



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

El entendimiento lingüístico en la Inteligencia Artificial: Una relación ambivalente con Descartes

Por: Rodrigo González F.⁶⁸
rodgonfer@gmail.com

*Awareness is like consciousness. Soul
is like spirit.
But soft is not like hard and weak is
not like strong.
A mechanic can be both soft and hard,
a stewardess can be both weak and
strong.
This is called philosophy or a world-
view.
RACTER, en *The Policeman's Beard is
Half Constructed*.*

Resumen

En este artículo se examina de qué forma los investigadores de la Inteligencia Artificial han asumido un desafío propuesto por Descartes: la imposibilidad de construir máquinas programadas que, al entender lenguaje, evidencien que son pensantes. Tal desafío, que se enmarca en la filosofía metafísica cartesiana, distingue entre cosa pensante y extensa, siendo imposible la existencia de pensamiento en esta última. El lenguaje evidencia la imposibilidad de la inteligencia de máquina, de hecho. Como se examina, al enfrentar el desafío cartesiano, dichos investigadores han debido suponer como verdadera parte de la teoría cartesiana, a saber, que el uso flexible de lenguaje implica la existencia de pensamiento. Por ello califico de ambivalente la relación que tiene la Inteligencia Artificial con el filósofo francés: por una parte, rechazan la imposibilidad en principio de la inteligencia de máquina; por otra, están de

⁶⁸ Profesor del Departamento de Filosofía de la Facultad de Filosofía y Humanidades de Universidad de Chile.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

acuerdo con Descartes en que el uso flexible de lenguaje es suficiente para mostrar que alguien piensa. Si bien el carácter ambivalente entre la IA y Descartes parece anecdótico, no lo es. El hecho de que aún se considere el Test de Turing como una evaluación adecuada de la inteligencia corrobora la compleja relación entre la IA y el pensamiento cartesiano.

Palabras clave: Entendimiento lingüístico, lenguaje, pensamiento, Test de Turing.

Abstract

*This article deals with the manner in which the AI researchers have taken a Cartesian challenge: the impossibility to create machines which, via linguistic understanding, evince thought. Such a challenge, which is part of the metaphysical Cartesian philosophy, distinguishes between *res cogitans* and *res extensa*. It is impossible for the latter to have thought. Language, then, provides evidence of the impossibility of machine intelligence. As I analyze, by facing the Cartesian challenge, those researchers have supposed to be true part of Descartes' theory, namely, that the flexible use of language entails thought. For this reason I consider that the relation between AI and Descartes is ambivalent: on the one hand, AI researchers reject that machine intelligence is impossible in principle; on the other hand, they agree with Descartes in relation to the importance of language when it comes to providing evidence of thought. Even though the ambivalent character of AI and Descartes seems anecdotic, it is not. The fact that some people still regard the Turing Test as an adequate evaluation of intelligence verifies the complex relation between AI and the French philosopher.*

Keywords: Linguistic understanding, Language, thought, Turing Test.

Introducción

Es usual que se fustigue a Descartes por haber legado

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

el problema mente-cuerpo a la filosofía contemporánea. En particular, se lo ataca por sostener que las substancias que pueden concebirse de manera separada, esto es, de forma independiente, son metafísicamente diferentes, pues dado que al menos la omnipotencia divina puede separarlas, son separables. Esto, que precisamente ocurriría con la mente y el cuerpo, se dice que provoca todos los problemas para justificar de qué manera ambas cosas se relacionan.

Sin embargo, no solo en relación con el problema mente-cuerpo los filósofos y científicos contemporáneos han vapuleado a Descartes. También lo han hecho en el ámbito de una nueva disciplina científica: la Inteligencia Artificial. En efecto, desde el albor de la misma, en el siglo XIX, se ha criticado a Descartes como “el” filósofo que, al separar la mente de lo físico, creyó imposible en principio mecanizar el pensamiento. Esto se explica por el siguiente argumento cartesiano: si la mente es no material, y las máquinas *solo* operan en virtud de mecanismos físicos, como los engranajes por ejemplo, es imposible en principio la inteligencia de máquina. Este argumento es lo que llamo *el desafío cartesiano* a dicha inteligencia, un concepto predecesor de la IA.

En este ensayo me aboco a mostrar que dicho desafío ha tenido un curioso efecto en los investigadores de tal disciplina: para responderlo, se han acostumbrado a asumir parte de la teoría cartesiana, i.e., que el uso flexible de lenguaje



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

implica entendimiento lingüístico, y que este implica pensamiento. Es por esto, argumento, que muchos investigadores de la Inteligencia Artificial consideran que el Test de Turing es adecuado para justificar que las máquinas programadas tienen tal entendimiento (MOOR, 1976 y COPELAND, 2016) y, en consecuencia, estados mentales.

El artículo está dividido en cuatro secciones. En la primera expongo el pensamiento metafísico cartesiano, el cual afirma que, en tanto el cogito es más fácil de conocer que el cuerpo, puede distinguirse *realmente* de este. Si esto es así, resulta imposible en principio que una máquina, o cualquier dispositivo físico, piense. El hecho de que máquinas no puedan usar el lenguaje de manera flexible corroboraría, empíricamente, tal imposibilidad. Luego, en la segunda sección, expongo de qué forma Turing, con su famoso test, enfrentó de lleno el desafío de Descartes, pero para hacerlo, tuvo que asumir como verdadera parte de su teoría, a saber, que la evidencia empírica de que las máquinas hablan es suficiente para sostener que piensan. En la tercera sección muestro cómo SAM, el Script Applier Mechanism de Schank y Abelson, pone en duda la negativa de Descartes de que puede existir inteligencia de máquina. Inspirándose en Turing, afirman que el entendimiento lingüístico es, al igual que el pensamiento, mecánico. Finalmente, en la cuarta sección discuto las dificultades que conlleva, desde el punto de vista de lo psicológico, la mecanización del pensamiento y del



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

entendimiento lingüístico. Me concentro, para estos efectos, en dos dificultades asociadas: la introspección como fundamento del entendimiento lingüístico, y el carácter no consciente de los algoritmos.

El desafío de Descartes: la imposibilidad de mecanizar lenguaje y pensamiento

Descartes es uno de los filósofos más influyentes, pero también uno de los más polémicos de todos los tiempos. Su distinción entre la substancia mental y corporal tiene una importante consecuencia con relación a la imposibilidad, en principio, de que el pensamiento se mecanice. Dicha distinción, *real* como la llama, puede captarse con precisión en el siguiente pasaje de las *Meditaciones Metafísicas*:

En primer lugar, puesto que ya sé que todas las cosas que concibo clara y distintamente pueden ser producidas por Dios tal y como las concibo, me basta con poder concebir clara y distintamente una cosa sin otra, para estar seguro que la una es diferente de la otra [...]

Por lo tanto, como sé de cierto que existo, y, sin embargo, no advierto que convenga a mi verdadera naturaleza o esencia otra cosa que ser cosa pensante, concluyo rectamente que mi esencia consiste en ser solo una cosa que piensa, o una substancia o esencia que consiste solo en pensar. Y aunque acaso, o con toda seguridad como diré enseguida, tengo un cuerpo al que estoy estrechamente unido, con todo, puesto que, por una parte, tengo una idea clara y distinta de mí mismo, en cuanto que soy solo una cosa que piensa –y no

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

extensa-, es cierto entonces que ese yo (es decir, mi alma, por la cual soy lo que soy), es enteramente distinto de mi cuerpo, y que puede existir sin él. (DESCARTES, 1985a, p. 54, AT VII, 78).

El argumento cartesiano mediante el cual se concluye que el pensamiento puede existir sin la corporalidad se denomina intuición modal. Pende de la potencia de concebir como diferentes dos cosas mediante el entendimiento, lo cual a su vez descansa en la omnipotencia de Dios. Para el caso, Descartes sostiene que del alma se tiene una idea clara y distinta, no extensa; luego, el alma no es el cuerpo y puede existir sin este. Tal argumento fundamenta el Dualismo cartesiano, una tesis metafísica que distingue alma y cuerpo. Es, por supuesto, polémico y ya los contemporáneos de Descartes lo objetaron por varias razones que no vale la pena explicar en detalle aquí.

Pese a las objeciones, el filósofo francés insiste en que hay otra diferencia fundamental entre el espíritu y el cuerpo: solo este último es divisible. Apoya su tesis dualista de la siguiente forma:

[...] Hay gran diferencia entre el espíritu y el cuerpo; pues el cuerpo es siempre divisible por naturaleza, y el espíritu es enteramente indivisible. En efecto: cuando considero mi espíritu, o sea, a mí mismo en cuanto que soy solo una cosa pensante, no puedo distinguir en mí partes, sino que me entiendo como una sola cosa y entera. Y aunque el espíritu todo parece estar unido al cuerpo todo, sin embargo, cuando se separa de mi cuerpo un pie, un brazo, o alguna otra parte, sé que no por ello se le quita algo a mi espíritu

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

(DESCARTES 1985a, p. 59, AT VII, 86).

Estos pasajes de las Meditaciones son claros en cuanto a la diferencia metafísica que existe entre la mente y el cuerpo. La primera, aunque está estrechamente unida a la corporalidad⁶⁹, es, existe, de un modo diferente a todo lo físico: lo hace solo siendo una cosa pensante. Tal cosa es unitaria, a diferencia de la corporalidad, que es divisible. Si esto es correcto, y la distinción *real* es correcta también, se sabe con certeza que un cuerpo no puede ser pensante, porque es metafísicamente imposible que lo sea. Cuando se piensa una cosa extensa, no hay nada pensante que el entendimiento capte. Así, tal como argumenta el francés en relación con su famoso ejemplo del trozo de cera, la esencia de este es ser una cosa extensa, una cosa solo percibida por la mente, cuya esencia es puro escrutinio mental (DESCARTES 1985a, p. 21, AT VII, 31). Ello no es una cuestión trivial, puesto que significa que no es posible construir una máquina, un objeto material, físico, que tenga la propiedad de pensar. No puede ser más claro el razonamiento metafísico de Descartes en este sentido. Tampoco puede ser más claro el Dualismo que propone, independientemente de la discusión académica que ha suscitado respecto de si es de

69. Que algo esté unido a otra cosa no implica que ambas cosas sean idénticas. Descartes aclara este punto en la Sexta Meditación, con la dis-analogía del piloto y el navío (DESCARTES 1985a, p. 56, AT VII, 81). Cabe destacar que cuando cito a Descartes también agrego la correspondencia AT en la edición Adams Tannery.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

substancias o de propiedades.⁷⁰

Con relación a las máquinas, él adiciona una razón específica y concreta sobre por qué es imposible en principio que piensen. Lo hace fundamentándose en la evidencia empírica que disponemos respecto de por qué toda máquina no puede usar signos convencionales lingüísticos. Es decir, de la incapacidad de ellas de usar lenguaje para expresar pensamientos. Este pasaje, del *Discurso del Método*, explica el criterio cartesiano para estar cierto de la existencia de pensamiento:

[...] Si hubiera máquinas que tuviesen los órganos y la figura exterior de un mono, o de cualquier otro animal irracional, no tendríamos ningún medio de reconocer que eran en todo de igual naturaleza que estos animales; al paso que si hubiera otras semejantes a nuestros cuerpos y que imitasen nuestras acciones cuanto moralmente fuera posible, siempre tendríamos dos medios seguros de reconocer que no por eso eran hombres verdaderos. El primero sería que *jamás podrían usar ni de las palabras ni de otros signos compuestos de ellas*, como hacemos nosotros para declarar a los demás nuestros pensamientos. Pues se puede concebir que una máquina esté hecha de tal manera que profiera palabras, y aun que pronuncie algunas con ocasión de las acciones corporales que causan algún cambio en sus órganos –como, por ejemplo, si se le toca en una parte que pregunte lo que se le quiere decir, y en otra, que grite que se le hace daño y otras cosas semejantes–, *pero no que arregle las palabras de diversos modos para responder según el sentido de cuanto en su presencia se diga* cómo pueden hacer aun los más estúpidos de los hombres. El segundo

70. Véase, por ejemplo, de qué manera Clarke (2003) argumenta que el Dualismo cartesiano es de propiedades, no de substancias, como la mayoría de los intérpretes proponen. Por razones de espacio solo menciono esta discusión.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

consiste en que, por más que estas máquinas hicieran muchas cosas tan bien o acaso mejor que nosotros, se equivocarían infaliblemente en otras, y así se descubriría que no obraban por conocimiento, sino tan solo por la disposición de sus órganos [...] (Descartes 1985b, pp. 139-140, AT VI, 56-57, énfasis mío).

Del criterio cartesiano para la adjudicación de pensamiento se deriva una consecuencia fundamental, a propósito de la discusión de este ensayo. Si se une la metafísica de Descartes a su criterio para la adjudicación de pensamiento o estados mentales, una máquina no solo no puede pensar, sino que no puede disponer de auténtico entendimiento lingüístico. Al carecer de *flexibilidad* para usar lenguaje, por tener una cantidad finita de respuestas posibles basadas en los engranajes, es evidente que las máquinas no entienden lenguaje, pues no pueden *responder* de diversas maneras significando lo mismo; luego, concluye el filósofo francés, no piensan.

En aras de la argumentación cartesiana, vale la pena insistir en que una máquina podría *simular* el uso de lenguaje y, por tanto, que piensa. El filósofo francés considera la posibilidad de que puedan construirse sofisticadas máquinas, mas estas no podrían igualar al ser humano en cuanto a la *flexibilidad* que va de la mano de la inteligencia genuina, la cual también da lugar a acciones inteligentes en tanto sensibles al contexto. Luego, aunque las máquinas simularan el uso de lenguaje, contarían solo como ingeniosos y sofisticados



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

artefactos, cuya conducta además, en lo global, no sería adecuada como la humana. Aunque exhibieran inteligencia *como-si* fueran humanas, por sus respuestas finitas, las máquinas solo *imitarían* la conducta lingüística humana, y más aún, las acciones concertadas mediante la inteligencia genuina.

La propuesta cartesiana, entonces, ha representado un desafío crucial para la Inteligencia Artificial, la cual desde sus comienzos ha intentado responder al francés.⁷¹ Justamente, la siguiente sección, que examina brevemente el Juego de la Imitación, exhibe de qué forma el Test que lo cobija enfrenta el desafío cartesiano con relación al entendimiento lingüístico. Pero, tal como sostengo, Turing, en aras de la argumentación, considera que el criterio cartesiano de atribución de inteligencia y mente es *cierto*. Luego, piensa, si se construyen máquinas que conversaran, se podría socavar la teoría cartesiana según la cual el pensamiento no puede mecanizarse en principio.

En el fondo algunos tienen razón al plantear que el Test de Turing es una suerte de experimento mental (por ejemplo, TITTLE, 2005, p. 50), uno de carácter refutatorio. Lo es, a mi juicio, porque para refutar que el lenguaje y la mente no pueden mecanizarse, se debe asumir, al menos en aras de la argumentación, que el criterio cartesiano para atribuir

71. En sus albores, en el siglo XIX y antes, solo existía el concepto de inteligencia de máquina. Varios investigadores de dicha disciplina concentraban sus esfuerzos para desarrollarla, y así refutar a Descartes. Destacan Babbage, Pascal, Schickard, Colmar, entre otros. La mayoría se dedicó a mostrar que era posible la mecanización del pensamiento matemático.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

pensamiento es correcto. Solo dicha suposición permitiría mostrar, mediante el Juego de la Imitación de Turing, que la imposibilidad en principio de que máquinas conversen y piensen es absurda.

Turing y su foco en el entendimiento lingüístico como signo de inteligencia

Entre los desarrollos tecnológicos de la Segunda Guerra Mundial hubo uno en 1943, COLOSSUS, el cual gatilló una verdadera revolución. Mediante esta compleja máquina, Alan Turing fue capaz de descifrar ENIGMA, el código secreto de los submarinos de la Alemania nazi. Dicho “hackeo” permitió a los aliados salvar numerosas vidas y, ciertamente, acortar un buen trecho la guerra. Sin embargo, si bien COLOSSUS podría contar como un computador programado, algunos cuestionan dicho estatus por no ser una máquina multipropósito, el carácter esencial que tiene todo computador, el cual es una máquina programada o programable (COPELAND, 1993).

Sin embargo, pese a los avances desde la Segunda Guerra Mundial, la sombra de Descartes continuaba acechando sobre la naciente Inteligencia Artificial. Turing, quien “tomó el guante del francés”, decidió enfrentarlo en la arena filosófica. Para estos efectos publicó en un ensayo seminal, “Computing Machinery and Intelligence” (TURING, 1950), el cual fue tan revolucionario como COLOSSUS. Lo fue porque abordó la



IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

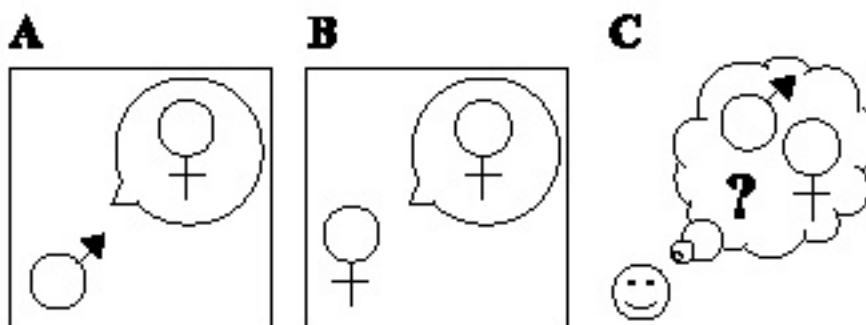
pregunta de si las máquinas programadas, los computadores, tenían estados mentales. Pero lo hizo, he aquí la novedad, proponiendo el reemplazo de la pregunta “Can a machine think?” por un test que recoge evidencia empírica a favor de la tesis de que los computadores entienden lenguaje y, luego, que piensan. Para eliminar la pregunta, la cual lleva a una discusión de conceptos y al uso de estos, se propone justamente el Juego de la Imitación, la esencia del test.

Dicho juego tiene tres etapas. En la primera, hay un hombre en una pieza A, una mujer en una pieza B y un juez cuyo sexo no importa. Cada uno tiene un rol específico. En el caso del juez, se debe determinar si está en presencia de un hombre o una mujer haciendo preguntas, mediante un teletipo, a las piezas A y B. La mujer responde de manera sincera las preguntas del juez con otro teletipo, mientras que el hombre se hace pasar por una mujer, y responde como si fuera tal del mismo modo que ella. Los interrogatorios, que duran 5 minutos, se concentran en preguntas simples para determinar el sexo de quienes están en las piezas A y B. La primera versión del juego de Turing puede esquematizarse de la siguiente manera:



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica



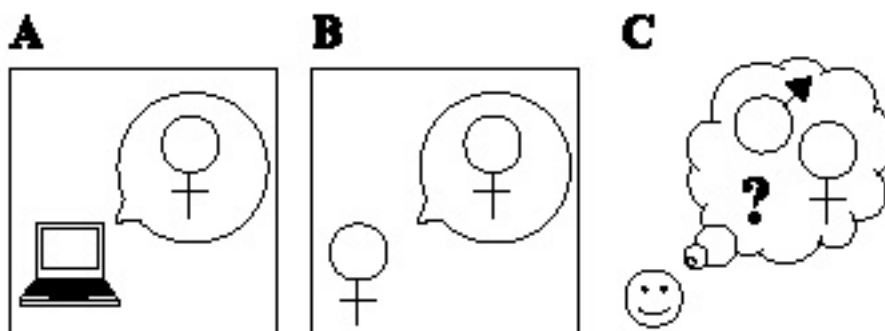
Ahora bien, vale la pena detenerse en una cuestión importante. Si el hombre logra engañar a una ronda de jueces, al menos a una parte de ellos, se dice que ha pasado el test. Un punto fundamental de la primera versión, y que ha sido sistemáticamente soslayada y malinterpretada por comentaristas, es que la identificación del sexo de los participantes no es un elemento que Turing haya elegido azarosamente (Cf. GONZÁLEZ, 2015). Dicha identificación no es gratuita, porque tiene el propósito de mostrar que la inteligencia femenina es separable de la corporalidad femenina. Luego, es posible tener corporalidad masculina y desempeñarse inteligentemente, tal como lo haría una mujer. Vuelvo sobre esta cuestión más abajo, ya que es no solo importante para entender el Funcionalismo de Turing, sino que también para explicar cómo este asume el criterio cartesiano de la inteligencia. Solo así es posible refutar la tesis sobre la imposibilidad de la Inteligencia Artificial.

Luego, en la segunda versión del juego, Turing pregunta: ¿qué sucedería si lo que hace el hombre en la pieza A lo hiciese un computador programado? Si la inteligencia es una

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

función separable de lo físico y, en particular, de lo biológico, entonces una máquina programada podría tener el mismo desempeño que una mujer. Claro, sería necesario que el computador tuviera la suficiente memoria y velocidad para que imitara a una mujer (y/o a un hombre, tal como queda claro en la segunda versión). Esta puede esquematizarse del siguiente modo:



Ahora bien, conviene hacer una aclaración importante. No estimo, respecto de la primera y la segunda versión, margen para interpretar que el hombre supere en inteligencia a la mujer, o lo contrario, pese a las teorías que consideran que el test es un proyecto emancipador de género. Recientes lecturas del Test de Turing han centrado el análisis en la cuestión del género (Genova 1994, Lassègue 1996), y han visto una agenda de Turing en este sentido. Sin embargo, es difícil que él haya tenido en mente un proyecto de esa índole, puesto que su objetivo, al menos declarado, era proporcionar una justificación filosófica sobre la plausibilidad de que se construyan máquinas pensantes.

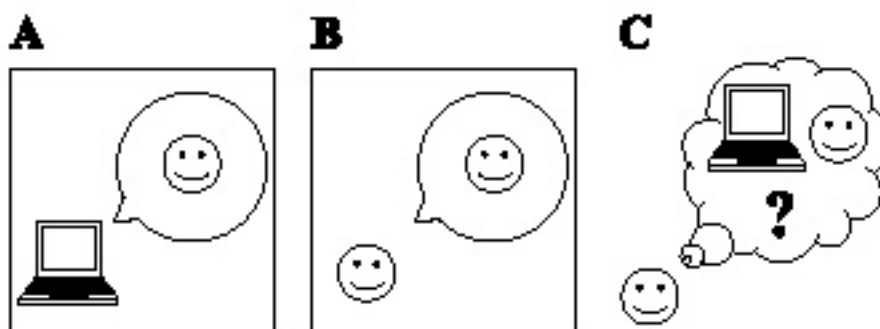
De hecho, una cuestión que apoya mi reticencia a

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

creer en el proyecto emancipador de género es que el propio Turing (1951b) simplificó su test; lo hizo, probablemente, porque la identificación del sexo de los participantes produjo algunos enredos en la interpretación de qué estaba detrás del juego. Dicha simplificación, que dio pie a “la versión estándar del test”, describe el siguiente escenario: un computador programado está en la pieza A, mientras que en la B hay una persona, cuyo sexo es irrelevante. Los jueces deben evaluar si están en presencia de una persona o de una máquina programada. Mientras que la persona responde sinceramente a las preguntas, el computador programado se hace pasar por un humano. Dado que nuevamente hay rondas de preguntas de 5 minutos, Turing sostiene que si el computador programado logra engañar al menos al 30% de los jueces, ha pasado el test. Lo ha hecho, según él, porque la máquina entiende lenguaje, y porque tal entendimiento implica la existencia de estados mentales o pensamiento, como hasta el más fanático de los cartesianos debería aceptar.

La versión estándar puede esquematizarse de la siguiente manera:





IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

Retomo un punto importante de arriba. El Juego de la Imitación que describe el test involucra una propuesta funcionalista profundamente anti-biológica. La capacidad de tener estados mentales es considerada por Turing como una propiedad funcional no biológica; es decir, como una función que no requiere instanciación en una configuración física específica, como la del cerebro.⁷² Por el contrario, él asume que un computador puede instanciarse en diversos materiales, tal como una Máquina de Turing, el concepto que permite entender qué es computar (TURING, 1936). Con ello él adhiere a un tipo de Funcionalismo, llamado posteriormente de Máquina (PUTNAM, 1967 y 1973), que considera la inteligencia como una mera función computable, ciertamente *separable* de la corporalidad y el cerebro. Para el lector atento, esto ciertamente recordará a la intuición modal cartesiana descrita en la primera sección. En efecto, el Funcionalismo de Turing es compatible con el Dualismo de Descartes, en el sentido de que ambos consideran que la inteligencia puede separarse del cerebro (Cf. GONZÁLEZ, 2011). Si en el caso de Descartes la inteligencia es separable del cuerpo mediante lo concebible, en el de Turing lo es mediante la

72. En este pasaje, Turing es claro en relación con su Funcionalismo anti-biológico. Dice, en este extracto de entrevista radial: “si alguna máquina particular puede ser descrita como un cerebro, solo debemos programar nuestra máquina digital para imitar a esta y también será un cerebro” (1951a: 112, trad. mía). Este extracto de la entrevista radial concedida a BBC en 1951 inspira una idea que nace a partir del Funcionalismo de Máquina de Turing: la posibilidad de alcanzar la inmortalidad mediante la transmigración del *software* a distintos *hardwares* (Rucker 1980).

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

programación de computadores.

Para algunos, incluso, el Funcionalismo de Máquina es lisa y llanamente un resabio de Dualismo, y en ese sentido, un heredero de la metafísica cartesiana. Tales suposiciones son bastante escépticas respecto de las posibilidades de la IA. Por ejemplo, Searle (1980) ha sido particularmente crítico porque tanto la IA como la Ciencia Cognitiva dan la espalda a la biología. Una de sus críticas más ácidas tiene como objetivo mostrar que la concepción de Turing de la mente, al suponer que una máquina programada puede instanciarse en cualquier material, *se aleja notoriamente de la cognición real humana*. Examinó la relación entre esta crítica y el entendimiento lingüístico en la sección final de este ensayo.

El argumento serleano contra el Funcionalismo de Máquina se expone de la siguiente manera:

[...] Ciertamente, la IA fuerte solo tiene sentido dado el supuesto dualista de que, en lo que concierne a la mente, el cerebro no importa. En la IA fuerte (y en el Funcionalismo también) lo que importa son los programas, y estos son independientes de su realización en máquinas; sin duda, en lo que respecta a la IA, el mismo programa podría realizarse por una máquina electrónica, una substancia mental cartesiana, o el espíritu hegeliano del mundo (SEARLE, 1980, p. 86, traducción mía).

Justamente las críticas serleanas a la IA fuerte tienen como objetivo el ensayo seminal que la inspira. En concreto, el clásico artículo de Schank y Abelson supone el



IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

diseño de un programa, SAM, el que supuestamente es capaz de entender lenguaje y, por tanto, de pensar. Con ello SAM asume, tal como he argumentado aquí, la teoría cartesiana de acuerdo con la cual, si un agente o una máquina entiende un lenguaje, entonces piensa.

SAM, el arquetipo del entendimiento lingüístico en la IA

El Test de Turing, pese a la discusión filosófica que ha generado y que probablemente seguirá generando, es considerado por muchos investigadores en Inteligencia Artificial como una marca distintiva y clara de la inteligencia de máquina. Ello porque el test es simple y eficaz en cuanto a la presentación de evidencia empírica: esta justificaría la tesis de que los computadores programados entienden lenguaje y, por tanto, que tienen estados mentales. Hay, en esta postura, una creencia de base, a saber, que si la máquina programada exhibe entendimiento lingüístico, entonces lisa y llanamente tiene mente.

Schank y Abelson (1977) son justamente simpatizantes del test. El programa que diseñaron, SAM (Script Applier Mechanism), fue inspirado por una teoría general del entendimiento. Una de acuerdo con la cual la gente entiende las situaciones en que se encuentra, y actúa en virtud de conocimiento organizado pasado. Cualquier proceso de entender un lenguaje requiere de conocimiento general, el cual permite que

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

una persona interprete el comportamiento de otra persona con base en necesidades y métodos compartidos y estándares (por ejemplo, por qué alguien se va a la cama). Pero también se interpreta en función de conocimiento específico, el cual hace posible la explicación y participación en situaciones experimentadas frecuentemente en la vida diaria (por ejemplo, por qué los boletos se muestran en la entrada de un estadio). Pero, ¿cuál es la conexión entre el conocimiento de base y la comprensión de una historia *nueva*?

De acuerdo con la teoría de Schank y Abelson, las historias nuevas, y que resultan comprensibles, hacen referencia a libretos [*scripts*] o a secuencias causalmente estandarizadas de eventos desde el punto de vista de un actor. Considérense, por ejemplo, las siguientes historias:

9. Flor entró a un pub y ordenó una caipiriña. Pagó la cuenta y se fue.
10. Fue fundamental que Flor no estuviera en la fiesta. El avaro de Agustín no había dejado nada de dinero de nuevo. La policía estaba haciendo controles rutinarios contra el manejo en estado de ebriedad.
11. Flor quería comprar un quitasol. Encontró un panfleto haciendo publicidad para una rebaja de temporada de playa en el bus. Los quitasoles se habían vendido todos.
12. Flor estaba en un supermercado. En la caja se dio cuenta de que había dejado su billetera en casa.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

Mientras que la historia 1 hace referencia a un libreto fácilmente identificable [PUB], que típicamente involucra una secuencia causal de eventos (deseo de tomar un trago, sentarse en el bar, ordenar el trago, pedir la cuenta, etc.), lo cual hace que los eventos narrados sean inteligibles, la historia 2, que carece de un libreto claro, exhibe una secuencia de eventos no relacionados. Por esta razón, la historia 2 tiene muy poco sentido y, sin información adicional o un gran esfuerzo, resulta francamente difícil de entender. Al tener sentido, la historia 3 traslapa dos libretos, [BUS] y [REBAJAS DE TEMPORADA], los cuales permiten que uno entienda de qué se trata la historia, y el final poco feliz para Flor. Del mismo modo que 3, 4 supone dos libretos [SUPERMERCADO] y [COMPRAR]. Pese a lo breve que es, resulta completamente inteligible y con pleno sentido.

A pesar de posibles complicaciones, tales como interferencias e interrupciones en la secuencia de eventos, el hacer corresponder libretos permite hacerse partícipes en situaciones públicas, y comprender narraciones sin la inclusión de detalles aburridos. En virtud de la internalización de libretos, *la gente es capaz de anticipar los eventos futuros de una historia, infiriendo información que no ha sido declarada explícitamente*. Según Schank y Abelson, esto último es un signo inequívoco de entendimiento genuino. Por ejemplo, para aquellos que realmente entienden la historias 3 y 4 es posible anticipar



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

la *frustración* de Flor, y la *vergüenza* de haber dejado su billetera en casa, respectivamente.

Sin embargo, la parte más interesante de su análisis es que, según ellos, la gente entendería historias tal como las máquinas programadas manipulan y procesan libretos, representaciones y conocimiento mundano (general y específico). En particular, ellos proponen como la gente y SAM entienden de modo similar así:

El entendimiento entonces, es un proceso mediante el cual la gente hace calzar lo que ve y escucha con grupos de acciones pre-almacenadas, las cuales ya han experimentado. La información novedosa se entiende en términos de información antigua. En esta teoría, *el ser humano es visto como un procesador* que solo entiende lo que ha entendido previamente. Nuestro programa basado en libretos, SAM, trabaja de esta forma [...] *Un ser humano que entiende está equipado con miles de libretos. Los usa casi sin pensar en los mismos* (SCHANK; ABELSON, 1977, p. 67, énfasis y traducción mías).

Ambos autores coinciden con Turing, quien considera la inteligencia como producto del desempeño de un computador programado. Para todos ellos los procesos mentales, tales como aquellos involucrados en el entendimiento, pueden descomponerse en operaciones mecánicas simples como *traer, recuperar, hacer calzar*. Quienes adscriben a la idea de que la mente es como un procesador informacional, creen que la capacidad de entender historias emerge de traerlas, recuperarlas y hacerlas calzar con libretos. Incluso, el hecho de que una máquina programada

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

responda a historias mediante tal procedimiento proporciona evidencia sólida de que tiene mente y, así, una variedad de estados mentales. Esto es lo que Searle denomina Inteligencia Artificial Fuerte, una aproximación que busca refutar con su famosa Habitación China (1980 y 1990). Dicho experimento mental muestra que es posible actuar con base en un programa, manipulando ideogramas chinos, y no entender nada. La moraleja es, entonces, que la sintáctica, o el programa, es insuficiente para la semántica, o el contenido mental.

Cabe destacar, a propósito de la discusión venidera, que no resulta del todo evidente que el procesamiento informacional de las máquinas programadas, cuando responden a preguntas sobre historias, pueda ser considerado como *evidencia suficiente* para la existencia de mente e inteligencia. Tal como analizo a continuación, correr programas que hacen que una máquina programada responda mediante el proceso de hacer coincidir libretos no implica que tales máquinas entiendan las historias y que posean vida mental.

Entendimiento lingüístico mecanizado: ¿psicológicamente plausible?

Hasta aquí resulta más o menos claro que, a pesar de lo que alguien podría pensar en la Ciencia Cognitiva, Descartes y Turing están en una suerte de relación teórica ambivalente. En efecto, si bien el primero niega en principio que las máquinas

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

puedan pensar, el segundo asume que el Funcionalismo de Máquina involucra que la inteligencia es separable del cerebro, y más importante aún, que usar lenguaje y tener entendimiento lingüístico implican tener mente y ser pensante. Tal como expongo más arriba, Turing opera tal como un experimentador mental que trata de refutar la teoría de Descartes acerca de la imposibilidad de inteligencia de máquina, que para el caso denomino T. Para llevarla al absurdo, se debe primero asumir que T es verdadera, lo cual ocurre en un experimento mental negativo o refutatorio (BROWN, 2007). Esto, tal vez, ha llevado a creer a los investigadores de la IA que el lenguaje y solo el lenguaje es *garantía* de que existe pensamiento, una tesis absolutamente cartesiana.⁷³ En ese sentido puede sostenerse que Descartes legó a la IA dicha garantía.

En esta sección me concentro en examinar si es psicológicamente plausible que el entendimiento lingüístico sea mecanizado. Con “psicológicamente plausible” quiero decir lo siguiente: ¿hay experiencias psicológicas que permitan evaluar si dicho entendimiento puede mecanizarse? Por ejemplo, no es psicológicamente plausible creer que uno es una molécula, porque las experiencias psicológicas asociadas a tal creencia son imposibles de imaginar, o bien, porque son una tergiversación lisa y llana de las experiencias de qué es ser como un humano.

73. Véase, por ejemplo, el programa de Colby (1975), PARRY. Dicho programa es una simulación de un paciente paranoico. Todo el énfasis de la conducta psicótica de PARRY se concentra en las respuestas de este, las cuales simulan el comportamiento lingüístico de un paranoico.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

Lo mismo sucede con la plausibilidad psicológica de qué es ser como un murciélago. Intentar serlo, como Nagel analiza (1974), lleva a imaginaciones en que uno, un humano, se *comporta* como si fuera un murciélago, en una suerte de conducta *Bat-Man*. De esta forma, mi tesis es que al apelar a experiencias psicológicas humanas se puede establecer si el entendimiento lingüístico puede llegar a mecanizarse.

En vista de este problema, la propuesta que defiende es que no es psicológicamente plausible mecanizar el entendimiento humano. Y no lo es por dos razones que parecen suficientemente convincentes. La primera es que la noción misma de entendimiento lingüístico es discutible y controvertida, porque requiere de un testigo que recurre a la *introspección*. Nadie puede decir que entiende un lenguaje L si en verdad no lo hace, salvo que esté mintiendo. Uno *sabe*, internamente, es decir desde el punto de vista de la primera persona, cuando entiende o no entiende un lenguaje. En ese punto tiene razón Searle (1990, p. 418): independientemente de toda la *elegante* discusión filosófica acerca de “S entiende L” (siendo L una lengua), es claro que uno y solo uno *sabe* cuándo entiende y cuándo no. Este sentido de entender, en consecuencia, descansa en la introspección.

La segunda razón es que, tal como uno puede argumentar con base en la Habitación China (SEARLE 1980, p. 418), los procesos algorítmicos, y derivadamente los

*IΦ-Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

sintácticos, no permitem que haya estados mentales que emerjan a partir de manipulaciones simbólicas. Al menos no lo hacen cuando uno implementa algoritmos, o cuando uno procesa sintácticamente de acuerdo con un libro de reglas, tal como sucede en la Habitación China. En efecto, a pesar de las múltiples objeciones a dicha habitación, resulta más o menos claro que hay una diferencia marcada entre la sintaxis y la semántica, en el sentido de que implementar un programa computacional sobre la base de manipulaciones simbólicas no es suficiente para que haya semántica. Incluso, de la manipulación de símbolos a partir de las instrucciones en un libro de reglas no emerge nada psicológicamente interesante, punto que recuerda, por supuesto, la primera razón que expongo aquí y que se analiza con más detalle a continuación.

Es posible argumentar en apoyo de las dos razones una cuestión crucial, a propósito de computar. No hay nada que uno pueda imaginar con relación a qué es ser como una máquina programada (BLOCK 1995, p. 270). Si uno implementa un algoritmo, nada psicológicamente interesante ocurre desde el punto de vista de las experiencias psicológicas de un agente cualquiera. Si esto último es correcto, SAM podría contar solo como un modelo para comprender mejor qué es el entendimiento lingüístico, toda vez que mediante el programa no es posible explicar el componente psicológico que va de la mano de dicho entendimiento, esto es, el proceso psicológico introspectivo mediante el cual



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

alguien que entiende *sabe* internamente que comprende una lengua.⁷⁴

Searle comprende cabalmente, como hablante nativo de inglés. Tiene una comprensión mucho menos perfecta del francés. Y del chino simplemente no comprende nada. En los tres casos el saber acerca de la comprensión es *interno* a él. Hay, en este sentido, una importante conexión entre el entendimiento y la semántica, punto clave de la Habitación China, que no debe dejarse de lado. Por ejemplo, Searle, al ser hablante nativo de inglés, sabe que comprende perfectamente que “insofar as” cuenta como “en cuanto a”. Como hablante de proeficiencia media del francés sabe que comprende que “Un demi” es una cerveza a presión. Finalmente, al ser incapaz de hablar chino, es igualmente incapaz de saber y comprender que gaseosa es “Sūdă”.

Sin embargo, es menester introducir un ejemplo final de un proceso de computar que no involucre ninguna experiencia psicológicamente relevante. El famoso algoritmo de Euclides servirá para este propósito (PENROSE, 1999, pp. 41-44), pese a

74. Alguien podría creer, y correctamente me parece, que saber internamente apela a una suerte de cartesianismo fuerte en relación con la evaluación de la existencia de entendimiento lingüístico. Descartes mismo propone que solo mediante el juzgar del cogito uno sabe que la cera es cera y que los humanos no son autómatas (DESCARTES, 1985a, p. 21, AT VII, 32). Cabe destacar, a propósito de lo que llamo cartesianismo fuerte, que Searle, pese a sus estertóreas declaraciones, es bastante cartesiano. Por ejemplo, y tal como argumento aquí, lo es cuando propone que los estados mentales intencionales tienen condiciones de satisfacción sabidas por el agente (Cf. SEARLE, 1983, p. 64). También cuando se examina en detalle la Repuesta del Sistema, la cual supone un elemento imposible de internalizarse: Searle mismo (GONZÁLEZ, 2012). La discusión sobre la relación entre Searle y Descartes da para un ensayo aparte, sin duda.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

que en estricto rigor no es tan original. No vale la pena que lo sea, toda vez que los algoritmos, o programas, operan todos de la misma manera.

Según Euclides, para encontrar el máximo común denominador de dos números, basta con seguir las siguientes reglas, expresadas mediante pasos finitos, siendo el tercero recursivo:

- i) Dividir número divisor, anotando resultado y remanente (R);
- ii) Si $R = 0$, *halt*;
- iii) Si $R \neq 0$, tomar divisor y remanente anteriores para volver a ejecutar paso 1.

Por ejemplo, para encontrar el máximo común denominador de 99 y 15, se debe operar así:

Número	Número divisor	Resultado	Remanente
99	15	6	9
15	9	1	6
9	6	1	3
6	3	2	0

Una vez implementados los pasos, es posible establecer que el máximo común denominador de 99 y 15 es 3. Es crucial, en este ejemplo, que no surge ninguna experiencia psicológica asociada a las matemáticas, ni tampoco a conceptos de esta disciplina. De hecho, el algoritmo puede ser seguido por alguien cuyos conocimientos matemáticos sean básicos, o por un experto, y en ninguno de ambos habrá nada *mental* que ocurra. Es

IΦ-*Sophia*

Revista eletrônica de investigação filosófica, científica e tecnológica

decir, en ningún agente que opere en función de las reglas del algoritmo de Euclides emergerá algo psicológico asociado a la implementación de dicho algoritmo, o programa. Lo mismo, *mutatis mutandis*, ocurre con el entendimiento lingüístico, tal como muestra la Habitación China de Searle: un hablante de inglés que corra un programa para conversar en chino no sabe qué se le pregunta y qué responde, menos tiene estados mentales asociados a la implementación del programa.

Este tipo de ejemplo muestra que Descartes parece estar en lo correcto cuando afirma que el lenguaje y la mente no pueden mecanizarse. Las razones son distintas a lo que el francés cree, sin embargo. En el caso de él, cree que ni lenguaje ni pensamiento pueden mecanizarse, porque los mecanismos son finitos y, por tanto, inflexibles, mientras que la razón es *res cogitans* flexible. En cambio, para Searle y quienes argumentan como él, razón y lenguaje no pueden mecanizarse mediante programas, porque no hay nada psicológicamente interesante que ocurra cuando se implementan estos. Luego, la Inteligencia Artificial no puede enseñarnos nada de la cognición inteligente real solo si opera con programas, los que, supuestamente, son capaces de conversar tal como los humanos, desiderátum impuesto por el propio Descartes.

Uno puede sintetizar la falencia de la IA para generar entendimiento lingüístico así: un ser humano puede ser como un programa computacional al correrlo, aunque no haya nada



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

psicológicamente interessante asociado a dicho proceso. Y si no hay nada psicológicamente interesante asociado a ser una máquina programada, esta simplemente no puede ser una mente. Difícilmente, entonces, puede tal máquina tener ligado dicho entendimiento, salvo que lo simule. Pero, eso es justamente lo que los investigadores de la denominada IA fuerte pretenden superar, mediante máquinas que conversen con humanos, tal como el desafío de Descartes establece. Como he mostrado aquí, dicho desafío se transforma en legado para los investigadores de la IA. El caso del entendimiento lingüístico y de los investigadores de tal disciplina recuerda entonces la situación en que, de tanto criticar, uno termina asumiendo lo menos pensado: la postura del contrario.

Referencias

- BLOCK, N. *“The mind as software of the brain”* In Heil, J. (ed.) **Philosophy of Mind: a Guide and Anthology**. Oxford: OUP, 1995. p. 267-274.
- BROWN, J. R., *“Counter Thought Experiments”* In **Royal Institute of Philosophy Supplement** 61, 82, p. 155-177, 2007.
- CLARKE, D. *Descartes’s Theory of Mind*. Oxford: Clarendon Press, 2003.
- COLBY, K. *Artificial paranoia*. New York: Pergamon Press, 1975.
- COPELAND, J. *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell, 1993.
- COPELAND, J. *“The Church-Turing Thesis”*. Disponible en: <http://plato.stanford.edu/entries/church-turing/>. Último acceso en 21.03.2016.
- DESCARTES, R. *“Meditations on First Philosophy”* In Cottingham, J.; Stoothoff, R.; Murdoch, D. **The Philosophical Writings of Descartes Vol. II**. New York: Cambridge University Press, 1985a.



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

p. 1-62.

DESCARTES, R. “*Discourse on the Method*” In Cottingham, J.; Stoothoff, R.; Murdoch, D. **The Philosophical Writings of Descartes** Vol. I. New York: Cambridge University Press, 1985b. p. 109-151.

GENOVA, J. “*Turing’s sexual guessing game*” In **Social Epistemology**, v. 8, n. 4, p. 313-326, 1994.

GONZÁLEZ, R. “Descartes, las Intuiciones Modales y la IA” In **Alpha** n. 32, p. 181-198, 2011.

GONZÁLEZ, R. “La pieza china: un experimento mental con sesgo cartesiano” In **Revista Chilena de Neuropsicología**, Vol. 7, edición especial, 1-6, 2012.

GONZÁLEZ, R. “¿Importa la determinación del sexo en el Test de Turing?” In **Aurora**, v. 27, n. 40 (enero-abril), p. 277-295, 2015.

LASSÈGUE, J. “*What kind of Turing Test did Turing have in mind?*” In **Tekhnema**, n. 3, p. 37-58, 1996.

MOOR, J. “*An Analysis of the Turing test*” In **Philosophical Studies** 30, p. 249-57. Reimpreso en Shieber, S. **The Turing test: Verbal Behaviour as the Hallmark of Intelligence**. Cambridge, Mass.: MIT Press, 1976. p. 297-306.

NAGEL, T. “*What is it like to be a bat?*” In **Philosophical Review** n. 83, p. 435-50, 1974.

PENROSE, R. **The Emperor’s New Mind**. Oxford: Oxford University Press, 1974.

PUTNAM, H. “*Psychological predicates*” In Capitan, W.; Merrill, D., **Art, Mind, and Religion**. Pittsburgh: University of Pittsburgh Press. Reimpreso en Heil, J. **Philosophy of Mind: A Guide and Anthology**. Oxford: Oxford University Press, 1967. p. 160-167.

PUTNAM, H. “*The nature of mental states*”, originalmente publicado como Psychological Predicates In Capitan, W.; Merrill, D., **Art, Mind, and Religion**. Pittsburgh: University of Pittsburgh Press. Reimpreso en Chalmers, D. **Philosophy of Mind: Classical and Contemporary Readings**. New York: Oxford University Press, 1973. p. 73-79.

RUCKER, R. **Software**. New York: Harper Collins, 1982.

SCHANK, R.C.; ABELSON, R.P. **Scripts, Plans, Goals, and**



IΦ-Sophia

Revista eletrônica de investigação filosófica, científica e tecnológica

- Understanding.** Hillsdale, N.J.: Erlbaum, 1977.
- SEARLE, J. “Minds, brains and programs” In *Behavioral and Brain Sciences* n. 3, p. 417-24, 1980.
- SEARLE, J. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press, 1983.
- SEARLE, J. “Is the brain’s mind a computer program?” In *Scientific American*, January 1990, p. 20-25.
- TURING, A. “On computable numbers, with an application to the Entscheidungsproblem” In *Proceedings of the London Mathematical Society*, series 2, v. 42, p. 231-65 (con correcciones en v. 43, p. 544-6), 1936.
- TURING, A. “Computing intelligence and machinery” In *Mind* LIX, n. 2236, Oct. 1950: p. 433-60. Reimpreso en: Boden, M. *The Philosophy of Artificial Intelligence.* Oxford: OUP, 1950. p. 40-66.
- TURING, A. “Can Digital Computers Think?”, tipeo de una entrevista radial en el tercer programa de BBC, del 15 de mayo de 1951. Número de referencia de los Archivos Turing: B.5. Reimpreso en Shieber, S. *The Turing Test: Verbal Behavior as the Hallmark of Intelligence.* Cambridge, Mass.: MIT Press, 1951a. p. 111-116.
- TURING, A. “Intelligent machinery, a heretical theory”, manuscrito inédito de una conferencia en la “51 Society” en Manchester, Inglaterra. Número de referencia de los Archivos Turing: B.4. Reimpreso en Shieber, S. *The Turing Test: Verbal Behavior as the Hallmark of Intelligence.* Cambridge, Mass.: MIT Press, 1951b. p. 105-109.
- TITTLE, Peg. *What if...Collected Thought Experiments in Philosophy.* New York: Longman, 2005.