

El pensamiento natural y las limitantes formales

Guillermo Morales-Luna
Sección de Computación
CINVESTAV-IPN
gmorales@cs.cinvestav.mx

25 de julio de 2002

Resumen

Abordamos, de manera más bien *naïve*, algunos aspectos de la Teoría de la Computación y la Filosofía. Distinguimos inicialmente las nociones de lenguaje, de modelo o idea, y de objeto. En la relación entre lenguajes y modelos están las nociones de consistencia, de completitud y de decisión de la noción de *inferencia lógica*. Los teoremas de Gödel imponen limitantes para que una teoría lógica sea completa o sea formalmente consistente. En el plano de la Computación estas limitantes se traducen a nociones de no-computabilidad. En el análisis de la inteligencia humana los teoremas de Gödel también imponen limitantes. Ufanos como somos, y seguros como estamos, no podemos aceptar vernos limitados como lo están las máquinas. Así pues resulta interesante ver cómo influyen los teoremas gödelianos en la inteligencia humana.

1 Modelos y objetos

El *pensar* es la actividad más importante de nuestro cuerpo y esto nos hace precisamente conscientes de nuestra existencia: *cogito, ergo sum*. Al pensar aprehendemos y aprendemos “la realidad”. De acuerdo con ciertos patrones de pensamiento, nos formamos modelos mentales de la realidad y después confrontamos “la validez” de esos modelos utilizando diversos procedimientos. El método científico puede dar cuenta de algunos de ellos. La semántica conceptual ciertamente está ligada con la organización funcional del pensamiento.

La diferencia entre “modelo” y “objeto” está muy bien ilustrada en el célebre cuadro de René Magritte, *Ce n'est pas un pipe*. En efecto, a lo que se refiere la frase en francés es una imagen de una pipa, y ésta no sirve para fumar. Una pipa, en cambio, bien que puede ser utilizada para fumar.

Wittgenstein decía que el significado de una palabra no es un objeto físico, ni siquiera una imagen mental. En 1928, George de Chirico, surrealista, decía en cambio que la palabra “estrella” está muy cerca de la idea de una estrella.

Estas reflexiones sobre lenguaje, idea y objeto se llevan naturalmente al ámbito de la computación. Un algoritmo, por ejemplo el cálculo de un punto fijo de una función polinomial o el del monto de los impuestos a ser devueltos a un causante en un cierto año fiscal o el del monto de los impuestos de una cierta transacción comercial, son motivados por “la realidad”, codificados en un lenguaje matemático y programados luego en un lenguaje de programación. Aquí “programa”, “procedimiento” y “noción calculada” son, en efecto, tres nociones distintas.

Desde un inicio, entonces, podemos distinguir a los tres componentes principales en una teoría del conocimiento: la *sintaxis* es el lenguaje de representación de las nociones; la semántica es la interpretación de las nociones, es la “idea” de los objetos descrita con los elementos de representación; la *pragmática* se refiere al ente que aprende, el cual observa los objetos y se forma modelos. La realidad, no es tan solo una población en Chiapas, es el universo de los objetos. La sintaxis y la semántica son las herramientas para construir modelos.

Es bien sabido que una teoría formal que sea consistente (es decir, libre de contradicciones) y que describa sus propios procedimientos de inferencia es, por un lado incompleta (habrá proposiciones que se cumplan en cualquier modelo de la teoría pero que no son demostrables en la teoría) y, por otro lado, no podrá demostrar su propia consistencia. Estos son los famosos teoremas de Gödel.

Cualquier noción razonable de “computabilidad” resultará por esto incompleta. ¿Qué hay sobre el proceso de pensamiento natural?

2 Consciencia y entendimiento

Evidentemente la función de pensar la realiza el cerebro. Desde mediados del siglo XX se han intensificado los esfuerzos por simular artificialmente varias de las tareas cerebrales.

La *Inteligencia Artificial Fuerte* (AIF) busca construir una máquina que piense. La *Inteligencia Artificial Débil* (AID) busca, en cambio, crear máquinas que hagan pensar a sus usuarios que piensan.

En enfoque de la AIF para analizar el proceso de pensamiento plantea que el cerebro es una computadora física con ciertos códigos de representación y con propios esquemas de procesamiento. En un segundo enfoque se distingue entre cerebro y mente. Acaso puesto de manera más bien burda, el cerebro es el *hardware* y la mente es el *software* del proceso de pensamiento. ¿Es la mente la versión moderna de la antigua noción de *alma*? No nos adelantemos.

Parece necesario un proceso de cómputo en la actividad del cerebro. Sin embargo, la noción de cómputo puede parecer no suficiente para describir todas las actividades mentales: La *intencionalidad* (planteamiento de objetivos, creencias, intenciones) y el *libre albedrío* plantean actitudes no-deterministas, muy poco formalizables. Ya no mencionemos los sentimientos o las sensaciones.

Algunos atributos de la mente son la *consciencia* y el *entendimiento*. Sin entrar en detalles técnicos, quedémonos con sus definiciones de diccionario o de sentido común. Si acaso, para precisar estos conceptos nos podría mos referir a los textos de Chalmers [1] y Tye [8]. “Entender”, por ejemplo, tiene al menos dos acepciones. Siempre se puede escribir un programa de computadora que calcule las dos raíces complejas de un polinomio de segundo grado con coeficientes reales. La computadora que corre el programa parece entender el procedimiento. Por otro lado, un alumno en una clase de álgebra elemental se encuentra con la fórmula para el cálculo de raíces y además de comprender el cálculo involucrado, se ha de esforzar para adquirir la noción de que los puntos resultantes son las abscisas de los puntos de cruce de la gráfica del polinomio dado con el eje de las x 's. Como dice Penrose en [5], si alguien considera irrelevante distinguir estas dos nociones de entender, entonces también lo ha de ser distinguir la noción de consciencia. Pero, aún con lo difícil de ser definidas, para muchas personas las nociones de consciencia y de entendimiento tienen sentido. Una pregunta natural es si acaso los programas de computadora pueden adquirir consciencia y entendimiento. Consideremos una máquina provista de un programa que juega ajedrez. Por un lado, su memoria de juegos previos o hipotéticos y los criterios de evaluación de situaciones actuales pueden hacer las veces de entendimiento, en tanto que su percepción de la partida, de las tiradas del oponente y su objetivo de ganar pueden hacer las veces de su consciencia. ¿Porqué no pueden ser considerados tales entendimiento y consciencia como reales? ¿Podría reconocer tal máquina jugadora de ajedrez que en una cierta partida va a perder? Aún para la mente humana éste puede ser un problema difícil. Efectivamente ciertos jugadores expertos pueden identificar situaciones de mate con varias jugadas de antelación. Plantearle el mismo problema a una máquina sería pedirle que calcule todas las posibles jugadas, a partir de la actual, y revise que ninguna de ellas es exitosa para ella. Esto equivaldría a probar que el objetivo de triunfar en la partida es inalcanzable desde la jugada actual. Visto este problema en un sistema formal equivale a demostrar la no-derivación en ese sistema de una cierta proposición. El célebre “segundo teorema de incompletitud” de Gödel afirma que los sistemas formales, como los programas de computadora, no pueden demostrar su propia consistencia y en consecuencia habrá proposiciones que no pueden demostrar como inalcanzables.

Penrose llega a afirmar que la inteligencia humana, es decir, su entendimiento y su consciencia, es superior a la de cualquier máquina por ese hecho: los humanos entendemos la incompletitud de los sistemas formales y las máquinas no la pueden realizar. Sin embargo, John McCarthy [3] señala que acaso la mente humana realiza la incompletitud porque razona sobre teorías, pero las teorías son meros objetos cuando se razona sobre ellas. Una máquina que razone sobre programas podría realizar la “no-programabilidad” de algunos problemas y aún así tendría un sistema formal incompleto. Puesto de una manera equivalente: si el sistema de razonamiento humano fuese formalizable entonces habría nociones verdaderas pero no-deducibles para la mente humana.

3 Consecuencias inmediatas del teorema de Gödel

Así pues, en una posición antropocéntrica, deberemos suponer que el razonamiento humano no es formalizable. En otras palabras, que supera los límites de lo computable. En un primer vistazo esto no parece descabellado. Diríamos que las creaciones del intelecto son fruto o motivo del placer estético. El arte, por ejemplo, manifiesta el talento de sus creadores. Sin embargo, una máquina que vaya colocando símbolos del alfabeto castellano uno tras otro, obteniendo siempre cadenas finitas de caracteres, al cabo de un número finito de pruebas habrá escrito *El ingenioso hidalgo Don Quijote de la Mancha*. O bien, variando píxeles de colores en un arreglo matricial, al cabo de un número finito de pasos podría obtener algo parecido a *Los fusilamientos del 3 de mayo* de Francisco de Goya, al *Cristo crucificado* de Velázquez o a cualquier otra pintura de El Prado. A pesar de lo pedestre de los argumentos anteriores, ellos hacen parecer “casi” computables a las creaciones artísticas.

Penrose demuestra por contradicción la no-formalización del razonamiento humano. Supóngase que hubiese un sistema formal F para el razonamiento de una persona, digamos *Fulano*. Entonces, necesariamente: toda proposición que *Fulano* considere verdadera ha de ser demostrable en F ; F ha de ser consistente y F ha de describir su propio mecanismo de inferencia. Al construir un enunciado autoreferente al estilo de Gödel, “La fórmula que tiene el índice que yo tengo es indemostrable en F ” (una observación pertinente aquí es que hay una correspondencia biunívoca entre fórmulas e índices, y otra es que el pronombre “yo” en el enunciado hace referencia a ese mismo enunciado) será verdadera y por tanto indemostrable. F entonces no puede existir.

Pero, como apunta McCullough [4], lo que el argumento anterior demuestra es que si el pensamiento fuese computable, entonces o bien es contradictorio o bien no puede formalizar sus mecanismos de razonamiento. Ciertamente, en cualquier sociedad humana se vive en contradicciones. En el mes de julio de 2002 hemos visto que un par de jóvenes palestinos hizo explotar bombas en sus cuerpos, en Tel Aviv, con lo que mataron de paso a dos ciudadanos israelíes. Menos de dos semanas después, el ejército israelí atacó “cuarteles” palestinos, mató a un cabecilla y a once personas más e hirió a otros 95 civiles. La primera acción se vió por segmentos importantes de la prensa mundial como un acto de terrorismo en tanto que la segunda como una “acción militar legítima”. El uso de las palabras pretende disminuir la contradicción resultante de las apreciaciones de los hechos. En lo individual, sobra decirlo, se puede vivir con muchas contradicciones lógicas. McCullough ilustra la inconsistencia del pensamiento con un experimento sencillo: A un individuo se le pide que apriete uno de dos botones marcados con “Sí” y “No” en función de la respuesta que considere correcta para cada pregunta que se le formule. Cuando se le pregunte “Apretará usted ahora el botón No” entonces el individuo caerá en contradicción.

En cuanto a los mecanismos de razonamiento se tiene que, a diferencia de los sistemas formales donde el conjunto de verdades es estático a manera de ideas platónicas, el conjunto de verdades aceptadas por un individuo es variable, e incluso puede ser no-monótono, y también está sesgado por creencias sociales. Ha habido diversos intentos a lo largo de la historia respecto a la formalización de las reglas del pensamiento. Sin embargo, dada la influencia notable de las relaciones sociales en el proceso de pensamiento, tenemos que a la fecha no hay una formalización universal. Tan no se tiene en el ámbito jurídico, por ejemplo, que es común dilucidar la corrección de un razonamiento por mera mayoría de votos de algunos ministros depositarios de la confianza social. Naturalmente, es necesario reconocer que algunos de los criterios para zanjar diferencias tienen que ver con el mantenimiento de la consistencia de los sistemas de deducción utilizados.

El teorema de Gödel establece que o bien no se puede formalizar el pensamiento o, si se hiciese, entonces no podría probar su consistencia. Este es un aspecto común de la inteligencia artificial y de la inteligencia natural. No es una limitante de ellas, es un rasgo inherente de todo sistema de razonamiento que puede razonar sobre sí mismo. Gödel en algún momento escribió que la combinación hilbertiana de materialismo con algunos aspectos de las matemáticas clásicas es imposible (demostrablemente). Las limitantes lógicas nos ponen pues al nivel de las máquinas. Ahora bien, cada humano tiene (o debería tener), en su interacción con el medio, la posibilidad de utilizar un lenguaje “natural”, una autonomía, un libre albedrío, una capacidad de resolución de problemas y una capacidad de razonamiento abstracto. En cambio, las computadoras no son creativas, pues no dan origen a nada, ellas sólo siguen las instrucciones dictadas por programas. Esta es tan sólo una diferencia trivial y determinante entre la inteligencia natural y la artificial.

Hasta aquí he mencionado que considero a las relaciones sociales como una dificultad para la formalización del razonamiento natural. Acaso su propia inconsistencia sea la clave del “no-sometimiento” del

pensamiento humano al teorema de Gödel. Penrose afirma, en cambio, que ese no-sometimiento se debe a que el pensamiento no es computable. Siendo todo cerebro un dispositivo físico (orgánico, biológico y vivo) la mecánica newtoniana determinista haría implausible la no-computabilidad. Así que la explicación física de ella se remonta a propiedades de la física del cerebro analizada en el marco de la mecánica cuántica (sobre todo porque en ella se ha visto ciertos patrones de cuasicristales que evidencian patrones no-computables de organización). En [2] y [6] se puede ver críticas y réplicas a este enfoque. Es acaso relevante mencionar que esta explicación ha sido acogida con mucho entusiasmo por asociaciones religiosas o místicas y por esto mismo se ha expresado reservas hacia ella desde su misma aparición (véase, por ejemplo, [7]). Parece que el antropocentrismo sigue siendo de gran influencia aún en los tiempos actuales.

4 Conclusiones

El razonamiento humano o bien es no-computable, o bien es contradictorio o bien sus reglas no pueden ser abarcadas en un sistema formal. El probar cada una de estas propiedades ha dado origen a intensos trabajos de investigación. Desde un punto de vista materialista elemental se puede suponer que el cerebro, como dispositivo físico que es, no puede ser superior a una máquina de Turing. Sin embargo, físicos de gran relevancia en la Física Moderna, como Roger Penrose, han planteado como hipótesis la no-computabilidad del pensamiento debido a la no-computabilidad de ciertos patrones vistos en la Física de partículas sub-atómicas.

Me parece, sin embargo, que hay que considerar la complejidad de las relaciones sociales y de la percepción que colectivamente nos hacemos del mundo para “probar” que, en efecto, nuestro pensamiento es contradictorio y no codificable en un sistema formal (que incluya sus propias reglas de deducción).

En este escrito sólo incluyo primeras impresiones personales y no les atribuyo validez “científica” alguna.

Referencias

- [1] D. Chalmers, *The conscious mind*, Oxford University Press, 1996 .
- [2] R. Grush, P.S. Churchland. Gaps in Penrose’s toilings. *Journal of Consciousness Studies* 2(1):10-29, (1995). Reprinted in *Conscious Experience*. Ed. by T. Metzinger. Berlin: Schöningh-Verlag. Reprinted in Paul Churchland and Patricia Churchland (1998) *On the Contrary: Critical Papers 1987-1997*. Cambridge, MA: MIT Press.
- [3] J. McCarthy. Awareness and understanding in computer programs: A review of “Shadows of the Mind” by Roger Penrose, *PSYCHE*, 2(11), July 1995. <http://psyche.cs.monash.edu.au/v2/psyche-2-11-mccarthy.html>
- [4] D. McCullough. Can humans escape Gödel?: A review of “Shadows of the Mind” by Roger Penrose, *PSYCHE*, 2(4), April 1995. <http://psyche.cs.monash.edu.au/v2/psyche-2-11-mccullough.html>
- [5] R. Penrose. *The shadows of the mind: A search for the missing science of consciousness*, Vintage, 1994.
- [6] R. Penrose, S. Hameroff. What gaps? Reply to Grush and Churchland. *Journal of Consciousness Studies* 2(2):98-111, (1995).
- [7] V. J. Stenger, The myth of quantum consciousness, *The Humanist*, 53 (3), pp. 13-15, May/June 1992. <http://www.phys.hawaii.edu/vjs/www/qmyth.txt>
- [8] M. Tye, *Ten problems of consciousness*, Bradford Books/MIT Press, Cambridge, MA, 1995.