



Tópicos del seminario

ISSN: 1665-1200

ISSN: 2594-0619

Benemérita Universidad Autónoma de Puebla, Seminario
de Estudios de la Significación

Rico Sulayes, Antonio

Hacia el reconocimiento de la subjetividad en los agentes
artificiales. Una delimitación del sujeto artificial en la semiótica

Tópicos del seminario, núm. 41, 2019, Enero-Junio, pp. 147-169

Benemérita Universidad Autónoma de Puebla, Seminario de Estudios de la Significación

Disponible en: <http://www.redalyc.org/articulo.oa?id=59460164007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Hacia el reconocimiento de la subjetividad en los agentes artificiales. Una delimitación del sujeto artificial en la semiótica

Antonio Rico Sulayes

Universidad de las Américas Puebla

Introducción

A propósito del peculiar caso de una orangutana, la cual ganó el fallo de un *hábeas corpus** en las cortes argentinas en 2014, Dorra (2016) se da a la tarea de interpretar el término *sujeto no humano*, utilizado para designar a la beneficiaria de esta decisión judicial (Cámara Federal de Casación Penal, 2014). De acuerdo con este autor, la relación entre los conceptos *sujeto* y *humano*, en el término en cuestión, no es de una simple negación, como implicaría la lectura estereotípica del *ítem* de polaridad negativa *no*. De manera alternativa, según el investigador, el significado del giro fraseológico podría interpretarse a partir de una lectura del *no* como un relacionante de carácter concesivo, que resultaría en la interpretación “sujeto, aunque no humano”. Esta lectura le conferiría la condición de sujeto a la orangutana y a los miembros de la clase animal, o al menos a algunos de ellos.

* Un *hábeas corpus* es un recurso jurídico cuyo objetivo es remediar de manera inmediata la detención o el arresto ilegal de un sujeto. Este recurso busca garantizar la libertad personal de los ciudadanos.

Más allá de la subjetividad lingüística (González de Requena, 2017) que se ocupa del sujeto involucrado en el lenguaje, a esta condición de sujeto conferida a algunos animales, en el caso legal mencionado le llamaré *subjetividad*. Esto lo haré para evitar llamarle *sujetidad* (término quizá más propio en el español), pero que podría resultar confuso por su asociación con el concepto de *sujetidad política*, relacionado con la caracterización del sujeto humano al rebasar su animalidad (Gasca, 2007). Conferirle la subjetividad o condición de sujetos a algunos miembros del reino animal resulta, más allá del complejo debate biológico sobre qué es un animal —señalado por Dorra (2016)—, en repercusiones legales inmediatas.

El efecto inmediato de postular la subjetividad animal en el ámbito legal se hace patente en el reciente caso de un oso en cautiverio en Barranquilla, Colombia. Apenas unas semanas antes de la redacción de este artículo, se le ha dado también el fallo de *habeas corpus* a dicho animal, por considerársele “sujeto sintiente no humano” (Corte Suprema de Justicia, 2017). De manera interesante, la definición de la Corte colombiana añade el concepto de *sintiente* al término de sujeto no humano y hace énfasis así en uno de los elementos que justamente llevan a Dorra a colegir la razón por la que hoy en día algunos animales comienzan a ser vistos como sujetos. Para este investigador, los humanos estamos dispuestos a conferirle la condición de sujeto a animales que cumplen con al menos tres características: 1) que expresen dolor y afecciones, 2) que parezcan interpretar un lenguaje y parezcan intentar comunicarse con nosotros, y 3) que nos dirijan su mirada. Así, la característica de sintiente recientemente añadida por la Corte colombiana es un elemento central, de acuerdo con Dorra (2016), de eso que nos lleva a considerar al menos a algunos animales como “superiores” a otros animales y poseedores de una subjetividad o condición de sujetos.

La discusión sobre la subjetividad del animal como sujeto no humano me trae inevitablemente a la discusión de otro tipo

de sujeto al que he llamado en otro trabajo investigativo *sujeto artificial* (Rico Sulayes, 2016). Como he argüido en aquel estudio, los sujetos artificiales tienen múltiples facultades que podrían hacerlos sujetos, es decir, que les permitirían también ostentar la subjetividad recién conferida a los animales en los casos legales citados. Para ejemplificar a qué me refiero por *sujeto artificial*, habría que señalar que este tipo de sujeto está representado de manera estereotípica por los agentes de diálogo para la recuperación de información, tales como Siri, Cortana y *Google Now*. Aunque estos agentes de diálogo son populares entre los usuarios avezados de la telefonía celular, existen ejemplos de este tipo de sujetos con comportamientos mucho más complejos y que más adelante discutiré. Así, aunada a su capacidad de escuchar, entender y responder a conversaciones con sujetos humanos para la resolución de problemas, algunos sujetos artificiales van más allá y tienen la facultad de observar y discriminar objetos (incluso objetos en movimiento), así como de identificar sujetos humanos y analizar las acciones de éstos para guiar sus propias acciones. Como ya lo he dicho en otro momento, la compleja actuación de los agentes artificiales los hace, sin duda, sujetos de conocimiento, de aprendizaje y de enunciación (como diría de sus estudiantes, “éstos sí, sujetos humanos” (Oberti, 2003). Esta primera descripción acercaría a los agentes artificiales a una condición de sujetos semióticos cabales; sin embargo, para darles dicho estatus a los agentes artificiales, haría falta teorizar algunos aspectos más sobre estos sujetos semióticos. Entre esos aspectos, se tendría que discutir de manera precisa cuáles son las particularidades de la subjetividad en este tipo de sujetos no humanos. Por ello, la caracterización de la subjetividad de los sujetos artificiales será el objetivo primordial de este trabajo.

1. Un programa sobre la subjetividad en los agentes artificiales

El concepto de *agente artificial*, como un agente que realiza tareas común o previamente realizadas por humanos y que las hace en interacción con los mismos, se encuentra bien establecido en la literatura científica (Coeckelbergh, 2015; Laukyte, 2017; Sellers, 2013). Frente al concepto de *agente artificial*, el de *sujeto artificial* requiere una caracterización más precisa. Yo he llamado artificial a este tipo de sujeto semiótico, no sólo porque su definición agencial coincide con la de los agentes artificiales, sino también por su inevitable asociación con la inteligencia artificial, disciplina que ha postulado la creación de agentes artificiales desde sus inicios mismos. En esa postulación, el elemento definitorio central de estos agentes o sujetos ha sido su inteligencia. Por tanto, una delimitación de la subjetividad en el sujeto artificial debería partir de este elemento también.

Dentro de la inteligencia artificial, los agentes artificiales han sido postulados de acuerdo con al menos dos objetivos a alcanzar: ser tan inteligentes como para parecer humanos (Turing, 1950) y precisamente por no humanos, ser incluso más inteligentes que aquellos en algunas tareas (Harari, 2016). En años recientes ambos objetivos han sido alcanzados bajo condiciones controladas y las implicaciones de esto para la definición de sujeto artificial ocuparán las siguientes dos secciones. En la primera de ellas mostraré cómo la postulación primera de los sujetos artificiales no fue de una inteligencia plena, sino de una inteligencia al menos en apariencia. En la sección posterior, discutiré cómo simultáneamente se ha perseguido un programa mucho más ambicioso para llegar a un sujeto artificial con una inteligencia “superior”, en este caso a la de los sujetos humanos. Derivado de los logros alcanzados por los sujetos artificiales actualmente, en una tercera sección volveré sobre la definición actualizada de los animales como *sujetos sintientes no humanos*, con el fin de buscar una fórmula paralela que funcione para el

sujeto artificial y que delimite qué tipo de sujeto no humano es éste. Finalmente, cerraré con un comentario, más a forma de discusión que de conclusión, sobre si las definiciones aparentemente dispares de los dos sujetos no humanos antes mencionados (agentes artificiales y animales) tienen un punto de encuentro en algunos de los avances recientes de la robótica de entretenimiento. Para cerrar esta sección, vale la pena destacar que el estudio del sujeto artificial, como he dicho antes (Rico Sulayes, 2016), debería ser una tarea central de una auténtica semiótica computacional.

2. Los agentes de la inteligencia artificial en apariencia inteligentes

El científico inglés Alan Turing, considerado como uno de los fundadores de la ciencia de la computación (Jurafsky & Martin, 2017), es también reconocido por algunos como el inventor de la primera computadora moderna (Copeland & Proudfoot, 2011). Aunque este último reconocimiento suele ser cuestionado, debido a la dificultad de separar a la computadora actual de sus predecesores, Turing es sin duda uno de los primeros teóricos tanto de la inteligencia artificial como de los agentes artificiales. Pocos años después de generar la propuesta que sería la base de su primera computadora (1945), Turing escribiría para la reconocida revista de filosofía *Mind* un artículo sobre el controversial tema de si las computadoras, de recién producción en aquel entonces, podrían llegar a ser inteligentes en un futuro (1950). Para elucubrar sobre la posible existencia de máquinas inteligentes, Turing parte de la premisa de que la pregunta “¿pueden las máquinas pensar?” es esencialmente errónea. Para este científico, contestar a esta pregunta basándonos en el significado estereotípico de los términos *máquina* y *pensar* no tiene sentido, especialmente si la aplicamos al ámbito de las computadoras. Algo obvio en los tiempos de Turing era que aquellos entes en los que se pensaba

de manera clásica cuando se hablaba de máquinas no hacían eso que se consideraba estereotípicamente como pensar. Por tanto, según este científico, cuestionarse sobre la posible inteligencia futura de las máquinas requeriría más bien hacer la pregunta aplicándole una prueba a éstas.

Con el objeto de diseñar una prueba de inteligencia para las computadoras, Turing (1950) plantea la interacción entre tres participantes: un sujeto humano, una máquina y un juez humano. En esta prueba, a la cual llama *juego de imitación*, el juez humano tiene la oportunidad de interactuar por separado con la máquina y el sujeto humano por un determinado lapso. Al final de este tiempo de prueba, el juez debe distinguir quién es el humano y quién la máquina. En el diseño de la prueba, la interacción debería estar mediada por una interfaz de texto que oculte la identidad de los dos sujetos interrogados, el humano y el no humano. Por supuesto, la máquina debe tener como objetivo aparecer como un sujeto humano. Turing parte del presupuesto de que este juego de imitación bien podría llevarse a cabo con otro tipo de sujetos, como por ejemplo un hombre y una mujer, donde el hombre intentara aparecer como la mujer. Entonces por extrapolación, la pregunta para juzgar la inteligencia de la máquina sería: ¿Se equivocaría el juez en la misma proporción al discriminar al sujeto humano de la máquina como cuando intentara distinguir al hombre de la mujer? Aunque lo correcto sería primero determinar la proporción de errores con que el juez se llega a equivocar al distinguir al hombre de la mujer, Turing no hace esto en su artículo. En vez de eso, el científico cierra la presentación de su prueba con una predicción. Turing vaticina que cincuenta años después de la escritura de su artículo las máquinas deberían tener suficiente capacidad de almacenamiento y de procesamiento de información para hacer que el juez se equivoque al menos el 30% de las veces que emite un juicio. Esto después de interactuar con la máquina y el humano por un periodo de cinco minutos con cada uno. Aunque la prueba de inteligencia de Turing es fácilmente cuestionable en térmi-

nos de validez, lo interesante es que la delimitación del agente artificial, en los inicios mismos de la disciplina que propuso su existencia, parte de una definición de éste en términos de lo que puede parecer. Es decir, se trata de un “sujeto *aparentemente* inteligente, aunque no humano”.

La temprana postulación de un agente artificial en Turing (1950) provocó una larga serie de intentos por generar el agente que pudiera pasar la prueba ahí descrita. Esto resultó incluso en la organización de encuentros internacionales donde diversos agentes compiten intentando pasar la eventualmente llamada prueba de Turing (Jurafsky & Martin, 2008). En esta larga lista de intentos existen al menos dos agentes artificiales particularmente interesantes por los resultados logrados con ellos. El primer agente que vale la pena mencionar es el agente interactivo de Weizenbaum, ELIZA (1966). Aunque desde el punto de vista de su diseño el agente ELIZA está constituido por una serie de reglas bastante simples, la impresión que logró tener en los sujetos humanos en los tiempos de su construcción no tenía precedentes. A través de transformaciones poco complejas en las oraciones proporcionadas por sus interlocutores, ELIZA puede producir preguntas o aseveraciones que imitan la interacción de un psicoterapeuta rogeriano. Para la psicoterapia rogeriana, en boga en los tiempos de la construcción de ELIZA, se debe dejar al paciente experimentar su propia sensación psicológica, intentando que éste se deje llevar por su propia reflexión (Hernández, 1963). Así, al menos con un interlocutor participativo que no cuestione la subjetividad o condición de sujeto de ELIZA, es fácil sostener conversaciones relativamente largas, como la reportada por Weizenbaum (1966: 36-37):¹

Los hombres son todos iguales.
¿DE QUÉ MANERA?

¹ Las líneas escritas en versalitas representan las intervenciones de ELIZA, mientras que las otras son las intervenciones de una mujer joven [La traducción es mía].

Siempre nos están molestando de una manera u otra.
¿PUEDES PENSAR EN UN EJEMPLO EN ESPECÍFICO?
Bueno, mi novio me hizo venir aquí.
TU NOVIO TE HIZO VENIR
Dice que muchas veces estoy deprimida.
SIENTO MUCHO SABER QUE ESTÁS DEPRIMIDA
Es cierto. Soy infeliz.
¿CREES QUE VENIR ACÁ TE AYUDARÁ A NO SER INFELIZ?
Necesito un poco de ayuda, eso parece cierto.
¿QUÉ SIGNIFICARÍA PARA TI OBTENER UN POCO DE AYUDA?
Quizá podría aprender a llevarme bien con mi madre.
DIME MÁS SOBRE TU FAMILIA
Mi madre cuida de mí.
¿QUIÉN MÁS EN TU FAMILIA CUIDA DE TI?
Mi padre.
TU PADRE
Tú eres como mi padre en cierta manera.
¿QUÉ PARECIDO ENCUENTRAS?
No eres muy agresiva, pero creo que no quieres que me dé cuenta de eso.
¿QUÉ TE HACE PENSAR QUE NO SOY MUY AGRESIVA?
No discutes conmigo.
¿POR QUÉ CREES QUE NO DISCUTO CONTIGO?
Me tienes miedo.
¿TE AGRADA CREER QUE TENGO MIEDO DE TI?
Mi padre tiene miedo de todos.
¿QUÉ MÁS PASA POR TU CABEZA CUANDO PIENSAS EN TU PADRE?
Los bullies.
¿ESO TIENE QUE VER ALGO CON EL HECHO DE QUE TU NOVIO TE HIZO VENIR AQUÍ?

Según reporta Weizenbaum (1976), el mérito de ELIZA fue que verdaderamente logró convencer a algunos de sus interlocutores de que se trataba de un sujeto humano. Según este autor, algunas personas continuaban conversando y creando lazos emocionales con ELIZA incluso después de que se les decía que era tan sólo un programa de computadora. Sin embargo, como fácilmente puede corroborar el lector, después de interactuar un poco con

ELIZA en alguna de sus implementaciones en línea,² si uno no coopera en la conversación es fácil romper la ilusión de que ELIZA es un sujeto inteligente.

Más recientemente en junio de 2014, durante una competencia organizada por la Universidad de Reading en la Real Sociedad de Londres para el Avance de la Ciencia Natural, el agente interactivo Eugene Goostman logró pasar la prueba de Turing, de acuerdo con la descripción original de ésta de 1950 (Press Team at the University of Reading, 2014). El hecho produjo simultáneamente una oleada de noticias que por un lado reconocían el logro, incluso en espacios prestigiosos como el diario *The Guardian* (Press Association, 2014) y la revista *Time* (Aamoth, 2014), y que por otro lado cuestionaban la validez del mismo (Aaronson, 2014; Lopatto, 2014; Sample & Hern, 2014). Las razones para reconocer el logro fueron simples. La competencia había respetado los criterios descritos por Turing (1950) para la prueba. El agente interactivo Eugene Goostman hizo que diez de treinta jueces que conversaron con él y un humano, lo identificaran a él como el sujeto humano después de interactuar individualmente con ambos por un lapso de cinco minutos. Es decir, el 33% por ciento de las veces los jueces fallaron identificando al agente artificial como el sujeto humano (Press Team at the University of Reading, 2014). También, cada uno de los treinta jueces que participaron en la competencia intervino en diez conversaciones (cinco con sujetos humanos y cinco con agentes artificiales), y por tanto, los cinco agentes artificiales evaluados interactuaron con treinta jueces distintos. En resumen, todos los criterios planteados por Turing en su prueba fueron cumplidos en la competencia de la Universidad de Reading. Además, la prueba se aplicó simultáneamente a varios agentes artificiales en una competencia de mayor envergadura. Entonces, ¿cuál fue el conflicto que llevó a algunos a no reconocer el logro de Eugene?

² En inglés existen múltiples implementaciones de ELIZA; una de éstas se puede encontrar en línea: <https://www.cyberpsych.org/eliza/>. Una versión en español se halla en <http://deixilabs.com/eliza.html>

En primer lugar, algunas características no consideradas en la prueba originalmente descrita por Turing fueron explotadas convenientemente por los diseñadores de este agente artificial (Aaronson, 2014; Lopatto, 2014). Por ejemplo, Eugene Goostman está diseñado para personificar a un adolescente ucraniano de trece años de edad. Por tanto, los errores de generación de lenguaje que comete a propósito y los que comete por error se pueden enmascarar detrás de su supuesto uso no nativo del inglés. Derivado también de la edad que personifica, Eugene cambia constantemente de tema arguyendo que no sabe nada de algún tema propuesto por su interlocutor o que dicho tema le resulta aburrido. Esto le permite evadir una conversación más compleja o cambiar el tema convenientemente cuando algún comentario de su interlocutor contiene cierta dificultad. En segundo lugar, para algunos de los detractores de Eugene (Aaronson, 2014; Lopatto, 2014), Turing (1950) sí describió un agente artificial con un uso nativo de la lengua y con la complejidad de razonamiento de un adulto educado. Esta descripción está implícita, según argumentan, en un extracto de diálogo imaginado por este científico, con el que intenta mostrar la interacción de un agente artificial que podría pasar su prueba. En dicho extracto, el agente imaginado por Turing tiene una conversación con un interlocutor humano sobre el tema de cómo componer la línea de un soneto y exhibe conocimiento de la métrica del verso inglés, así como del significado y asociaciones metafóricas de las palabras. En su crítica al desempeño de Eugene, Lopatto (2014) decide utilizar los comentarios hechos por el interlocutor humano imaginado por Turing (1950) y se los alimenta a la interfaz en línea de este agente artificial.³ El resultado es que Eugene cambia constantemente de tema y las repuestas equivocadas que da sólo se justifican en el presupuesto de que es un adolescente no nativo del inglés.

³ Lamentablemente la interfaz en línea de Eugene fue quitada de internet después de que este agente ganara la prueba en la competencia organizada por la Universidad de Reading.

Finalmente, para los detractores de Eugene Goostman existe un último pero fundamental problema en la descripción misma de la prueba de Turing. Quienes se oponen a reconocer el triunfo de Eugene (Aaronson, 2014; Lopatto, 2014; Sample y Hern, 2014) coinciden en pensar que Turing únicamente expuso el diseño de su prueba en un intento por esbozar un instrumento que permitiera identificar un programa de cómputo con un alto poder de manejo de información y de razonamiento, y que justamente esas características no distinguen a Eugene. Si los detractores de Eugene están en lo cierto, tendríamos que revisitar la definición de *sujeto artificial* que anteriormente sugerimos como un “sujeto aparentemente inteligente, aunque no humano”. Siguiendo la reflexión de los opositores de Eugene, el agente artificial imaginado por Turing (1950) tendría algo más que una aparente inteligencia. De hecho, este agente tendría al menos algunos aspectos de la inteligencia, como la acumulación y la capacidad de procesamiento de una amplia gama de datos. Ahora bien, por encima de la discusión de si Turing en realidad quiso proponer o no un agente artificial con una inteligencia más refinada que la exhibida por Eugene, es claro que la agenda de la inteligencia artificial ha tenido el objetivo paralelo de crear agentes artificiales que no sólo equiparen, sino que superen a los humanos en algunas tareas. La siguiente sección se ocupa de presentar algunos agentes que han alcanzado esta meta teniendo como escenario, nuevamente, la competencia entre sujetos humanos y sujetos no humanos.

3. Los logros de la inteligencia superior en los agentes artificiales

Si recordamos la definición antes dada de agente artificial, como aquel que lleva a cabo tareas propias de los humanos en interacción con los mismos, se puede deducir que los agentes que lleguen a superar a los sujetos humanos deberían hacerlo bajo

dichas condiciones. En ese sentido, uno de los ámbitos en que se han alcanzado los logros más significativos de los agentes artificiales es el almacenamiento y la recuperación de grandes cantidades de datos, así como las inferencias que se pueden realizar a partir del análisis complejo de éstos —facultades que justamente exhibe el agente imaginado por Turing (1950) en su diálogo. Esta sección discutirá la aparición de un par de agentes artificiales, Deep Blue y Watson, que han superado a los humanos en estos aspectos y que lo han hecho justamente en pruebas de inteligencia en las que los sujetos humanos han estado compitiendo por mucho tiempo.

Ahora bien, la decisión de argüir sobre la inteligencia superior de los agentes artificiales presentando a los agentes Deep Blue y Watson no obedece únicamente a que éstos han ganado pruebas compitiendo en ciertas tareas con sujetos humanos. Más allá de dichas pruebas, estos agentes representan hitos comúnmente mencionados por quienes ensalzan los avances recientes de la inteligencia artificial (Harari, 2016; Jurafsky & Martin, 2017). Precisamente con esta agenda investigativa en su extenso trabajo científico de corte histórico, Harari (2016) sugiere que los agentes artificiales superarán en prácticamente todas las actividades laborales a los humanos en un futuro no muy lejano. A pesar de esta perspectiva potencialmente desoladora para muchos miembros de la sociedad, este autor destaca que los agentes artificiales no se están acercando a los humanos en todos los ámbitos. Para Harari, los agentes artificiales se han ido mejorando de manera inequívoca en su inteligencia, entendida por este autor como el almacenamiento y análisis complejo de grandes cantidades de datos. Sin embargo, según Harari, estos agentes se han quedado esencialmente estáticos, desde sus inicios en 1950, en la evolución de una conciencia y en la experiencia de emociones y sentimientos. Es decir, para este autor el agente artificial actualmente sería “un sujeto inteligente, aunque no humano, ni consciente, ni sintiente.” En este sentido, habría que señalar que detrás de la primera parte de esta definición, *sujeto inteligente*, existió el

objetivo implícito de llevar la inteligencia del sujeto artificial por encima de la humana. Por eso no resulta sorprendente que los hitos de esta superación hayan utilizado, de forma similar a como propuso Turing, pruebas entre sujetos humanos y no humanos.

El primer gran hito en la demostración de una inteligencia superior de los agentes artificiales, según Harari (2016), tuvo lugar en 1997 con la victoria de Deep Blue, un sofisticado jugador de ajedrez diseñado por IBM, sobre el entonces campeón mundial de este juego, Garri Kasparov. El hecho de que este importante logro de la inteligencia artificial se diera en una competencia pública entre un sujeto humano y uno artificial obedeció, de acuerdo con Harari, al prestigio que tenía este juego en la década de los ochenta. Dicho prestigio convertía al ajedrez, en el imaginario popular, en insignia de la superioridad humana en el ámbito del razonamiento.

Según describe IBM (2011b), el desarrollo del agente artificial Deep Blue implicó un largo proceso que comenzó con el trabajo de la tesis doctoral de Fenghsiung Hsu en 1985. Más tarde, uno de los compañeros de Hsu, Murray Campbell, se uniría al proyecto y los dos serían contratados en 1989 por IBM Research. En esta empresa, un equipo de investigadores trabajaría por varios años para finalmente lograr el triunfo de Deep Blue en una contienda que de hecho fue la revancha de un enfrentamiento el año anterior, donde el sujeto humano, Kasparov, le había ganado al agente artificial. Ambos eventos tuvieron un escenario tradicional, con una mesa y un tablero normales. El único elemento extra era un monitor donde Deep Blue mostraba sus jugadas y un sujeto humano las ponía en práctica en el tablero tradicional. La contienda final, compuesta por seis juegos individuales, fue televisada mundialmente y produjo una reacción sin precedentes para un juego de ajedrez.

Ahora bien, dada la descripción del evento, es claro que algunos elementos aún faltaban para pensar en un sujeto artificial demasiado complejo. Más allá de la interacción de Deep Blue asistida por el sujeto humano, quien ponía en práctica sus

jugadas en el tablero, el razonamiento de este agente artificial estaba claramente limitado al universo de posibles eventos del ajedrez. Si pensamos en la descripción que he ofrecido inicialmente para los sujetos artificiales, como sujetos de conocimiento, de aprendizaje y de enunciación, Deep Blue no sólo mostraba un conocimiento y aprendizaje de una aplicación limitada, sino además una enunciación también escasa. Desde el punto de vista lingüístico, Deep Blue exhibía una limitada capacidad de escuchar, entender y responder a conversaciones con sujetos humanos. Aunque estas facultades podrían parecerles triviales a los entusiastas del ajedrez, como he mostrado en la discusión sobre la prueba diseñada por Turing, detrás de la concepción de un agente artificial inteligente ha habido desde sus inicios el anhelo de una interacción lingüística compleja, y dicha interacción estaba aún ausente en Deep Blue.

Poco después, sin embargo, el diseño de un sujeto artificial con un desempeño lingüístico altamente complejo alcanzó un nuevo hito con la construcción de Watson, otro agente artificial también diseñado por un equipo de IBM. Para Harari (2016), este agente representa un hito en la evolución de la inteligencia artificial por las profundas repercusiones que podría tener en el mercado laboral este tipo de tecnología. Una vez más en una competencia entre sujetos humanos y no humanos, Watson, un agente artificial experto en la recuperación de información, ganó el juego de *Jeopardy!*, un concurso televisivo donde los participantes contienden respondiendo preguntas sobre temas varios como historia, literatura, ciencia, geografía y cultura general. El triunfo de Watson tuvo lugar mientras competía con los dos jugadores humanos que más veces habían ganado en la historia del juego (IBM, 2011a). Desde 2010, reporta el equipo desarrollador, Watson era capaz de ganar el 70% de los juegos en que competía contra ganadores previos del juego. Esto sucedía en prácticas de entrenamiento que se realizaban en las oficinas de IBM. No obstante, a principios de 2011 Watson compitió en la televisión contra los dos jugadores con más triunfos en la historia

del concurso, Ken Jennings y Brad Rutter, y les ganó. Según los desarrolladores de Watson, llegar a este logro presentó principalmente dos retos, los cuales resultan particularmente interesantes para nuestra discusión.

Como reconoce el equipo de IBM, uno de los mayores retos en la construcción de Watson fue lidiar con la ambigüedad de las preguntas del juego, las cuales suelen incluir múltiples juegos de palabras (2011a). Debido a esta característica del juego, contestar a muchas de las preguntas requiere no sólo la simple recuperación de conocimientos concretos, sino también una elevada sagacidad en el uso del lenguaje. Otro de los retos principales en la construcción de Watson, según sus creadores, era generar un sistema de evaluación interna que le permitiera a este agente producir un nivel de confianza para sus respuestas. Este sistema de evaluación era requerido para evitar perder demasiados puntos al dar respuestas poco plausibles. Es decir, además del avanzado manejo del lenguaje, Watson fue diseñado con un elaborado sistema de creencias para decidir sus acciones, un elemento que anteriormente he descrito como central en los sujetos artificiales (Rico Sulayes, 2016). Así, con su elaborado entendimiento del lenguaje y su capacidad de juzgar la pertinencia de sus intervenciones, Watson exhibe un desempeño lingüístico muy superior al de Eugene y más cercano al ideal que, según los detractores de este último agente (Aaronson, 2014; Lopatto, 2014; Sample & Hern, 2014), Turing esbozó en su diálogo imaginado con un agente artificial inteligente (1950). Como he mencionado, en este diálogo dichos detractores ven implícitas en el agente facultades como la acumulación y el procesamiento de una amplia gama de datos, y el razonamiento complejo a partir de éstos. Es decir, las características de la victoria de Watson nos llevan una vez más a la concepción inicial del agente artificial. De esta manera, el camino que se ha recorrido para llegar hasta los logros recientes de la inteligencia artificial nos debería permitir hacer la delimitación del sujeto artificial en su estado actual.

4. La delimitación del sujeto artificial a partir de las definiciones de los sujetos no humanos

Entre los adjetivos que expresan cualidades del sustantivo existen los explicativos, que expresan una cualidad intrínseca definitoria del nombre (como *suave* cuando califica a *susurro* o *verde* a *follaje*), y los calificativos, que expresan una cualidad extra y por tanto restrictiva (como *rojo* cuando se aplica a *automóvil*). Desde esta perspectiva, el adjetivo añadido por la Corte colombiana cuando califica al oso de Barranquilla como “sujeto *sintiente* no humano”⁴ debería ser visto como explicativo, pues *sintiente*, como dice Dorra (2016), es una de las cualidades que nos llevan a considerar a algunos animales como superiores y, por ende, como sujetos. Así, el término usado por esta Corte podría ser interpretado como “por sujeto sintiente, aunque no humano”, donde el sustantivo *sujeto* implica la capacidad de sentir. Siguiendo esta lógica, en esta sección ofreceré finalmente una fórmula equivalente que delimite a los sujetos artificiales en su estado actual de desarrollo.

Como he mostrado a lo largo de este trabajo, la definición de *sujeto artificial*, en términos de la fórmula que he ido esbozando para los animales, tiene de manera similar una trayectoria con varios estadios. En primer lugar, Turing (1950) parece ofrecer una conceptualización inicial de los agentes artificiales que correspondería a la fórmula “sujeto aparentemente inteligente, aunque no humano”. Sin embargo, los detractores del primer agente artificial que ganó la prueba descrita por este autor abogan por una conceptualización más compleja, algo más cercano a la fórmula “sujeto *parcialmente* inteligente, aunque no humano”. Como recién mencioné, esta inteligencia parcial debería incluir la capacidad de acumular y procesar una gran cantidad de datos y analizarlos de manera compleja. Así, dada la precisión del párrafo anterior sobre el adjetivo explicativo y calificativo, y el

⁴ El énfasis en *sintiente* es mío.

argumento de los detractores de Eugene que piensan que la inteligencia imaginada por Turing no lo era sólo en apariencia sino en esencia, podríamos ajustar un poco más la última fórmula. De esta manera, podríamos colegir que Turing imaginó un agente que era “por sujeto parcialmente inteligente, aunque no humano”. Más tarde, con la revisión de los logros de la inteligencia artificial, hemos visto que Harari (2016) ofrece una visión más elaborada de los agentes artificiales que podría resumirse en la fórmula “sujeto inteligente, aunque no humano, ni consciente, ni sintiente”. Esta fórmula la podríamos acotar aún más para pensar en un agente que es “por sujeto inteligente, aunque no humano, ni consciente, ni sintiente”. Es decir, con esta precisión la inteligencia sería esencial y definitoria del sujeto. Finalmente, pensando en la inteligencia superior de algunos agentes en tareas específicas, una última fórmula podría describir a agentes como Deep Blue y Watson (quizá también a Siri, Cortana y muchos más): “aunque no humano, por sujeto incluso más inteligente, aunque no consciente, ni sintiente”. Me parece que ésta es la fórmula que mejor delimita al sujeto artificial según los logros de los agentes artificiales en la actualidad. Ahora bien, la compleja fórmula derivada de toda la discusión anterior me lleva a una última pregunta: ¿en verdad la inteligencia artificial ha soslayado en lo absoluto el proyecto de un sujeto *sintiente* no humano? Con una reflexión sobre dicha pregunta, más que una respuesta a la misma, cerraré el presente trabajo.

5. Discusión: por un programa de los agentes artificiales sintientes

Derivado de la delimitación del sujeto artificial en la sección anterior, que niega su propia capacidad de sentir, he terminado dicho apartado preguntándome si en verdad la inteligencia artificial ha ignorado la posibilidad de desarrollar un sujeto artificial sintiente. Ciertamente los grandes logros en los agentes

artificiales que buscan una inteligencia superior a la humana están encabezados por agentes que no exhiben la facultad de experimentar emociones o sentimientos. Sin embargo, Sellers (2013) argumenta que si no se modelan emociones en los agentes artificiales éstos no podrán alcanzar algunos objetivos cruciales en su interacción con los sujetos humanos, por ejemplo, la credibilidad de sus reacciones o de su comportamiento. Siguiendo esta lógica, no es sorprendente que uno de los ámbitos donde más se ha trabajado en el desarrollo de agentes artificiales que exhiben emociones y sentimientos sea la robótica de entretenimiento. Enfocada en sus inicios al desarrollo de robots que pudieran coexistir con los sujetos humanos en tareas de entretenimiento (Velo, 2002), esta área de la robótica tenía como objetivo original la construcción de robots autónomos con tres habilidades principales: la percepción del ambiente, el entendimiento y razonamiento de los estímulos percibidos, y la realización de acciones como respuesta a dichos estímulos. En la actualidad, esta área ha recorrido un largo camino y ha evolucionado hacia el desarrollo de mascotas robots altamente complejas. Entre los múltiples robots interactivos que existen hoy, me gustaría mencionar de manera rápida dos que me parecen particularmente interesantes para una discusión sobre los agentes artificiales sintientes.

El primer robot que quisiera mencionar es Zoomer,⁵ cuyo personaje más común es un perro, aunque también los hay en otras versiones como gato y chimpancé. Zoomer es un robot con diversos sensores, desplazamiento por ruedas, producción de sonidos que emulan ladridos y reconocimiento facial y de voz en varias lenguas. Esta última característica es utilizada para identificar, después de un periodo de entrenamiento, hasta quince comandos que producen diversas reacciones en el robot. Zoomer también cuenta con un par de pantallas de *led* en los ojos que se

⁵ En la siguiente liga se puede encontrar una introducción a Zoomer en español, que incluye varios videos, entre otros recursos: <https://www.zoomerpup.com/international/index.php?lang=mx>

utilizan para indicar al usuario algunos elementos de la interacción, como cuando Zoomer está esperando una instrucción, o para intentar mostrar emociones ante algunas reacciones, como cuando se toca su pecho en señal de caricia. Entre los comandos a los que Zoomer responde, existen los clásicos trucos esperados en los perros mascota, como “dame la patita” y “siéntate”, pero también existen algunos comandos para restablecer el canal de comunicación, como “mírame”, y para intentar crear vínculos afectivos en el usuario, como cuando se le dice “te quiero” y el robot responde con una serie de arrumacos.

En segundo lugar, mencionaré al robot Cozmo. Éste no representa a una mascota sino más bien a un compañero de juego. Cozmo⁶ es un pequeño camión con tracción de oruga y una pala mecánica que puede participar en una serie de juegos con el usuario. Igualmente, Cozmo cuenta con una compleja gama de sensores, que no son visibles como los de Zoomer u otros robots de entretenimiento, y que le permiten desplazarse de manera compleja, así como manipular de manera muy fina una serie de objetos. Igualmente, Cozmo cuenta con reconocimiento facial y de voz, emisión de sonidos que emulan un lenguaje humanoide y una pantalla donde despliega una mirada con una amplia gama de expresiones que evocan sorpresa, enojo, concentración, y aburrimiento, entre otras varias emociones posibles.

Considerando la descripción recién hecha de Zoomer y Cozmo, es claro que estos dos robots han rebasado el plan inicial de la robótica de entretenimiento. Estos robots pretenden algo más que simplemente percibir el ambiente, entender los estímulos y reaccionar a ellos mientras coexisten con los sujetos humanos en una tarea de entretenimiento. Es obvio que los dos robots descritos tienen como uno de sus objetivos primordiales emular sentimientos y evocarlos en sus usuarios. Además, esos robots persiguen esta tarea utilizando las tres características que Dorra (2016) ve como esenciales para el reconocimiento de la subjetivi-

⁶ Cozmo no tiene una página de internet en español. La versión oficial en inglés se encuentra en <https://www.anki.com/en-us/cozmo>

dad animal por los sujetos humanos: la expresión de afecciones, la utilización de un cierto tipo de lenguaje para intentar comunicarse con el usuario y el direccionamiento de su mirada hacia éste. A pesar de que esos robots no tienen la capacidad de sentir, parecen tener un objetivo similar al que se interpretó literalmente en la prueba de Turing (1950), si bien en dicho caso este objetivo se manifiesta en el plano de los sentimientos y no en el de la inteligencia. Es decir, estos robots bien podrían ser considerados como “sujetos *aparentemente* sintientes, aunque no humanos”. Conforme ésta y otras características más sean incorporadas a los agentes artificiales con una inteligencia superior, habrá que seguir revisitando la delimitación de *sujeto artificial* como parte de los objetivos centrales de una semiótica computacional.

Referencias

- AAMOTH, Doug (9 de junio 2014). “Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test”. *Time*. Disponible en: <http://time.com/2847900/eugene-goostman-turing-test/>
- AARONSON, Scott (12 de junio 2014). “My Conversation with ‘Eugene Goostman,’ the Chatbot that’s All Over the News for Allegedly Passing the Turing Test”. *The Blog of Scott Aaronson*. Disponible en: <https://www.scottaaronson.com/blog/?p=1858>
- CÁMARA FEDERAL DE CASACIÓN PENAL (18 de diciembre 2014). Causa CCC 68831/2014/CFC1. Capital Federal, Ciudad Autónoma de Buenos Aires. Disponible en: <http://www.saij.gob.ar/descarga-archivo?guid=rstuvwfa-llos-comp-uest-ol4261110pdf&name=14261110.pdf>
- COECKELBERGH, Mark (2015). “Artificial agents, good care, and modernity”. *Theoretical Medicine and Bioethics*, vol. 36, núm. 4, pp. 265-277.

- COPELAND, Jack & PROUDFOOT, Diane (2011). "Alan Turing, Father of the Modern Computer". *The Rutherford Journal. The New Zealand Journal for the History and Philosophy of Science and Technology*, núm. 4. Disponible en: <http://www.rutherfordjournal.org/article040101.html>
- CORTE SUPREMA DE JUSTICIA (26 de julio de 2017). Radicación núm. 1700122130002017-00468-02. Bogotá, D.C. Disponible en: <https://redjusticiaambientalcolombia.files.wordpress.com/2017/07/habeas-corpus-oso-chucho.pdf>
- DORRA, Raúl (julio-septiembre 2016). "Alteridad y proximidad: para una semiótica del cuidado". *Elementos*, núm. 103, vol. 23, pp. 3-12. Disponible en: <http://elementos.buap.mx/num103/pdf/Elem103.pdf>
- GASCA SALAS, J. (2007). "Alteridad, corporalidad social del sujeto y politicidad" (Artículos y Miscelánea). *Mundo Siglo XXI. Revista del Centro de Investigaciones Económicas, Administrativas y Sociales del Instituto Politécnico Nacional*, núm. 8, pp. 97-108.
- GONZÁLEZ DE REQUENA FARRÉ, Juan Antonio. (2017). "Fronteras del sujeto. Un enfoque liminar de la subjetividad lingüística". *Lenguaje*, núm. 45, vol. 1, pp. 89-113.
- HARARI, Yuval Noah (2016). *Homo Deus: breve historia del mañana*. México: Peguin Random House.
- HERNÁNDEZ, Isauro (1963). "La psicología terapéutica rogeriana". *Revista de Psicología*, núm. 8, pp. 173-179.
- IBM (2011a). "A Computer Called Watson". *IBM 100. Icons of Progress*. Disponible en: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/watson/>
- _____ (2011b). "Deep Blue". *IBM 100. Icons of Progress*. Disponible en: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/>
- JURAFSKY, Dan & MARTIN, James H. (11 de octubre 2017) [en preparación]. *Speech and Language Processing: An Introduction to Language Natural Processing, Computational Linguistics*,

and Speech Recognition. Disponible en: <https://web.stanford.edu/~jurafsky/slp3/>

(2008). *Speech and Language Processing: An Introduction to Language Natural Processing, Computational Linguistics, and Speech Recognition*. Upper-Saddle River, Nueva Jersey: Pearson/Prentice Hall.

LAUKYTE, Migle (2017). “Artificial agents among us: Should we recognize them as agents proper?” *Ethics and Information Technology*, núm. 19, pp. 1-17.

LOPATTO, Elizabeth (10 de junio 2014). “The AI That Wasn’t: Why ‘Eugene Goostman’ Didn’t Pass the Turing Test”. *The Daily Beast*. Disponible en: <https://www.thedailybeast.com/the-ai-that-wasnt-why-eugene-goostman-didnt-pass-the-turing-test>

OBERTI, Liliana (2003). “Semiótica de las pasiones y sujeto de conocimiento”. XI Jornadas de Reflexión Académica. Buenos Aires: Universidad de Palermo, pp. 81-83.

PRESS ASSOCIATION (9 de junio 2014). “Computer simulating 13-year-old boy becomes first to pass Turing test”. *The Guardian*. Disponible en: <https://www.theguardian.com/technology/2014/jun/08/super-computer-simulates-13-year-old-boy-passes-turing-test>

PRESS TEAM AT THE UNIVERSITY OF READING (8 de junio 2014). “Turing Test success marks milestone in computing history”. University of Reading Press Releases. Disponible en: <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>

RICO SULAYES, Antonio (2016). “Las funciones de creencia en la lingüística computacional: una reconsideración de los modos del creer”. *Tópicos del Seminario. Revista de Semiótica*, Los modos del creer, núm. 36, pp. 139-161.

SAMPLE, Ian & HERN, Alex (9 de junio 2014). “Scientists dispute whether computer ‘Eugene Goostman’ passed Turing test”. *The Guardian*. Disponible en: <https://www.theguardian.com/technology/2014/jun/09/scientists-disagree-over-whether-turing-test-has-been-passed>

- SELLERS, Michael (2013). "Toward a comprehensive theory of emotion for biological and artificial agents". *Biologically Inspired Cognitive Architectures*, núm. 4, pp. 3-26.
- TURING, Alan M. (1945). *Proposed electronic calculator*. Report for National Physical Laboratory.
- _____ (1950). "Computing Machinery and Intelligence". *Mind*, núm. 9, pp. 433-460.
- VELOSO, Manuela M. (2002). "Entertainment robotics". *Communications of the ACM*, núm. 45, vol. 3, pp. 59-63.
- WEIZENBAUM, Joseph (1966). "Eliza—A computer program for the study of natural language communication between man and machine". *Communications of the ACM*, núm. 9, vol. 1, pp. 36-45.
- _____ (1976). *Computer Power and Human Reason: From Judgement to Calculation*. Nueva York: W.H. Freeman and Company.