). Pero aquí nos enfrentamos a un segundo problema: no hemos obtenido un resultado, sino 25.440.000.000 usos de un mismo símbolo ¿Cómo sabrá el programa en que contexto se le ha pedido ‘procesar’ la variable D?. Block quizás replicaría que un operador humano tendría idéntico problema (que Weizenbaum podría caracterizar con la siguiente frase: “*the house blue it*”¹⁰). Pero lo que Weizenbaum, Block y Searle saben, es que lo más probable es que el programa sólo reconozca combinaciones de caracteres alfanuméricos que procede a clasificar (y que en un nivel más bajo —el ‘Lenguaje de Máquina’— son enormes cadenas de 0 y 1). A cada caracter corresponde un largo código en notación binaria y, a su vez, las palabras son aún más largas cadenas formadas por los caracteres así codificados (lo cual incluye toda clase de signos ortográficos, caracteres en otros alfabetos —como el chino— y también espacios vacíos). Pero esta precisión nos ayuda en nada a que La Máquina de Coser ‘sepa’ lo que es D —suponiendo que esto es posible— (excepto que es una secuencia, indeterminada pero única, de 0 y 1). Como dijimos: encontró 25.440.000.000 variables en 23 segundos. Debemos darle un criterio adicional. Puede buscar en el diccionario de la RAE ¿Pero porque confiar sólo en la RAE cuando existen también otros diccionarios? ¿Quizás debería buscar en todos los diccionarios online y hacer una síntesis estadística? ¿Pero porque confiar en los diccionarios cuando el significado funcional de D es una

⁹ Esta aplicación ya existe, aunque sólo como un prototipo a menor escala (examina todos los ‘homes’ en español). Pueden verse videos en dos direcciones: <http://es.youtube.com/watch?v=QCav3LkjH-c> y <http://es.youtube.com/watch?v=5d3NNDDjdwk&feature=related>.

¹⁰ Weizenbaum. Op. Cit. Pág. 185.

propiedad emergente derivada de su dimensión sintáctica? Descartemos esta opción y supongamos que nuestro sistema puede procurarse dos métodos para saber lo que es D. El primero, es preguntarle a la RAE (pero esta opción ya está descartada, porque confiamos más en Block que en la RAE); el segundo, más acorde con el funcionalismo, es preguntarle a toda la World Wide Web: la panacea de las propiedades emergentes y de las estructuras relacionales. Este último método consistirá en lo siguiente: la Máquina de Coser buscará en toda la Web la variable D (ello es lógicamente posible, mas tecnológicamente muy lento y sobre todo muy costoso). Una vez que disponga del monstruoso número, nuestra máquina procederá a examinar todos los sitios que encontró (que con toda probabilidad superará los 25.440.000.000 encontrados por Google porque, verbigracia de los experimentos mentales, incluiremos también las páginas carentes de enlaces con otras páginas). El criterio a emplear será el siguiente: la máquina someterá a un cálculo a todas las palabras que aparecen junto a D, aislando toda frase en donde esté D (cuya extensión está precisamente delimitada por el punto anterior a la frase y el punto con el cual concluye). Posteriormente, procederá a hacer un ranking viendo cual es la palabra más frecuentemente asociada a D (para facilitarle el trabajo cometeremos un pequeño ‘pecado semántico’: le daremos una lista de todos los conectores, ya que de lo contrario es muy probable que la palabra más asociada a D resulte ser un artículo como por ejemplo ‘la’). Ahora bien: la máquina ha encontrado esa palabra. ¿Qué hacer con ella? Necesitamos una definición. Acorde con el funcionalismo implementaremos un ‘motor de escritura’ exclusivamente relacional. Es así como operará para producir sus definiciones ‘semánticas’: lo primero que hará será imprimir en pantalla la frase más frecuente con la variable D. Le seguirá la frase más frecuente con la palabra que aparece con mayor frecuencia junto a D (D2). La tercera frase será la más frecuente con la palabra —D3— que aparece con mayor frecuencia asociada a la palabra —D2— que aparece con mayor frecuencia asociada a D. Repetiremos recursivamente este proceso hasta agotar toda la Web. A esta segunda instancia es a la que le corresponde con mayor propiedad la rúbrica de ‘Máquina de Coser’, ya que de un modo parecido a como hace Searle dentro de la Pieza China, no hace más que unir símbolos (o cáscaras de símbolos) sin atender a su dimensión semántica. Por otra parte, puede entenderse como lo que Block llama un *Cruncher* de símbolos (en el sentido de ser un ‘desmenuzador’ de frases que aísla cada palabra) ¿Qué saben el motor de búsqueda y el motor de escritura que genera las definiciones? Saben —‘saben’ y no saben— precisamente lo que Block supone que es el saber acerca de la dimensión

semántica del lenguaje: que su significado emerge de su uso.

A lo más el programa podría ‘saber’ que ‘guerra’ (con minúscula) aparece frecuentemente en oraciones donde lo acompaña ‘Irak’ (lo cual no excluye, a favor de Block, el ‘saber’ de ‘Guerra’ que le proporciona la RAE (y todos los diccionarios de los que podrá servirse para sus cálculos estadísticos): que esta palabra aparece en un párrafo acompañada de muchas otras, y que estas son las que constituyen su definición. Aún así, es muy probable que encuentre muchas más frases donde ‘guerra’ aparece vinculada a ‘Irak’ que a cualquiera de las palabras que la acompañan en el diccionario de la RAE (como la Guerra del Pacífico o las miles de otras ‘guerras’ en minúscula). Por lo tanto, de acuerdo a la caracterización funcionalista, habría que admitir que la palabra ‘guerra’ está vinculada a la palabra ‘Irak’. Podríase categorizar la información para que la Máquina de Coser ‘sepa’ distinguir las palabras de los ‘diccionarios’ de todas las demás en cualquier otro cómputo estadístico que le pidamos (es decir, que compute todo, o determinada parte de esa totalidad, menos lo que está en los ‘diccionarios’). Hemos vuelto a ‘pecar’ a favor de Block. Sin embargo, darle esa ‘ayuda’ sería contradecir a Block. El único problema es que si no somos nosotros quienes le decimos que tiene que distinguir lo que encuentre en la RAE de lo encuentre en los demás dominios es muy difícil que sepa distinguir por sí sola. Por otra parte, si la ‘engañáramos’ ingeniándonoslas para que compute que ‘Irak’ = ‘Suiza’, el carácter relacional que otorga algo de sentido a ‘guerra’ + ‘Irak’ se transformaría en una aplicación cuyos gigantescos cómputos estadísticos están destinados a medir la frecuencia con que se informa sobre las atrocidades cometidas por las tropas de Bush en suelo suizo; donde presumiblemente existen redes de ‘Al-Qaeda’ y donde se ocultan (las palabras) ‘Muqtada’ y ‘Al-Sadr’ (que tampoco nuestra máquina podría diferenciar de ‘Muqthada’ que si bien tiene un número menor de ocurrencias que ‘Muqtada’, tiene la frecuencia suficiente como para ser tomada por el nombre de otra persona y no por el nombre mal escrito del primero). Hay que agregar que yo creo que ese es el nombre mal escrito — ‘Muqthada’— precisamente porque su número de frecuencias es menor, pero que a diferencia de Google al menos admito la posibilidad de que la mayoría escriba mal ese nombre. Lo que tengo claro es que ninguno de los dos es un líder integrista de nacionalidad suiza. Para dar un ejemplo (algo) menos drástico: podemos suponer que la variable ‘D2’ deja de ser ‘Irak’ y que eso inevitablemente modificará la cadena de frases relacionales que nazcan desde ‘D2’. El programa no tiene como ‘saberlo’ salvo que vayamos refinando más y más las categorías en relación a los vínculos (pero eso no es

Inteligencia Artificial —ni la aplicación de un funcionalismo duro—sino la intervención del ‘factor humano’). Nuestra máquina no podrá distinguir que ‘algo anda mal’, como presumiblemente podría hacerlo John Searle, si siguiendo el ‘script egipcio’ se encuentra con un escuadrón de Al Qaeda y no con la Esfinge de Gizeh.

Ciertamente la auto-organización se puede forzar cada vez más. Por ejemplo: podemos producir un programa que reconozca raíces comunes a ciertas palabras, lo cual reforzaría la idea de Block. De lo que no se sabe es de un programa capaz de darse a sí mismo esta clase de criterios relacionales, progresivamente más y más ‘semánticos’ —de engendrar programáticamente a un programador— y de prescindir completamente de un operador humano. Ciertamente existen intentos, pero los resultados son más bien magros y rústicos.

No obstante, creo que el interesante corolario de los mecanismos descritos por Block ayudan a comprender más que la ‘inteligencia humana’ —el ‘modelo computacional de la mente’— el funcionamiento ‘inteligente’ (o ‘no inteligente’, por carecer de ‘intencionalidad’, como afirma Searle) de ciertos mecanismos que ya existen en el ‘modelo computacional de la computación’: los ‘circuitos lógicos’, la descomposición recursiva, las ‘bases de datos’, las ‘librerías’, los ‘diccionarios’ y la ‘memoria’ RAM. Como el mismo Block reconoce, la respuesta a la hipótesis del cerebro como una máquina sintáctica a la conducción de una máquina semántica es una hipótesis empírica, que no puede ser resuelta por la especulación filosófica. Al otro lado, Searle admite la posibilidad de lo mental implementado en otra substancia que la del cerebro humano como problema fundamentalmente empírico. Pero de ello no se puede concluir que la Inteligencia Artificial haya logrado —ni vaya a lograr— emular lo mental mediante la implementación del funcionalismo (precisamente por que el funcionalismo no atiende a la dimensión material de la semántica mental, que parece ser el cerebro, y quizás la inconmensurable red causal que supone el mundo físico que interactúa con él)¹¹. Paradojalmente la Filosofía de la Mente aparece como el límite de una cuestión mental que no puede ser resuelta en términos meramente especulativos.

Creo que el pequeño experimento mental de la Máquina de Coser, si bien no es capaz de ‘falsear empíricamente’ que la mente es solamente un operador sintáctico, demuestra que la hipótesis de Block es menos falsa que drásticamente parcial, ya que si

¹¹ “*Perhaps other physical and Chemicals processes Could produce exactly these effects; perhaps, for example, Martians also have intentionality but their brains are made of different stuff. That is an empirical question, rather like the question wheter photosynthesis can be done by something with a chemistry different from that of clorophyl*”. Searle. Op. Cit. Pág. 11.

las propiedades semánticas de ‘guerra’ fueran solamente una propiedad emergente de sus relaciones sintácticas habría que concluir que ‘guerra’ e ‘Irak’ se condicionan mutuamente en términos más relevantes que cualquier definición de diccionario, que además la definición de ‘guerra’ cambiará considerablemente cuando se introduzcan las variables ‘Irán’, ‘bomba’ y ‘atómica’, y por último que cuando se le pregunte a alguien en la calle que entiende por guerra, este comenzará a hablar de Irak y en un tiempo más de Irán. De acuerdo a un criterio reductivo y extremadamente ‘inmanentista’ y ‘relacionalista’ sin duda que esto es algo posible y real, pero no creo que la totalidad de la semántica se agote en esta clase de reducciones. Una máquina podrá funcionar así, pero resulta difícil definir la subjetividad humana en términos de un inmanentismo absoluto, o al menos tan ‘intensamente inmanente’ como la Máquina de Coser. Además no está claro como la Máquina de Coser podría ‘cotejar’ sus definiciones ‘relacionales’ y aquellas provenientes de su ‘diccionario’ si no es a través de más estadística relacional y en virtud de más ‘pecados semánticos’ de parte de los programadores.

Es curioso que los partidarios del ‘modelo computacional de la mente’ (o los ‘conexionistas’) no se hayan interesado en Internet en donde la ‘semántica relacional’ encontraría un verdadero paraíso (por no decir un ‘universo paralelo’). Aún así, el hecho de que Google encuentre en 23 segundos 25.440.000.000 resultados para la variable D, y que yo no pueda acceder a ese mar de información, nos sugiere que de tener las máquinas —las redes de máquinas— ‘inteligencia’ ésta es sin duda muy distinta de la humana. Las máquinas carecen de algo así como ‘subjetividad’, salvo que se reduzcan las causas genéticas de la subjetividad a la existencia individual (desconociendo que ciertas patologías mentales se heredan genéticamente y que, además, la existencia condicionante de los seres humanos está repleta de estados mentales sin contenido intencional). ¿Cómo se llega a ser lo que se es sin estados mentales sin contenido intencional? Por otra parte, es perfectamente posible que la Máquina de Coser, que hemos esbozado, sea objeto de las mismas objeciones de Searle, las cuales se mantendrían incólumes en tanto continúe siendo una máquina sintáctica (como de hecho lo es: si bien las palabras no son ideogramas, si son ‘strings’ de letras que representan un objeto concreto: por ejemplo ‘Irak’).

¿Pero cual es la intencionalidad que Searle extraña en el funcionalismo? Ciertamente Block está dispuesto admitir una cosa a favor de Searle, y es que el análisis

funcional no explica la intencionalidad¹². Por otra parte, es el mismo Block quién afirma que puede haber inteligencia sin intencionalidad, lo que no queda claro es cómo adecua esta idea al carácter antropométrico de su concepto de inteligencia. ¡Nada menos que con un *Swampman*!¹³. Sin embargo, como observa Searle, en este carácter antropométrico habita una tremenda contradicción. Si lo que importa es el análisis funcional y lo que no importa es la implementación material —el cerebro—, entonces el ‘funcionalismo’ padece una suerte de ‘dualismo’. Si lo mental es separable de lo material, si se puede implementar lo mental en una máquina electrónica o en cualquier otro soporte que cumpla con las condiciones básicas de lo que el funcionalismo supone que es lo mental, entonces, por una parte, el cerebro es un mero accidente de la evolución, y por la otra, parece contradictoria la pregunta de ‘si puede pensar una máquina’ (ya que como vimos esas máquinas que son los cerebros serían un mero accidente evolutivo en donde lo mental habría encallado más por azar que por necesidad). La pregunta es si ‘puede pensar un programa’ que es otra clase máquina; a saber: una máquina lógica como la Máquina Universal de Turing. La respuesta de Searle es un tanto menos dualista: una máquina ‘puede pensar’ y esa máquina es el cerebro. En el ‘análisis funcional’ la inteligencia es literalmente ‘desmaterializada’ y secuestrada de las causas materiales que supone su residencia en el cerebro y los estímulos que éste recibe y procesa, arbitrariamente reducidos a puras relaciones formales *ex nihilo*. Block nunca llega a describir con la suficiente minuciosidad el recorrido de los estímulos desde los ‘transductores’ a los ‘procesadores primitivos’, sólo dice que tienen lugar y que los símbolos se generan en virtud de la función que cumplen¹⁴. Lo que no queda claro es si el procesador primitivo trata con dígitos o con otra clase de representaciones mentales, y sí así fuera no queda claro como un ‘procesador primitivo’ podría ‘procesar’, por ejemplo, la imagen de Ned Block (o que es lo que realmente procesa en relación o en lugar de la imagen de Ned Block). En otras palabras: no está claro cómo se codifica la imagen de Ned Block, sólo está claro que debería codificarse mediante una milagrosa conjunción de ‘procesadores primitivos’. Pero decir esto no es decir suficiente. Así, se define al cerebro por una parte como un computador, pero por la otra se lo define como una implementación material irrelevante. Pero para Searle el punto no es solamente ese. El punto es también que su causalidad

¹² Block. Op.Cit., Pág. 263.

¹³ *Ibíd.* Pág. 263.

¹⁴ *Ibíd.* Pág. 270.

para producir intencionalidad no residiría, según Searle, en ser la ‘instanciación’ de un programa, y que en consecuencia tampoco residiría ahí su capacidad de tener estados mentales. Como dijimos, la cuestión de la causalidad es aquí relevante, en la medida en que no parece formalizable por un algoritmo (que además tendría no sólo que ‘emular’ a una mente inteligente, sino probablemente todo el universo causal en que esa mente se desenvuelve y llega a ser lo que es). ¿Forman parte los contenidos no intencionales de este universo?

Además, las propiedades formales de una máquina sintáctica manejando una máquina semántica no serían en sí mismas constitutivas de intencionalidad (ahora la palabra ‘máquina’, con algo más de culpa, se nos ha hecho un tanto más incómoda y literalmente demasiado pesada: el ‘funcionalismo’ padece un prejuicio materialista y por otra parte pretende desembarazarse del cerebro). De esta manera, estas máquinas lógicas, que son los programas inteligentes, carecerían de todo poder causal, salvo cambiar de estado de acuerdo con el algoritmo del programa (algo similar a los autómatas finitos de Block¹⁵). ¿Pero que pasa con el ‘algoritmo’ del mundo causal que no reside en el programa sino en el mundo que lo condiciona o que, como diría Varela, “en el mundo que lo enactúa”¹⁶? Toda otra propiedad causal queda inmediatamente excluida, presa en una especie de ‘solipsismo’ del ‘script’ y del ‘computacionalismo’ (y no queremos insinuar que el funcionalismo excluye de raíz la causalidad, pero al menos el texto de Block parece insuficiente para dar cuenta de las distintas ‘interfaces’ que supone la interacción del mundo causal con un procesador primitivo, o con muchos de ellos). Dicho de otra manera: el programa no posee los medios para entrar en relación con una causalidad que no esté prevista en su algoritmo (que por definición es finito y que, además, no debería ser confundido con las posibilidades potencialmente infinitas de las combinaciones coherentes posibles de sus elementos). No hay nada que indique que a la mente humana funcione del mismo modo en su totalidad. Es quizás aquí donde el argumento de los estados mentales sin contenido intencional pueda cobrar alguna relevancia, puesto que las causas, por ejemplo, del dolor o del frío, son causas que no residen dentro de quien las padece —no son previsibles—, sino que son reacciones a un entorno determinado y dependen directamente de éste. Si hacen 15 grados centígrados un sujeto proveniente del Caribe se abrigará, en tanto que es probable que un esquimal

¹⁵ Ibid. Pág. 273.

¹⁶ Al respecto véase: Varela, Francisco J. *Conocer. las Ciencias Cognitivas: tendencias y perspectivas. Cartografía de las ideas actuales*, 1988, Ed. Gedisa, Barcelona, España.

no lo haga. Es evidente que para tener hambre es condición necesaria tener un cerebro (o un ‘programa’), pero es igualmente cierto que nadie sabe del momento exacto en que comenzará a sentir hambre, ni tampoco del potencial infinito de las causas y cadenas causales posibles como para implementar una simulación que de cuenta del medio ambiente en que la inteligencia humana tiene lugar. Al menos para una simulación dentro de una máquina esto resulta muy difícil de implementar. Incluir el potencial causal (probablemente infinito) de los estímulos posibles implicaría una formalización del universo completo. Como sabemos hay buenas teorías, y buenas evidencias, que sostienen que lo real no es completamente formalizable (especialmente en fenómenos concernientes al caos y al azar que por principio es no formalizable). No existe un script de lo infinito ¿Cómo podría existir? Buen ejemplo de lo aleatorio son las turbulencias que no se pueden prever. No basta con simular a una entidad ‘inteligente’ si no se lo provee de un contexto para que ejercite su inteligencia (y hasta es dudoso pensar que algo podría llegar a ser inteligente sin un correlato contextual). Resulta muy difícil explicarse la dimensión propiamente actual de los procesos intelectivos (que Searle localiza en las secuencias sinápticas) prescindiendo de la idea de un marco contextual en donde éstas tienen al menos parte de su origen causal más probable. Aquí es donde la cuestión de la intencionalidad pareciera volverse más problemática, sobre todo si ha de residir en la sintaxis simbólica, como querría Block. Esta es por definición sólo acerca de sí misma (al menos dentro de un programa). Además, la manipulación simbólica carece de intención desde el momento que deja de ser manipulación simbólica, en la medida en que ni Searle dentro de la pieza china —ni tampoco la Máquina de Coser procesando toda la World Wide Web— manipulan realmente símbolos, sino secuencias de cálculos cuyo significado no importa más que de acuerdo a un principio que siempre es atribuido por un tercero al *output* generado. La semántica de los mismos no es un asunto material (pues materialmente son tinta china sobre papel o combinaciones de 4 y 7 volts). Es un asunto de relaciones de cómputo que carecen de dimensiones cualitativas más allá de la mera computación. Como señala el mismo Block, es el operador humano quién mapea los fenómenos físicos y los transforma en fenómenos semánticos. En sentido estricto sólo poseen una sintaxis pero no poseen semántica. En el caso de Searle son ‘formas’; en el caso de la Máquina de Coser son cadenas de códigos binarios, que se computan a una gran velocidad según parámetros lógicos y árboles categoriales. Bastaría que los intérpretes humanos desaparecieran para que rápidamente se manifestara que estos ‘símbolos’, como querría Block, sólo existen

en la medida que se los interpreta, así como un billete pasaría a perder el ‘valor’ que tiene si la Humanidad desapareciera de la faz de la Tierra para no ser más que un trozo de celulosa. Aquí, curiosamente, el argumento que distingue entre ‘clases funcionales’ y ‘clases naturales’ es revertido por Searle en contra del funcionalismo, puesto que la ‘clase funcional’ no puede ser lo que es sin un marco contextual —es decir ‘funcional’— que la contenga y que el computacionalismo evidentemente no provee más que en el ámbito del cómputo de valores. La noción de algoritmo es por principio exclusiva de toda causalidad que no esté prevista en él. Además ¿quién ‘lee’ el cómputo? ¿la computadora o el usuario? Lo único que provee el computacionalismo es una ilusión muy confusa acerca de lo que es el ‘procesamiento de información’ en la medida en que usa estos términos reduciéndolos a una mera manipulación de cálculos que como vimos ni siquiera puede llamarse ‘símbolos’ porque carecen de espesor semántico propio (no son acerca de nada). Y pocas señales hay de que el programa pueda atribuirse alguna consistencia semántica más allá del cálculo. Generalmente, ésta surge de un operador humano que le asigna un valor a cada variable. Pero está ‘codificación’ no proviene de la máquina, salvo que se trate de argumentos exclusivamente matemáticos (por no decir computacionales: de más y más cálculos). Son cadenas de 4 y 7 volts, de gatos persiguiendo ratones¹⁷ o de máquinas hidráulicas que imitan redes neuronales. Verdadero para los cálculos pero falso para los símbolos. El hecho de que estos procesos sean utilizados como significantes de cuestiones que transcurren fuera de las computadoras no es, en estricto rigor, un asunto de las computadoras sino de quienes las utilizan. Es importante destacar que la dimensión propiamente ‘simbólica’ les es otorgada ‘fuera’ de la máquina y no es un producto de la misma máquina. La máquina no es ‘reflexiva’ en este sentido: es una computadora. Quizás es ese el hecho que explica porqué un gato que sale disparado detrás de un ratón podría equivaler a un golpe de corriente de 7 volts (porque convenimos que ambos ‘significan’ 1 —y ello no implica desconocer que para eso tenemos que ‘darnos’ ciertas condiciones propiamente ‘computacionales’, vale decir, ciertas relaciones matemáticas). Por cuestiones de ingeniería práctica se ha convenido reducir estas relaciones a 0 y 1; y a 4 y 7 volt. Es precisamente un problema matemático, y lógico matemático, lo que habilita distinguir entre lo ‘semántico’ y lo ‘sintáctico’. Por cierto: los argumentos matemáticos existen en potencia (y sin que se requiera de máquinas). Pero no está claro

¹⁷ Ibíd. Pág. 258.

cual sería su vínculo con lo semántico en la teoría de Block (más allá de que ciertos símbolos estarían en una relación computacional los unos con los otros). Resulta sintomático que el mismo Block nunca explique el tránsito de lo matemático a lo semántico en el sentido de la actividad simbólica de los lenguajes naturales. Podrá hablarnos de ‘transductores’, de ‘diccionarios’ y de que creer o desear algo es estar en una relación ‘computacional’ con ese algo¹⁸. La verdad es que lo único medianamente probable es que si alguien desea algo posee un estado mental con ese contenido intencional, junto a otros que carecen de contenido intencional. No está claro si éstos están dispuestos en una cinta de una Máquina Universal de Turing y si desfilan secuencialmente frente la puerta del deseo —o simultáneamente, como querría el ‘conexionismo’. Siendo no menos fantasioso: quizás a cada cosa corresponda una red que se comunica ocasionalmente con la red del deseo; pero este modelo no nos permite salir del ámbito de la sintaxis. La pregunta es como es que un elemento dentro de un sintagma adquiere su significado en relación a la exterioridad. La aproximación propiamente funcionalista parte de una cuestión de sentido común: lo que le otorga sentido semántico a las representaciones mentales es cómo funcionan y como se comportan. Asumiendo incluso la plausibilidad de este argumento, tan razonable (en realidad el argumento es de Fodor), no queda claro porque una base de datos, plena de relaciones sintácticas, va a tener ‘comprensión de lenguajes’ por el mero hecho de computarlas de acuerdo a un algoritmo que ni siquiera es capaz de procurarse a sí mismo. No deja de ser indicativo que todos los ejemplos de procesamiento simbólico que da Block se refieran a la sintaxis de argumentos matemáticos (sea en binario, en hexadecimal o finalmente en decimal que es el modo más familiar de *inputs* y *outputs* matemáticos para nosotros). Resulta absurdo negar que estos cómputos tienen un sentido matemático, o al menos que son capaces de expresarlo. Lo que nunca queda claro es cual es el tránsito de lo matemático a lo semántico fuera de lo matemático (de hecho Block sólo es capaz de explicarlo mediante mapeos de notación binaria a decimal, y viceversa). Si el lenguaje humano fuera un conjunto de relaciones matemáticas habría que admitir que el modelo funcionalista —el ‘modelo computacional de la mente’— es capaz de dar cuenta de un procesamiento de argumentos semánticos efectivo. Sabemos, por sentido común, que esto es posible en una calculadora y que lo matemático es esencialmente computable. Pero no sabemos

¹⁸ *Ibíd.* Pág. 268.

cómo dichos esquemas pueden explicar la comprensión de los lenguajes naturales. Además ello no anula la réplica de Searle: que a pesar de eso la máquina lógica carece de intencionalidad (y que la intencionalidad es una condición necesaria para la comprensión de lenguaje). Quizás la intencionalidad comprende un ‘mapeo’ que Block no describe con suficiente exactitud. Además existe un problema adicional que éste no considera (o que considera demasiado poco): resulta difícil reducir la dimensión semántica de los lenguajes naturales a todos los argumentos matemáticos conocidos. Si es que la semántica emerge de la computación, no está claro como es eso posible en una Máquina de Turing sin que intervenga un programador que le asigne variables numéricas a cada palabra. Es cierto que hay ‘verdades’ matemáticas que se mantienen inmutables a pesar de poder ser presentadas por muchos argumentos y en muchos sistemas posibles, pero esas ‘verdades’ no agotan ningún modelo de comprensión de lenguaje natural¹⁹. Hasta el momento nadie ha sido capaz de proveernos de dicho modelo. Al menos nadie ha logrado que una máquina ‘aprenda’ a asignar por si misma un número a una variable imprevista sin salir del ámbito de la estadística. ¿Es la estadística lo que explica la comprensión de lenguaje en los seres humanos? Creo que el experimento de la Máquina de Coser demuestra al menos una cosa: que en el origen de toda Inteligencia Artificial no hay un programa sino un programador (y hasta el momento nadie ha inventado un programa que programe programadores y que no sea un modelo ejemplar muy rústico de sistemas autoorganizados). Es muy difícil decir que la Máquina de Coser tiene una comprensión semántica de lo que es ‘Michelle’. Podrá manipular esa cadena de cálculos que están ahí en lugar de ‘Michelle’ (por que nosotros lo decretamos arbitrariamente). Podrá tener un registro discreto (pero monstruoso) de todas las cadenas donde está contenida esa cadena que llamamos ‘Michelle’ y hasta una categoría para cada registro, pero eso no es suficiente para hacer de ‘Michelle’ un contenido intencional, al menos como lo concibe Searle. En este sentido Searle admite dos posibilidades: o bien se le niega intencionalidad a la máquina programada (en la medida en que la intencionalidad simbólica no es provista por ella); o bien se le puede atribuir intencionalidad al ‘procesamiento de información’ que hace el computador (lo cual, por cierto, implicaría atribuírsela también a un termostato, como querría McCarthy²⁰, y así asumir que muchos subsistemas no cognitivos adquirirán

¹⁹ Ibid. Pág. 270.

²⁰ Véase al respecto: McCarthy, John, *Ascribing Mental Qualities to Machines*, en **Philosophical Perspectives in Artificial Intelligence**, 1979, Ringle, Martin (ed.), Humanities Press, Stanford, CA.,

competencias cognitivas).

Quisiera concluir este trabajo con una cita de Jorge Luis Borges: “Hume notó para siempre que los argumentos de Berkeley no admiten la menor réplica y no causan la menor convicción”²¹. Quizás radique ahí lo propiamente alucinógeno del ensayo de Block. En lo que se refiere a la semántica, sus especulaciones rozan apenas la cáscara del sentido (bajo el supuesto de que el sentido es apenas una cáscara).

Demian Schopf, agosto de 2008.

USA.

²¹ Borges, Jorge Luis, *Tlön, Uqbar, Orbis Tertius*, en *Ficciones*, 1956, Emecé, Buenos Aires, Argentina.

Bibliografía:

Block, Ned, *The Computer Model of the Mind*, En Brook, Andrew y Stainton, Robert. **Knowledge and Mind. A Philosophical Introduction.**, 2000, MIT Press, Cambridge, Mass, USA.

Searle, John R., *Minds, Brains, and Programs*, *The Behavioral and Brain Sciences*, vol. 3. 1980, Cambridge University Press, USA.

González, Rodrigo, *The Chinese Room: are programmed computers minds?* en *The Chinese Room Revisited: Artificial Intelligence and the Nature of Mind*, 2007, Katholieke Universiteit Leuven, Institute of Philosophy, Centre for Logical and Analytic Philosophy, <https://repository.libis.kuleuven.be/dspace/handle/1979/939>, Leuven, Bélgica.