La mente no funciona así. Alcance y límites de la psicología computacional

JERRY FODOR

Siglo XXI, Madrid, 168 págs.

Trad. de José Luis Gil Aristu

Donde la teoría computacional de la mente falla

Ángel Alonso-Cortés

1 abril, 2004

Los programas científicos, como los programas y partidos políticos, son obra de personas con intereses y designios, con líderes que pugnan por llevarse la dirección del programa (o del partido), los becarios, las ayudas financieras de las agencias públicas, y todo lo demás. El resultado probable de esta situación es que los programas ganadores deriven en una escolástica estéril y, al final, se desplomen como un guiñol.

Jerry Fodor es un teórico de la ciencia cognitiva y seguidor del programa chomskyano sobre la organización modular y computacional de la mente desde finales de los años sesenta. Fodor ha escrito un libro magro, pero robusto y valiente. Después de examinar concienzudamente los argumentos de otros miembros de la escuela, como Steven Pinker ¹, que ha encolado el neodarwinismo con el programa computacional de Chomsky (la Nueva Síntesis), concluye su libro así: «De momento, lo que nuestra ciencia cognitiva ha descubierto sobre la mente es, ante todo, que no sabemos cómo funciona» (pág. 135).

¿Cuáles son las razones que ofrece Fodor por las que el programa cognitivo-computacional, que él llama modularidad masiva (MM), no es un modelo de la estructura de la mente humana? La idea de que un modelo computacional es un modelo real de la mente humana nació con las ideas del matemático británico Alan Turing (1915-1952). A él se refiere Fodor continuamente, destacando que el tipo de computación que ha servido para modelizar los procesos cognitivos es la computación Turing que caracteriza la «máquina de Turing». En 1937 este matemático ² estableció los fundamentos de una máquina ideal que permitía computar números y funciones matemáticas. Una computación Turing es un conjunto de operaciones que combinan los símbolos (o componentes) atómicos paso a paso, formando una línea de símbolos L_n , a partir de una línea anterior L_{n-1} . La línea L_{n+1} formada a partir de L_n no tiene relación con L_{n-1} . Esto quiere decir que las líneas o computaciones son operaciones locales, pues no tienen en cuenta líneas que no sean inmediatamente adyacentes. Más tarde, en 1950, Turing ³ diseñó unas pruebas (conocidas como el Test de Turing) para demostrar cómo esta máquina (lógica), que sólo manipula símbolos, aunque no los significados, era indistinguible en su actuar del razonamiento humano. Nació así la inteligencia artificial.

Pocos años después, entre 1955 y 1960, Noam Chomsky (1928-) desarrolló las gramáticas formales, para configurar la estructura gramatical de las lenguas naturales se demostró que los lenguajes formales generados por estas gramáticas (inconsensibles, consensibles, etc.) eran -débilmente- equivalentes a distintos autómatas y éstos que podrían ser generados por máquinas de Turing. Chomsky, ya desde 1959, exigía que las gramáticas formales que describen la estructura del lenguaje natural fuesen, además de métodos de formalización, psicológicamente reales y, en último lugar, que tuviesen realidad biológica. Cuando a esto se le añade que la capacidad del lenguaje es completamente autónoma de cualquier otra y que constituye un módulo del edificio cognitivo, se obtiene la primera columna de la psicología cognitiva computacional.

En los últimos diez años esta idea de autonomía o modularidad cognitiva del lenguaje humano ha sido ampliada a otras capacidades cognitivas: el razonamiento, la percepción (visual y musical), la memoria y la toma de decisiones. Se obtiene así la modularidad masiva, donde la mente resulta en un conjunto de módulos «sin ventanas». Además, algunos biólogos y psicólogos (Pinker, Cosmides y Tooby) afirman que esos módulos son una adaptación biológica, resultado de la selección natural darwiniana. Se obtiene así la segunda columna en que descansa el programa cognitivista de que se ocupa Fodor. La conjunción de la modularidad masiva con el adaptacionismo es lo que Fodor llama la Nueva Síntesis.

El principal contraargumento que esgrime contra el modularismo masivo es que el razonamiento

humano, un módulo cognitivo esencial, no realiza operaciones locales o computaciones tipo Turing. Por el contrario, el razonamiento humano, esgrime el autor, es también holístico (o abductivo, como también lo llama). Esto quiere decir que es un razonamiento que tiene en cuenta el conjunto total de creencias del que razona y, en consecuencia, no se ajustaría a operaciones de computación local, paso a paso, del tipo Turing. La inferencia holística, sostiene Fodor, no resulta de la combinación sintáctica de átomos o componentes, sino de un sistema de creencias del razonador.

A esto se pueden añadir otros dos tipos de razonamiento que no satisfacen la computación local del tipo Turing: la inferencia bayesiana (o probabilista) y el razonamiento borroso, que el autor no menciona. La inferencia bayesiana no es estrictamente holística, sino una función del grado de creencia subjetiva en una proposición, mientras que el razonamiento borroso se basa en que los conceptos que emplea el razonador son vagos; por ejemplo, «bajo» en la proposición «Fodor es bajo». Y, desde luego, se pueden aportar argumentos contra la modularidad masiva desde el módulo del lenguaje, que Fodor mantiene intocado como paradigma de la modularidad de la mente. Así, un efecto holístico se consigue en el reciente programa minimalista de Chomsky 4, que selecciona una derivación computacional de entre el conjunto global de computaciones. Además, el fenecido programa de la semántica generativa (activo entre 1965 y 1975) descubrió que algunas derivaciones computacionales de oraciones gramaticales necesitaban información transderivacional, es decir, información no local.

Pero, además del holismo, en el módulo computacional del lenguaje intervienen otras limitaciones que hacen de la computación del lenguaje un extraño objeto. En la computación Turing, las reglas se aplican mecánica y ciegamente. En el lenguaje, las reglas están sometidas a restricciones del tipo «No muevas de la posición en que se encuentra un componente C (una frase) si hay otra frase entre la posición que se va a mover y el lugar al que se mueve C».

Otra limitación de la computación gramatical transgrede palmariamente el modelo clásico de computación local, que Fodor asume para el módulo lingüístico. En la computación Turing no se obtiene más información de la que contiene el léxico de que se parte. Las operaciones computacionales son, en términos matemáticos, monótonas. El lenguaje humano transgrede frecuentemente esta propiedad. Es el caso, frecuente en la computación de la forma sonora de las frases, en que crea información nueva (sílabas, acentos, entonación) no contenida en el léxico de partida.

La componencialidad del lenguaje, axioma de los programas computacionales, afirma que la forma y el significado de una frase están determinados por la combinación de rasgos primitivos (fonológicos, sintácticos y semánticos) que se componen o articulan como un mecano. Este axioma deja a la intemperie numerosas formaciones gramaticales o, de lo contrario, requiere una complejidad en la gramática que se aviene mal con la computación Turing. Pensemos en frases como (1) «Dime la hora», (2) «Sé ir a su casa». La frase (1) es una pregunta («Dime qué hora es»). La mera combinación de «la hora» con «di (me)» no resulta en una composición semántica de la que salga la pregunta indirecta. De forma parecida, el significado de (2) no se obtiene de la composición de «saber» con «ir». El significado de la frase (2) es «saber el camino a, saber por dónde hay que ir», significado que contiene más que la suma de sus partes. Es decir, de nuevo el holismo, omnipresente en el lenguaje y en el universo.

En el cuento «La escritura del dios», contenido en *El Aleph*, Borges acusa la presencia del holismo en el universo como una dificultad insuperable. El narrador se encuentra en una cárcel, esperando la muerte. En la celda vecina había un jaguar, cuyas manchas contienen un mensaje, que el narrador es incapaz de descifrar porque las cosas, el universo, forman una trama total:

«¿Qué tipo de sentencia (me pregunté) construirá una mente absoluta? Consideré que aun en los lenguajes humanos no hay proposición que no implique el universo entero; decir *el tigre* es decir los tigres que lo engendraron, los ciervos y tortugas que devoró, el pasto de que se alimentaron los ciervos, la tierra de que fue madre el pasto, el cielo que dio luz a la tierra. Consideré que en el lenguaje de un dios toda palabra enunciaría esa infinita concatenación de los hechos, y no de un modo implícito sino explícito».

Finalmente, aparece en el módulo del lenguaje la teleología o finalismo. Esto quiere decir que, en el modelo computacional de la gramática, las reglas o instrucciones que llevan desde los componentes iniciales al resultado final (una sílaba, una frase) parecen estar no ya ordenadas, sino guiadas por una forma ideal que impone el lenguaje tanto a los sonidos como a los conceptos y a la combinación de los componentes. Así, la forma sonora de las palabras prefiere sílabas con la disposición consonante-vocal (CV), y prefiere palabras, como sucede en español, que tengan un ritmo trocaico (CV-CV, con la primera sílaba CV acentuada), y prefiere frases que tengan el objeto contiguo al verbo del que dependen, y otras muchas preferencias universales que no tienen que ver con la computación Turing.

Es posible que sí, como concluye Fodor, no sabemos cómo funcionan la mente y el lenguaje, la causa de esta ignorancia sea el modularismo computacional, incapaz de alcanzar el Aleph. Y, a pesar de la valiente crítica a que Fodor lo somete, se aferra al modularismo como Moisés a las tablas de la ley.

Por lo que respecta a la traducción, ésta mantiene el estilo informal del autor, norteamericano, que parece perseguir aliviar así al lector. Hay, sin embargo, dos términos del original inglés en los que el traductor ha marrado el blanco. El sustantivo *fitness*, de la teoría biológica darwiniana, es vertido sistemáticamente como «aptitud», a secas. Su traducción normativa, de acuerdo con el *Vocabulario científico y técnico*, de la Academia de Ciencias, es «eficacia biológica», o bien «eficacia reproductiva». El otro término es el adjetivo inglés *relevant*, que aparece traducido sistemáticamente como «relevante» y su derivado como «relevancia». *Relevant* significa «pertinente» (derivado, «pertinencia»). El adjetivo español «relevante» significa, según el Diccionario de la Real Academia, «importante, sobresaliente». Dada la importancia y el uso abundante de estos dos términos en el texto, la lectura se hace en estos casos incomprensible.

¹. De cuyo libro *El instinto del lenguaje* (Madrid, Alianza, 1996) apareció una crítica nuestra en *Revista de Libros*, núm. 0, diciembre de 1996, pág. 45.

². «On computable numbers, with an application to the Entscheidungsproblem», *Proceedings of the London Mathematical Society*, serie 2, vol. 42 (1936-1937), págs. 230-265.

³. «Computing machinery and intelligence», *Mind*, vol. 59, núm. 236 (1950), págs. 433-460. ⁴ Del que nos ocupamos en *Revista de Libros*, núm. 80, julio-agosto de 2003, págs. 31-36.