

¿Cerca de 2001?

Turing, Clark, Kubrick,
Orwell y las tecnologías
del lenguaje

Octavio Augusto Sánchez



LA LITERATURA Y EL CINE DE CIENCIA FICCIÓN han prometido, durante años, un mundo donde la mayor parte de nuestras actividades cotidianas se vuelven tan sencillas como sólo pedirle a una máquina que limpie la casa, nos prepare un café, o nos ayude a tomar decisiones para que todo sea hecho al instante. En esas historias la máquina suele ser una computadora de enormes dimensiones, pero invisible. Un ejemplo referencial clásico de estos enormes equipos es, sin duda, la HAL 9000.

Esta máquina, protagonista de *2001: Odisea del espacio*, muestra todos los deseos que tenemos algunos seres humanos sobre la tecnología. Controla todo lo que ocurre dentro de la nave, la temperatura, el movimiento, la navegación. Pero, además, es capaz de comunicarse con los humanos de abordo de distintas maneras: mediante sus *tablets* o con la interfaz de habla. Los usuarios le hablan y HAL entiende lo que le dicen, y les responde. Incluso, esta supercomputadora es capaz de apreciar, estéticamente, los dibujos realizados por un ser humano y de producir una opinión al respecto. Quizá uno de los aspectos más impresionantes de esta máquina es la capacidad de reconocer sentimientos generados en su contra como consecuencia de sus actos.



Fotograma de 2001: Odisea del espacio, 1968

Ver películas como esas, leer artículos fantásticos sobre cómo podría ser el futuro a partir de equipos que puedan leer e interpretar nuestras ondas cerebrales, y la creciente disponibilidad de artilugios tecnológicos, nos hacen preguntarnos: ¿nos hallamos cerca de 2001?

Para contestar esta pregunta quizás sería conveniente revisar otra novela que, si bien no es de ciencia ficción, hablaba del futuro cuando fue escrita. Me refiero a *1984* de George Orwell. En esa obra, quizás el dispositivo que se puede recordar con mayor claridad es el *Telescreen*. Ese sistema capaz de saber lo que las personas hacían y veían en todo momento. Sin embargo, no es el único dispositivo con carácter futurista que se presenta en la obra. Uno de los que recuerdo con más frescura es el *Speakwrite*.

Speakwrite era un sistema al que se le dictaba y las palabras se transcribían sobre el papel requerido para los cambios. Pues bien, estos sistemas sí existen actualmente, desarrollados principalmente para el inglés, aunque ya existen versiones en español. Son los

llamados sistemas de reconocimiento de habla. Estos sistemas reconocen palabras dictadas por un usuario y las transcriben, casi como por magia, en el monitor. Uno se da cuenta de que en realidad no es magia lo que hace funcionar a estos aparatos porque, aunque se ha avanzado bastante en esta área, siguen produciendo muchos errores.

Estos sistemas de reconocimiento de habla en realidad son la base para muchos otros programas que tienen como intención el control de los equipos informáticos mediante la voz. Actualmente es posible hasta cierto punto controlar algunas computadoras y teléfonos celulares. Sin embargo, este proceso requiere un entrenamiento de la máquina que la mayoría de los usuarios no estamos dispuestos a realizar. En cuanto empezamos a hablarle a la computadora nos sentimos raros, y nos sentimos peor cuando la maquina no entiende lo que le decimos. Pero, a pesar de ello, es justo reconocer que en algunas ocasiones resultan bastante cómodos. Por ejemplo, si queremos buscar algo

en Internet, podemos dictarle la palabra o frase que buscamos a nuestro teléfono y lo buscará por nosotros.

Esto me lleva a mencionar otra tecnología que fue originalmente presentada como ciencia ficción: un buscador de información. Unos personajes discuten acerca de una cita, que resultó ser de Shakespeare. Uno de ellos menciona que si estuviera dentro de su nave podría obtener el autor y el libro de esa cita en cuestión de minutos. En la realidad, los sistemas de recuperación existentes, como Google, pueden encontrar una cita exacta en menos de un segundo. Es decir, los sistemas actuales superan a aquel que se había predicho en la ficción. En 1984 aparece la aterradorra descripción de un departamento encargado de hacer una de las cosas más atroces. Este departamento producía, para los proles de Oceanía, periódicos, libros de texto, películas, programas de *Telescreen*, novelas completamente artificiales, incluso existía la *pornsec*, la sección encargada de producir pornografía. Lo



característico de estas “producciones” literarias es que se realizaban completamente por métodos mecánicos, es decir, sin la participación humana. En la actualidad no es posible hacer este tipo de literatura sin la intervención humana, sin embargo, existen diferentes áreas de investigación en este aspecto, prototipos en diferentes universidades de máquinas que escriben cuentos, discursos políticos, artículos científicos e, incluso, algunas que “componen” piezas musicales. De momento no conozco a nadie que esté trabajando en la producción automática de pornografía, pero sin duda parece una oportunidad interesante para alguna mente creativa. Los logros de estos dispositivos escritores son aún muy pequeños, pero ya se pueden producir frases con sintaxis correcta.

Además, otras tecnologías que no recuerdo haber leído en 1984, pero que se relacionan y desarrollan actualmente, son los programas de computadora capaces de reconocer opiniones y sentimientos hacia alguna entidad específica. Estos son capaces de leer lo que se publica en Internet y entregar al usuario un indicio del sentimiento general que tiene un determinado grupo hacia él. Los usuarios pueden ser marcas específicas o cualquier persona interesada en saber lo que se dice sobre ellos.

Otra de las que no aparece en la novela es la traducción automática. En Oceanía no se necesitaba tal, todos hablaban inglés, aunque se imponía el *Newspeak* sobre el *Oldspeak*. Pero estos sistemas son de los que, posiblemente, sean de mayor necesidad en la actualidad. Cada vez podemos encontrar más información en diferentes lenguas. Con la invención de la web y el consecutivo incremento en el uso de internet, las lenguas entran en contacto entre sí con mayor frecuencia que antes, lo que pone de manifiesto una mayor necesidad de sistemas capaces de traducir en tiempo real de una lengua a otra sin errores.

Y ¿qué tienen todas estas tecnologías en común? Todas usan como datos de entrada y de salida las len-



Fotograma de 1984 de Michael Radford

guas humanas, es decir, el lenguaje natural. Por tanto, se ha dado en llamarlas con el creativo nombre de “Tecnologías del lenguaje”.

Y ¿quién desarrolla esta tecnología? ¿Se desarrolla sólo en Estados Unidos? ¿Existe en idiomas diferentes al inglés? Todas estas preguntas tienen sus respectivas respuestas. Las Tecnologías del lenguaje se desarrollan principalmente en las universidades y centros de investigación de todo el mundo. Gran parte se ha realizado para el inglés —pues nació en ese país, con propósitos bélicos.

Los primeros dispositivos se realizaron con el fin de traducir los mensajes rusos interceptados cuando comenzó la Guerra Fría. En ese entonces se decidió aprovechar la experiencia adquirida en la decodificación de mensajes durante la Segunda Guerra Mundial. Los primeros intentos asumieron que el ruso era una codificación particular del inglés, por lo que el mensaje sólo debía decodificarse para lograr su traducción. Siguiendo este procedimiento se estrellaron con los problemas inherentes a cualquier lengua natural: ambigüedad, polisemia, homonimia, anáforas, anacolutos, etcétera. Por lo que dedujeron que esos métodos no eran suficientes y decidieron contratar lingüistas. A partir de entonces, la lingüística, las matemáticas, la filosofía y la computación se encuentran muy unidas en el campo de la inteligencia artificial, en un subcampo conocido como procesamiento del lenguaje natural.

Como en muchos lugares del mundo, en México se desarrolla activamente esta área. Hoy existen diversos grupos en las universidades mexicanas que investigan sobre tecnologías del lenguaje. Entre estas instituciones educativas se encuentran la UAM, la BUAP, el IPN, el INAOE y la UNAM. En esta última se encuentra el Grupo de Ingeniería Lingüística del Instituto de Ingeniería. Grupo que desde 1999 conjunta los esfuerzos de lingüistas con ingenieros en computación y con profesionales de otras áreas del conocimiento con el fin de desarrollar nuevas tecnologías del lenguaje.

Entre otros, se desarrollan sistemas de detección de similitud de textos, que pueden servir para identificar plagio académico; extracción de información de trabajos académicos con el fin de identificar a diferentes investigadores que trabajen un mismo problema abordado desde diferentes puntos de vista; extracción de conocimiento a partir de textos comunes donde se obtienen definiciones y términos con el fin de entender de qué manera se presentan nuevos conceptos.

Como podemos ver, aunque la meta es construir una computadora que sea mucho más inteligente que HAL 9000 y sea capaz de pasar la prueba de Turing —un juez humano engañado por una máquina que finge ser humana—, las diversas tecnologías del lenguaje que tenemos ahora se encuentran lejos de ese rendimiento. Por tanto, para responder a la pregunta original, en 2012, ¿qué tan cerca nos hallamos de 2001?, yo diría que con dificultad estamos cerca de 1984. ▀