

## **El autómata con discernimiento de Torres Quevedo: Un antecedente del modelo de Turing**

Manuel Garrido

El nombre del ingeniero Leonardo Torres Quevedo (1852-1936), uno de los más grandes genios españoles del pasado siglo XX, no es tan conocido como debiera serlo actualmente entre sus compatriotas. Si no son muchos los que saben que alcanzó reputación internacional en la construcción de dirigibles y transbordadores (de los que es famoso ejemplo el aún existente *Spanish Aerocar* en las cataratas del Niágara), son menos los que tienen noticia de que treinta años antes del nacimiento de la moderna teoría de la computabilidad y de los primeros ordenadores y cuarenta antes de que se formulase el programa de investigación de lo que hoy llamamos “inteligencia artificial”, Torres Quevedo publicó un profético ensayo sobre autómatas, a continuación reproducido en las páginas de este número de *Teorema*, que puede parangonarse sin ningún desdoro con el ensayo fundacional sobre la metáfora del ordenador publicado por Alan Turing en 1950, cuatro años antes de su muerte.

Turing fue un investigador solitario que se movió en los ambientes científicamente privilegiados de Cambridge y Princeton y que vivió el nacimiento y el desarrollo del ordenador digital. También Leonardo Torres Quevedo fue un investigador solitario, pero su medio ambiente era el páramo español, aunque el nivel económico de su familia le permitió contactar seriamente con la ciencia francesa. Él había nacido en 1852 y cuando publicó sus *Ensayos sobre Automática* (1914) había cumplido ya los sesenta. Sus puntos de referencia no pudieron ser como en el caso de Turing, nacido en 1912, ni el esplendor de la lógica matemática de los años treinta del pasado siglo con el teorema de Gödel o la tesis de Church ni la tecnología creadora de los años cuarenta protagonizada por Wiener o von Neumann. Su punto de referencia tuvo que ser forzosamente Babbage, otro gran solitario de un siglo anterior, el XIX, que gastó su vida y su fortuna en la construcción de un “ingenio analítico” o máquina diferencial que hoy es pieza de museo.

Torres Quevedo realizó entre otras cosas la proeza tecnológica de sustituir los métodos mecánicos de Babbage por el nuevo instrumental que le proporcionaba la electrotecnia y así pudo alardear que donde Babbage había fracasado el triunfó. Pero donde mejor se puede medir el vuelo de su mente

es en las especulaciones futuristas de esos *Ensayos*, que representan como ha escrito su biógrafo, el profesor Santesmases, la más ambiciosa y audaz de sus obras. Torres Quevedo comparte con Turing la rara conjunción de la creatividad tecnológica con el interés filosófico y lo superó en la realización efectiva de soluciones parciales a la mecanización del juego de ajedrez. El análisis de su “autómata con discernimiento” nos permite establecer un parangón nada desventajoso con el genio creador de Turing en su propuesta de la metáfora del ordenador.

#### *El modelo de Torres Quevedo*

Una de las principales críticas de que ha sido objeto el modelo propuesto por Turing en su ensayo “¿Puede pensar una máquina?”, críticas que son extensivas al programa de investigación de la inteligencia artificial en la medida en que se inspire en semejante modelo, es que este último carece de la posibilidad de establecer un contacto real con el mundo exterior. Esta crítica invita a reflexionar sobre el hecho de que todo el potencial operativo de una máquina universal de Turing se limita al recuento y reordenamiento de símbolos o piezas de cálculo sin suministrar ni poder suministrar información real sobre el mundo. Es, por así decirlo, una máquina exclusivamente *sintáctica*, sin capacidad de aportar dato o contenido semántico alguno. Padece respecto del mundo de un solitario aislamiento, de un *solipsismo*, como se dice en lenguaje filosófico, que se asemeja mucho a la falta de conexión con la realidad física que sufre el sujeto cognoscente concebido por Descartes como pura “cosa pensante” o *res cogitans*.

Esta limitación está en consonancia con la intención del proyecto de Turing, cuya apuesta aceptaba el desafío de simular la actividad de la mente humana por procedimientos no biológicos. El resultado, deliberadamente pretendido, resulta ser así la producción de un simulacro de mente “desencarnada” o separada del cuerpo biológico. La conciencia de esta limitación dio lugar en los años sesenta, segunda década de la investigación en inteligencia artificial, a la proliferación de programas que pretendían suministrar “contenido semántico” al ordenador introduciéndole léxicos y diccionarios que pudiesen complementar con un simulacro de “significado” sus operaciones de deducción y de cálculo, que son puramente sintácticas. Y en las décadas subsiguientes la obsesión de los investigadores llegó más allá en el intento de simular artificialmente la percepción de la realidad mediante programas de “reconocimiento de formas” perceptuales que doten al ordenador de un simulacro de conocimiento sensorial. Minsky ha diseñado programas que simulan el conocimiento global o giestáltico. Y John Mc Carthy ha llegado a desarrollar una especie de lógica del sentido común o “lógica de la circunscripción” que pudiera permitir al ordenador desenvolverse con éxito en contextos donde juega un papel esencial el cambio de circunstancias. En progresiva conjugación con la robótica, los

programas de inteligencia artificial han tenido que pugnar por liberarse del solipsismo y el inmovilismo inherente al modelo de Turing.

En el modelo de *autómata con discernimiento* propuesto por Torres Quevedo en sus *Ensayos sobre Automática* no hay sin embargo, en contraposición al modelo de Turing, ni la más remota sombra de solipsismo ni de cartesianismo. La necesidad de que el autómata se comunique con el mundo exterior y opere sobre él, es decir, el nexo de la inteligencia artificial con la robótica es algo que queda manifiestamente consignado por el ingeniero español en este pasaje de sus *Ensayos*, en donde prescribe que el autómata disponga no sólo de “sentidos”, que le suministren noticia sensorial del mundo, sino también de “miembros y “energía” que posibiliten sus movimientos y acciones exteriores:

Los autómatas deberán tener *sentidos*: termómetros, brújulas, dinamómetros, manómetros, [...] aparatos sensibles a las circunstancias que deben influir en su marcha.

La impresión recibida por cada uno de estos aparatos se traduce, generalmente, por un movimiento, por ejemplo, el desplazamiento de una aguja sobre un limbo graduado.

Los autómatas deberán tener *miembros*. Las máquinas o los aparatos capaces de ejecutar las operaciones que les sean encomendadas. La *orden* de ejecutar una operación será transmitida al aparato encargado de realizarla por procedimientos muy sencillos, aunque se trate de operaciones complicadas; esto se ve, por ejemplo, en algunos relojes, en los cuales una pieza que se dispara permite que se ponga en marcha un mecanismo, el cual actúa sobre muñecos que ejecutan diversos movimientos.

Los autómatas deberán tener la *energía* suficiente: los acumuladores, las corrientes de agua, los depósitos de aire comprimido que han de suministrársela a las máquinas destinadas a ejecutar las operaciones necesarias.

Pero más importancia tiene todavía a este respecto el hecho de que Torres Quevedo no concibe a su “autómata con discernimiento” como una mera máquina abstracta de cálculo lógico y matemático. Anticipando la necesidad de la lógica de la circunscripción propuesta por Mc Carthy, el autómata con discernimiento debe ser capaz de un conocimiento de las circunstancias de su entorno que le permita desarrollar una conducta de adaptación al mundo:

Además, se necesita —y éste es el principal objeto de la Automática— que los autómatas tengan *discernimiento*, que puedan en cada momento, *teniendo en cuenta las impresiones que reciben, y también, a veces, las que han recibido anteriormente*, ordenar la operación deseada. *Es necesario que los autómatas imiten a los seres vivos, ejecutando sus actos con arreglo a las impresiones que reciben y adaptando su conducta a las circunstancias.*

Generalizando, podría decirse que la diferencia esencial entre el modelo de Turing y el modelo de Torres Quevedo está en que el primero se orienta a simular automáticamente el *pensamiento* humano y el segundo la *acción* humana como evidencia este pasaje inicial de los *Ensayos*:

La denominación de autómatas se aplica a menudo a una máquina que imita la apariencia y los movimientos de un hombre o de un animal. Se trata, generalmente, en el caso de una máquina que lleva en sí misma la fuente de energía que la hace marchar (un resorte, por ejemplo) y ejecuta ciertos actos, siempre los mismos, sin recibir ninguna influencia exterior.

Hay otra clase de autómatas que ofrecen un interés mucho más considerable: los que imitan, no los gestos, sino las acciones del hombre, y algunas veces pueden reemplazarle. El torpedo automóvil, que *sabe* maniobrar para llegar al término de su carrera; la balanza que pesa las piezas de moneda para *elegir* las que tienen peso legal, y mil aparatos más, muy conocidos, pueden servir como ejemplo de autómatas de esta última especie.

Se hallan otros mucho más interesantes en las fábricas. El progreso industrial se realiza principalmente sustituyendo al trabajo del hombre por el trabajo de la máquina; poco a poco llegan a hacerse mecánicamente la mayor parte de operaciones primitivamente ejecutadas por obreros, y se dice que una fabricación ha llegado a ser automática cuando puede ser completamente ejecutada por medio de máquinas.

La idea de una máquina, o una clase de máquinas, con capacidad de decisión para alterar el curso de las cosas indujo a Torres Quevedo a proponer una ampliación de la teoría tradicional de máquinas con una nueva rama, a la que denominó *Automática*, señalándole como objetivo la tarea de examinar “los procedimientos que pueden aplicarse a la construcción de autómatas dotados de una *vida de relación* más o menos complicada”.

#### *La anticipación del argumento del “cuarto chino”*

Recientemente, el filósofo norteamericano John Searle ha ideado una brillante crítica del llamado “test de Turing” (la apuesta de que en el curso de unas décadas un ordenador sería capaz de competir airoosamente durante varios minutos en el conocido “juego del mentiroso” con cualquier ser humano). Para Searle este test adolece del defecto de las teorías conductistas, que confunden lo que parece ser con lo que es. Y para evidenciar este aserto diseñó su ya famoso “argumento del cuarto chino”. Imaginemos una cabina cerrada y comunicada con el exterior con sendos buzones de entrada y salida, por el primero de los cuales admite preguntas en chino y por el segundo emite respuestas en ese idioma. Mirándola desde fuera, podría decirse, y un conductista podría admitirlo, que esa cabina posee conocimiento del idioma chino. Pero si mirándola por dentro descubriésemos que aloja en su interior un individuo que dispone de un copioso repertorio de preguntas y respuestas

en chino numeradas correlativamente y que se limita a cotejar la pregunta entrante con las de su repertorio hasta identificarla con una de ellas y a emitir acto seguido la respuesta numerada correspondiente, resultaría al punto manifiesto que la cabina en cuestión no tiene nada que ver con una persona que sepa realmente chino.

Ahora bien, volviendo a la alternativa de Torres Quevedo, impresiona advertir que el argumento del cuarto chino está previsto e inteligentemente valorado en sus *Ensayos sobre Automática*. Me refiero a un pasaje en el que discute la opinión de Descartes (nunca mencionada por Turing) de que un autómata jamás sería capaz de mantener un diálogo razonable. Después de aludir a dicha opinión de Descartes recordando que éste “admite fácilmente que el autómata pueda hablar, pero no concibe que pueda hablar razonablemente”, y para rebatirla, Torres Quevedo diseña mentalmente un escenario mecánico que es reproducción exacta de la cabina searleana:

Imaginemos una máquina [...] en la cual, en vez de tres conmutadores, haya miles o millones, si hace falta; y que, en vez de tres o cuatro posiciones diferentes, cada conmutador tenga una posición correspondiente a cada uno de los signos de la escritura (letras, cifras, signos de ortografía, etc.).

Se comprende perfectamente que se puede, valiéndose de estos conmutadores, escribir una frase cualquiera, y hasta un discurso más o menos largo; esto dependerá del número de conmutadores de que se disponga.

A cada discurso corresponderá una posición del sistema, y, por consiguiente, un electroimán.

Podemos suponer que éste dispare un fonógrafo sobre el cual se halle inscrita la respuesta a la pregunta que ha provocado su movimiento, y de este modo tendremos una autómata capaz de discutir *de omni re scibile*.

Ciertamente, el estudio preliminar de todas las preguntas posibles, la redacción de la respuesta a cada una de ellas, y, finalmente, la construcción de una máquina semejante, no sería una cosa muy llana; pero no sería mucho más difícil que la construcción de un mono, o de otro animal bastante bien imitado para que pudiera ser clasificado por los naturalistas entre las especies vivas.

Finalmente, y para rematar matizándola su refutación de Descartes, el ingeniero español recoge la profunda idea de Ada Lovelace, hija de Lord Byron y consejera espiritual de Babbage, de que lo que la máquina hace es lo que su constructor, el hombre, ha dispuesto que haga:

No hay entre los dos casos la diferencia que veía Descartes. Pensó sin duda que el autómata, para responder razonablemente, tendría necesidad de hacer él mismo un razonamiento, mientras que, en este caso, como en todos los otros, sería su constructor quien pensara por él de antemano.

Creo haber mostrado, con todo lo que precede, que se puede concebir fácilmente para un autómata la posibilidad teórica de determinar su acción en

un momento dado, pesando todas las circunstancias que debe tomar en consideración para realizar el trabajo que se le ha encomendado.

Se puede concebir igualmente a un autómata que *obre con una finalidad*, un autómata que realice una serie de actos *tendiendo* a conseguir un objeto determinado.

Volviendo la vista atrás, vale la pena recordar las apuestas tecnológicas de Turing y Torres Quevedo, respectivamente formuladas por uno y otro en los dos párrafos que transcribo a continuación:

Yo creo que en unos cincuenta años será posible programar ordenadores [...] para que jueguen tan bien el juego de la imitación que un interrogador medio no tenga más del 70 % de probabilidades de acertar en la identificación de su interlocutor después de cinco minutos de diálogo [...]. Y creo que para finales del actual siglo XX el uso de las palabras y la opinión general de la gente educada habrá cambiado tanto que se podrá hablar de máquinas pensantes sin miedo a que los demás nos contradigan [Turing, en su ensayo de 1950].

El Sr. Torres sostiene que [...] al menos en teoría casi todas las operaciones de una vasta gama de ellas podrían ser realizadas por una máquina, incluso aquellas de las que se supone que precisan la intervención de una considerable capacidad intelectual [...] [de una entrevista a Torres Quevedo realizada por la revista *Scientific American* en 1915].

Descartes, Gödel y Penrose han aportado argumentos teóricos de peso que ponen serias trabas de principio a las pretensiones de ambas apuestas. Pero con todas las reservas mentales que esos argumentos nos aconsejan tomar, uno no puede dejar de constatar el hecho o los hechos de que ya hace décadas que el trivial programa “Eliza” engañó a más de un enterado en psicoanálisis, de que en nuestros días el programa “Pensamiento profundo” ha puesto en jaque a un campeón de ajedrez de nivel no ya comarcal sino mundial, y de que hace veinte años que los ordenadores digitales despertaron de su inicial solipsismo cartesiano para comunicarse unos con otros y con nosotros mismos permitiéndonos, a través de Internet y la Tela de Araña Universal, un ejercicio alucinante del método del hipertexto que nuestros padres estaban lejos de soñar. Dificilmente se desprende de estos hechos que, en la práctica, las apuestas de Turing y Torres Quevedo hayan dejado de cumplirse.

#### NOTA BIBLIOGRÁFICA

El escrito *Ensayos sobre automática* de Leonardo Torres Quevedo fue originalmente publicado en la *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, XII, 391-419. En el mismo año apareció también como folleto en

la Imprenta del Renacimiento de Madrid, y en versión francesa al siguiente en la *Revue Générale des Sciences*. Ha sido asimismo incluido en el volumen de escritos de Torres Quevedo editado por Francisco González de Posada en la Biblioteca de la Ciencia Española, Fundación del Banco Exterior, Madrid, 1992.

El ensayo de Alan Turing de 1950 ha sido traducido al castellano con el título *¿Puede pensar una máquina?*, en la Colección Cuadernos Teorema, Valencia 1974.