



“NÃO SOU UM ROBÔ”

o que os teste de linguagem nos
permitem dizer sobre isso?

“I'M NOT A ROBOT”

what do language tests allow us to say about it?

Gabriel de Albuquerque Barbosa Baumann¹
Universidade Federal do Rio Grande do Norte

¹ Mestrando em Filosofia pela Universidade Federal do Rio Grande do Norte (UFRN), e também Bacharel em Filosofia (2024), em Ciências e Tecnologia (2016), Engenharia Mecânica (2018) e mestre em Engenharia Mecânica (2021) pela mesma instituição.

E-mail: gabriel.baumann.108@ufrn.edu.br.

Lattes: <http://lattes.cnpq.br/9013391108821022>. Orcid: <https://orcid.org/0009-0004-5408-848X>.



RESUMO: Neste artigo, argumentamos que um *Teste de Linguagem* não é suficiente para o reconhecimento da inteligência como uma propriedade intrínseca tanto de máquinas como de seres humanos. Explicamos que defender o contrário decorre de uma noção cartesiana de "natureza humana", que leva à confusão entre as interpretações ontológica e epistêmica desses testes: não se segue do fato que "eu sei o que sou" que "eu sei que algo é ou não é como eu". Assim, qualquer *Teste de Linguagem* que vise a identificação de algum *marco* ontológico da humanidade (racionalidade, pensamento, intencionalidade) apresenta esse mesmo problema, que, em última instância, leva ao solipsismo e ao *Problema das Outras Mentes*. A fim de evitar esse desfecho, temos de admitir a possibilidade de existência da Inteligência Artificial no sentido forte ou assumir que, contraintuitivamente, talvez não saibamos com máxima clareza e distinção se somos humanos ou máquinas.

Palavras-chave: Testes de Linguagem. Inteligência Artificial. Natureza Humana. Solipsismo. Problema das outras mentes.

ABSTRACT: In this paper, we argue that a Language Test is not sufficient for recognizing intelligence as an intrinsic property of both machines and human beings. We explain that defending the opposite stems from a Cartesian notion of "human nature," which leads to confusion between the ontological and epistemic interpretations of these tests: it does not follow from the fact that "I know what I am" that "I know whether something is or is not like me." Thus, any Language Test aimed at identifying some ontological hallmark of humanity (rationality, thought, intentionality) presents this same issue, which ultimately leads to solipsism and the Problem of Other Minds. In order to avoid this outcome, we must either admit the possibility of the existence of strong Artificial Intelligence or accept that, counterintuitively, we may not know with maximum clarity and distinction whether we are human or machine.

Keywords: Language Testing. Artificial Intelligence. Human Nature. Solipsism. Problem of Other Minds.



INTRODUÇÃO

Alan Turing (1950) sugeriu que a pergunta "as máquinas podem pensar?" fosse sem sentido. Porém, admitiu que, no futuro (em relação àquela época), não seria estranho que falássemos de máquinas pensando sem a expectativa de sermos contestados, porque elas seriam indistinguíveis linguisticamente de nós seres humanos. Ele vislumbrava que, com os avanços técnicos e tecnológicos, da engenharia e da ciência da computação, logo pudéssemos falar em algum tipo Inteligência Artificial (IA). De fato, apenas seis anos depois de seu artigo *Computing Machinery and Intelligence*, em uma conferência ocorrida na Dartmouth College, em Hanover, New Hampshire, o campo da Inteligência Artificial (IA) foi operacional e paradigmaticamente definido (Bringsjord; Govindarajulu, 2024).

Essa área de pesquisa, conforme sugerem Russell e Norvig (1995, 2002, 2009), pode ser definida nos termos de seus objetivos. Com isso, os autores apresentam quatro categorias de abordagens à IA: *pensar como humanos*, *pensar racionalmente*, *agir como humanos* ou *agir racionalmente* (Russell; Norvig, 2016). Estas duas últimas são focadas no comportamento observável de um sistema (agente), com o intuito de desenvolver uma IA que aja como seres humanos ou de acordo com algum modelo de racionalidade. Já as duas primeiras abordagens visam compreender e implementar os fundamentos daquilo que seria o pensamento humano ou racional, a fim de produzir, por consequência de uma suposta cognição intrínseca, um comportamento que possa ser dito inteligente em um sistema.

Mas que critérios podem ser utilizados para identificar uma IA como realmente inteligente, e não como algo que apenas parece ser? Embora tenha sido impressionante ver os programas *Deep Blue* e *AlphaGo* vencerem os mestres Gary Kasparov (xadrez) e Lee Sedol (*Go*), isso não parece ser suficiente para chamar um programa desse tipo de inteligente. A questão permanece: qual será o *marco* da inteligência da IA²?

Nesse contexto, a proposta de Turing é um experimento mental potente, porque, além de convidar os cientistas e pesquisadores a produzir uma IA que passe em seu teste, mobiliza questões filosóficas relevantes acerca daquele que é considerado, pela tradição filosófica, principalmente cartesiana, ao longo da história, o *marco* de humanidade (logo, de racionalidade e inteligência): a

² Devemos, neste ponto, separar entre problemas técnicos e filosóficos em torno da IA. Os primeiros se referem a como deve ser construída e projetada uma máquina para exibir um comportamento que qualificariam como inteligente. Essa pergunta pode ser tecnicamente respondida por qualquer abordagem conforme as categorias descritas. Por outro lado, os problemas filosóficos se referem a questões ontológicas, epistêmicas e conceituais que orbitam os fundamentos, as conclusões e experimentos mentais ou empíricos das pesquisas em IA.



capacidade de performar jogos de linguagem. É a partir do Teste de Turing que podemos nos indagar com direito: será o comportamento observado suficiente para denotar algo a respeito do funcionamento interno do sistema? Será a linguagem suficiente para indicar inteligência? O que significa dizer que, porque usam a linguagem como nós, as máquinas pensam? O que diríamos de nós, caso concluíssemos que máquinas podem pensar?

Diante disso, dependendo de como respondemos a tais questões, temos duas vias alinhadas com os enfoques das categorias descritas por Russell e Norvig (2016) e suas respectivas consequências: IAs não são meros algoritmos implementados para produzir um determinado efeito, mas são de fato sistemas pensantes dotados de estados cognitivos; ou se não há nada de especial em IAs capazes de comunicação como a nossa, é necessário repensar o estatuto da linguagem, o que há de diferente entre meu uso e o do *ChatGPT*, por exemplo.

Por isso, tendo em vista a variedade de problemas filosóficos, nosso artigo se dedica a investigar os fundamentos e as consequências filosóficas dos *Testes de Linguagem* para IAs, mas que também traz implicações para nós seres humanos. Primeiramente, faremos uma breve exposição de noções filosóficas associadas à relação entre linguagem e *natureza humana* em Heráclito e Aristóteles, bem como entre linguagem e mundo em Platão, com o objetivo de mostrar que as discussões levantadas por nosso artigo remontam a uma intuição clássica, que ganha mais força no pensamento de René Descartes. O pensamento do filósofo francês será fundamental para o desenvolvimento de nossa argumentação, porque ele é explícito em defender como a capacidade linguística tem sido considerada pela tradição filosófica o *marco* da humanidade. Com isso, pretendemos expor de que modo a intuição de Turing está alinhada com a de algumas noções filosóficas na história. Depois, na seção seguinte, tratamos dos *Testes de Linguagem* de Turing, Descartes e John Searle, apresentando seus princípios, suas motivações e algumas razões pelas quais seus resultados podem nos confundir quando tentamos responder a pergunta ignorada por Turing: *máquinas podem pensar?*⁹ Por último, avaliamos as consequências filosóficas do que foi desenvolvido e concluímos que a IA, em seu estágio técnico atual, abala um dos *marcos* tradicionalmente reconhecidos para a qualidade de humano: a linguagem, pois a capacidade linguística não é mais uma prerrogativa humana. O que deveremos fazer em um futuro não muito distante para dizer: "não sou um robô"?



1 A LINGUAGEM: MARCO DE HUMANIDADE

O termo *marco(s)* pode ser compreendido, de maneira geral, em dois sentidos: como propriedade intrínseca essencial; ou como um indicador externo com implicações ontológicas.

Alinhado com o primeiro, Tim Crane (1998) nos esclarece o uso do termo quando ilustra que, para Jean-Paul Sartre, é da natureza da consciência ser intencional, de tal modo que, se deixar de ser intencional, deixará de ser uma consciência (Sartre, 1995). Além do filósofo francês, Franz Brentano também tem uma tese sobre a intencionalidade como o *marco* do mental neste sentido, porque ela é uma característica exclusiva dos fenômenos mentais, não havendo manifestação física do que chamava intencionalidade (Brentano, 2014). Aqui a intencionalidade é intrínseca e remete ao que é propriamente mental.

Baseado nisso, mas alinhado com o segundo sentido de *marco*, Pernu (2017) propõe uma nova abordagem ao problema mente-corpo à luz dos "cinco *marcos* do mental": intencionalidade, consciência, livre-arbítrio, teleologia e normatividade. Ele explica que os *marcos* que propõe devem ser considerados como sinais ou sintomas da mentalidade. Em outras palavras, caso algo possua os *cinco marcos* (indicadores de caráter verificacionista), ali deve ter uma mente que os produziu.

Neste artigo, para nossos propósitos, falaremos de *marco* como se referindo a condições suficientes para indicar a presença ou existência de algo que não pode em si mesmo ser verificado por acesso direto. É mais forte do que um indício, pois funciona como um sinal não apenas epistêmico, mas também metafísico-ontológico, a partir do qual, visualizados certos efeitos, temos a implicação de que uma determinada causa certamente os precede. A fim de compreender o uso da expressão no contexto da nossa discussão interdisciplinar entre as Filosofias da Mente, da Linguagem e a Inteligência Artificial, a leitora deve considerar que, quando falamos de *marco(s)* do mental, referimo-nos às condições suficientes para que possamos concluir válida e corretamente que em determinada situação há uma mente e em outra não.

Dito isso, é importante destacar que, neste trabalho não discutiremos a intencionalidade ou as propostas de *marcos do mental* na literatura. Não faremos uma argumentação embasada na história. É suficiente, para os nossos propósitos, compreender o uso do termo, pois ele será empregado na elaboração da nossa hipótese de que a linguagem (ou pelo menos a capacidade de performar jogos de linguagem) tem sido tomada como *marco* de um traço da humanidade na medida em que tem sido considerada, de certa forma, um traço do mental por alguns pensadores.



1.1 Logos, razão e linguagem: perspectivas da antiguidade

Nesta subseção, trataremos brevemente dos pensamentos de Heráclito, Aristóteles e Platão, com o intuito de corroborar nossa hipótese, por meio de uma tradição filosófica que associa a linguagem e o pensamento à pretensa *natureza humana*. A linguagem, neste tópico, é dita no sentido amplo: capacidade de organizar, comunicar e expressar o pensamento³.

Enrique Hülz (2013) explica que o termo grego *logos* envolve noções da linguagem, do conhecimento, da realidade e da ação, bem como está permeado por um senso comum da racionalidade. Segundo o autor, não se deve reduzir *logos* à meras questões linguísticas (West, 1971; Johnstone, 2014), porque Heráclito utiliza a expressão com referência a diversas coisas (Johnstone, 2014), inclusive para fundamentar (*grounds*) a ideia de *natureza humana* na dimensão metafísica do *logos* (Hülz, 2013).

Alguns indícios disso estão nos fragmentos B1 e B 114 (Hülz, 2013). Hülz (2013) apoia-se neles para alegar que, de acordo com o filósofo antigo, o termo se refere ao aspecto racional da realidade (em sentido tanto ontológico como epistemológico), de tal maneira que seria justo dizer que *logos* se refere ao princípio do ser. Essa interpretação é reforçada por Hussey (1982), consoante o qual, para Heráclito, o *logos* expresso é uma declaração da essência. De acordo com o filósofo de Éfeso, há uma unidade vital entre ser e discurso, palavra e conteúdo, motivo pelo qual Hülz (2013) conclui que *logos* é a linguagem do real.

Já em Aristóteles, o *logos* está no centro da atitude política, pois é ele que produz ações que levam ao que é benéfico e justo (Trott, 2010). Por isso o *logos* é o fundamento da famosa definição do ser humano como animal político por natureza. Isso é assinalado na *Política* (Pol. 1253a9-18), onde Aristóteles explica que a razão pela qual o ser humano é um ser político em sentido pleno, enquanto outros animais não o são, é óbvia: ele possui a palavra, o discurso (*logos*). Os outros animais têm voz, são capazes de expressar prazer ou sofrimento, porém, nós, mediante o *logos*, conseguimos "tornar claro o útil e o prejudicial e, por conseguinte, o justo e o injusto" (Pol. 1253a9-18). Nisso consiste a diferença entre a mera voz (que todos os animais possuem) e o *logos* humano para Aristóteles (Aygün, 2018).

Até este ponto, vimos que o *logos*, seja exclusivamente humano, como é para Aristóteles, ou parte do universo, como para Heráclito, está intimamente ligado à nossa capacidade linguística de organizar, comunicar e expressar o pensamento. A linguagem nos permite manifestar o *logos*: esta capacidade é, para ambos os autores, característica da natureza humana. Contudo, é importante

³ Optamos por trabalhar com esta caracterização ampla aqui, pois caracterização semelhante de linguagem é fornecida pela literatura mencionada. Para nossos propósitos, nesta subseção, isso é suficiente.



destacar que linguagem e *logos* não se confundem, pois este não se reduz ao mero discurso. É na medida em que o discurso se apresenta de acordo com algum conteúdo do conhecimento (Heráclito) ou uma forma de pensamento visando uma finalidade (Aristóteles) que ele manifesta o *logos*. Embora isso implique na não identificação entre linguagem e *logos*, o uso da linguagem pode ser compreendido como *marco* da *natureza* humana, porque é nosso traço distintivo usá-la para expressar o *logos*.

Essa conclusão e os tópicos que evocaram ambos os pensamentos nos remetem à tensão existente entre linguagem e realidade no *Crátilo*, diálogo platônico que problematiza a atividade de nomear. Nesse texto, Hermógenes e Crátilo discutem, tendo Sócrates como mediador-provocador, acerca da qualidade dos nomes de serem meramente convencionais (Crat. 383a-386e), ou naturalmente determinados pela adequação com o modo de ser daquilo a que se referem (Crat. 386e-390e).

O que nos interessa destacar com isso é somente a tensão revelada pelo texto de Platão entre a linguagem (e seus usos) e a realidade, que pode remeter às noções heraclítica e aristotélica do *logos*. Ela nos permite vislumbrar a densidade dos problemas filosóficos que orbitam a linguagem: será que as palavras carregam ou revelam aspectos ontológicos daquilo que nomeiam? Será que quem nomeia, por meio da linguagem, tem algum tipo de acesso ao ser das coisas? Se sim, poderia, pelo uso da linguagem, destacar os *marcos* do que existe? Será que a linguagem denota aspectos ontológicos acerca daquele que fala (se é humano ou possui *logos*)? Será ela um *marco* da humanidade? É possível pensar *logos* separado da linguagem? E linguagem separado de *logos*? Ao que tudo indica, não é o mero discurso linguístico que marca a qualidade de humano, mas o *logos*. Porém, isso só pode ser averiguado linguisticamente. Então, como reconhecer que alguém fala por meio do *logos*?

1.2 Linguagem e alma: a perspectiva moderna cartesiana

Nesta subseção, trataremos do raciocínio do filósofo francês René Descartes, a fim de explicitar de que modo a linguagem humana adquire um estatuto ainda mais especial, porque denota que no corpo existe alma. Com esse intuito, passaremos por pontos específicos de suas fisiologias, teoria das paixões, dúvida metódica e do seu dualismo de substâncias. Aqui, a linguagem deve ainda ser compreendida no sentido amplo, como anteriormente. A distinção entre sintaxe e semântica, bem como a discussão contemporânea em torno desse assunto, não entram em jogo nas assertivas cartesianas.

Na primeira parte de *As Paixões da Alma*, Descartes fala das paixões em geral visando a natureza do ser humano, a partir da distinção entre mundo físico e mundo espiritual (mental), que fundamenta o raciocínio do autor e culmina na capacidade linguística como *marco* de humanidade.



Descartes (1973, pp. 227-228) explica que se algo que ocorre em nós e acontece igualmente em corpos inanimados (sem espírito, sem mente), então isso deve ser atribuído exclusivamente ao mundo físico; e que, ao contrário, aquilo que existe em nós, mas não pode ser atribuído somente ao corpo, deve pertencer à alma. Seguindo esse raciocínio, ele afirma que o corpo, bem como todo o mundo material, é regido pelas leis e princípios da física, do movimento e da geometria, enquanto o espírito tem seu domínio próprio e independente. Por isso o filósofo francês conclui que não é possível "que o corpo pense de alguma forma" e que "toda espécie de pensamento que existe em nós pertence à alma". Este é um aceno ao seu dualismo entre *res cogitans* (substância pensante) e *res extensa* (substância material).

A separação entre corpo e alma é alcançada por Descartes no projeto epistemológico apresentado nas *Meditações Metafísicas* (1973), cujo princípio é a dúvida como método. Na *primeira meditação*, o filósofo estabelece a dúvida hiperbólica por meio do princípio metafísico do *gênio maligno*, capaz de enganar irrestritamente (p. 99).

Entretanto, na *segunda meditação*, Descartes nota que, ainda que haja um *gênio maligno* a lhe enganar absolutamente, a proposição "*eu sou, eu existo*", é necessariamente verdadeira todas as vezes que a enuncio ou que a concebo em meu espírito" (1973, p. 100). Dessa primeira certeza, amplamente conhecida como *cogito ergo sum* (eu penso, eu existo), após uma reflexão sobre a natureza do seu ser, o filósofo conclui que é uma coisa essencialmente pensante (1973, p. 103): o ser humano existe como *res cogitans* (substância pensante). A partir desse ponto, o autor, com o puro raciocínio, faz distinções acerca do que pode dizer a respeito de seu próprio espírito num movimento auto reflexivo (meditação), que culmina em outra certeza: a alma é mais fácil de conhecer do que o corpo (1973, p. 106)⁴. Os motivos para isso estão relacionados ao contraste entre o acesso direto e imediato que temos dos conteúdos dos pensamentos conscientes e a incerteza acerca daquilo que é externo, material e depende dos sentidos para ser conhecido⁵. Isso vale para todos os objetos, incluindo o corpo.

Por essa razão, a existência das coisas materiais somente é alcançada na *sexta meditação*, após a fundamentação de um plano epistemológico de ideias claras e distintas assegurado pela ideia de Deus (*terceira meditação*). Contudo, para chegar a essa conclusão, Descartes, por conta do raciocínio que desenvolve nessa obra, deve se comprometer com outra tese: a distinção real entre alma e corpo, ou

⁴ A independência epistêmica entre alma e corpo, baseada em uma noção de transparência dos estados mentais ao sujeito que os possui conscientemente, será usada por Descartes para afirmar a sua independência ontológica.

⁵ Esse raciocínio cartesiano e a assimetria entre os planos de conhecimento do "interno" e do "externo" serão motivo da confusão entre os planos ontológico e epistêmico nos *Testes de Linguagem*. Por isso, a discussão acerca das *Meditações* é necessária: nosso argumento não dependerá apenas do dualismo, e sim da dinâmica entre a metafísica e a epistemologia cartesianas, cuja lógica, quando aplicada aos *Testes de Linguagem*, leva, irremediavelmente, ao solipsismo - ou pelo menos ao *Problema das Outras Mentes*. Agradecemos ao avaliador anônimo pela observação.

entre espírito e matéria, e suas respectivas propriedades. Ele defende que, tendo conhecido com certeza a si próprio como algo cuja natureza é de essência unicamente pensante, segue-se o fato de ter uma ideia clara e distinta de si como algo inextenso e que, portanto, o corpo, é uma outra coisa, algo extenso que não pensa (Descartes, 1973, p. 142)⁶. Isso implica que "minha alma, pela qual eu sou o que sou, é inteira e verdadeiramente distinta de meu corpo e que ela pode ser ou existir sem ele" (Descartes, 1973, p. 142). A existência do corpo é concebida apenas pela plausibilidade da inferência de que para sensações e faculdades corpóreas deve haver uma causa de mesmo tipo, física, material, extensa e corpórea (Descartes, 1973, p. 143).

Expostas as dimensões metafísica e epistêmica do dualismo cartesiano, mostraremos que elas são suficientes para estabelecer um *Teste de Linguagem* como meio seguro para determinar se os animais ou os autômatos poderiam pensar como nós e para resolver o *Problema das Outras Mentes*, porque o sistema de Descartes habilita a linguagem como *marco* da razão, algo exclusivo e necessário à natureza humana.

Na primeira parte do *Discurso do Método*, o filósofo afirma que aquilo que distingue os seres humanos dos animais é a capacidade para razão. Nossos corpos, como *res extensa*, estão sob as leis da mecânica e têm os mesmos princípios fisiológicos, mas nós somos, como *res cogitans*, racionais. Logo, Descartes conclui que se houvesse máquinas com órgãos ou corpos como "de qualquer outro animal sem razão, não disporíamos de nenhum meio para reconhecer que elas não seriam em tudo da mesma natureza que esses animais" (1973, p. 68). Todavia, "se houvesse outras que apresentassem semelhança com os nossos corpos e imitassem tanto nossas ações quanto moralmente fosse possível, teríamos sempre dois meios muito seguros para reconhecer que nem por isso seriam verdadeiros homens" (Descartes, 1973, p. 68): (1) nunca poderiam usar as palavras como nós para declarar os pensamentos; e (2) nem seriam capazes de desempenhar as mesmas funções que nós, em geral, somos capazes de realizar.

Sobre (1), Descartes imagina que seja possível conceber que uma máquina, feita como uma boa réplica do corpo, profira poucas palavras e até mesmo diga algumas a respeito de sensações decorrentes de estímulos físicos. Ele supõe que, ainda que houvesse máquinas que, quando tocadas em uns pontos específicos, exibam comportamentos como perguntar por instruções, gritar de dor ou outras coisas semelhantes (Descartes, 1973, p. 68), não seria possível que elas pudessem rearranjar as palavras tão variadamente de modo a responder ao sentido de tudo o que lhe for dito em sua presença. Mas isso é algo que podem fazer "os homens mais embrutecidos" (Descartes, 1973, p. 68). O mesmo vale para o caso de animais como os papagaios, pois, apesar de poderem proferir algumas palavras, não podem

⁶ Aqui está: da independência epistêmica, segue-se a ontológica (conforme Descartes).

rearranjá-las de maneira complexa, a fim de dar testemunho do pensamento por meio do discurso (Descartes, 1973, p. 69). Por isso é um erro, ele afirma, confundir nosso uso da palavra com o dos animais ou autômatos, porque este está para "os movimentos naturais, que testemunham as paixões e podem ser imitados pelas máquinas assim como pelos animais" (Descartes, 1973, p. 69).

Assim, conclui Descartes, pode-se:

[...] conhecer a diferença existente entre os homens e os animais. Pois é uma coisa bem notável que não haja homens tão embrutecidos e tão estúpidos, sem exceção mesmo os insanos, que não sejam capazes de arranjar em conjunto diversas palavras, e de compô-las num discurso pelo qual façam entender seus pensamentos; e que, ao contrário, não exista outro animal, por mais perfeito e felizmente engendrado que possa ser, que faça o mesmo (Descartes, 1973, p. 69).

Já a respeito do meio (2) para reconhecer a razão em outrem, o filósofo diz acreditar que autômatos e animais possam desempenhar algumas tarefas melhor do que os seres humanos, porém, isso não tem grandes implicações, pois falhariam invariavelmente para outras e o relógio também marca as horas com mais precisão do que nós (Descartes, 1973, p. 69). Isso seria suficiente para concluir que esses corpos agem conforme a disposição de seus órgãos, e não pelo seu conhecimento, pois as máquinas realizam ações e apresentam disposições particulares. Assim, seria necessário construí-las para inúmeras e muito diferentes situações, o que Descartes julga impossível (1973, p. 69). Diferentemente, os seres humanos dispõem da razão para reger suas ações e seus comportamentos, o que justifica nossa capacidade de desempenhar múltiplas e diversificadas funções (Descartes, 1973, p. 69).

O filósofo francês é muito claro ao defender a relação estreita entre linguagem e razão: porque é expressão do pensamento, denota racionalidade, uma característica essencial e distintiva do ser humano como *res cogitans*⁷. Mesmo pessoas nascidas surdas ou mudas são capazes de se fazer entender por meio de sinais que criam, pois desenvolvem uma linguagem (no sentido amplo) própria. A palavra, enquanto expressão racional, para Descartes, é testemunho da alma (mente) humana e, portanto, um

⁷ Um opositor hipotético poderia alegar: "Isso parece contrariar a hipótese em discussão pelo autor de que a linguagem é a marca da racionalidade humana, dado que aqui ela é apenas parasitária ao pensamento, sendo a *res cogitans*, e não a linguagem, a marca distintiva do humano. Podemos ainda assumir que sendo a linguagem uma expressão física (seja falada ou escrita) deveria ser regida por aquelas leis materiais da *res extensa*". A ele responderíamos: não contraria, pois, em Descartes, no tocante à *res cogitans*, não há cisão entre pensamento e linguagem. Além disso, não devemos confundir a mecânica da fala (âmbito da *res extensa*) com a linguagem entendida no sentido expresso pelo filósofo francês (algo exclusivo da *res cogitans*): é um erro "confundir as palavras com os movimentos naturais, que testemunham as paixões e podem ser imitados pelas máquinas assim como pelos animais" (Descartes, 1973, p. 69). Descartes defende claramente que somente a *res cogitans* é capaz expressar o pensamento por meio da linguagem, de tal modo que, para ele, um *Teste de Linguagem* (e a existência de Deus) é suficiente para resolver o *Problema das Outras Mentes*. Agradecemos ao avaliador anônimo pela observação.



marco daquilo que é qualificado como humano⁸ (1973, p. 70). Por isso, um Teste de *Linguagem* é, para ele, suficiente para determinar quando se trata de ser humano ou animal (ou autômato).

No entanto, há algo que deve ainda aqui ser destacado: nas *Objeções e Respostas às Meditações*, o filósofo francês alega que a alma, mediante a reflexão e a meditação sobre o que ela é, pode experimentar que pensa, mas não experimentar se os animais têm pensamento ou não, por isso, "nada disto pode descobrir senão quando, examinando as operações dos animais, remonta dos efeitos às causas" (Descartes, 1973, p. 194). Isso significa que a clareza e a distinção do conhecimento acerca de si próprio como uma *coisa pensante* são garantidas pelo acesso direto e privilegiado que temos acerca dos nossos próprios conteúdos mentais, uma espécie de transparência. Já no caso dos animais, somente podemos observar suas operações, seus comportamentos, e, a partir desses efeitos, inferir as causas. Por que motivos isso não vale para as outras pessoas? Por que o *Problema das Outras Mentes* pode ser superado pela aplicação de um *Teste de Linguagem*? Para Descartes, porque é impossível que uma máquina faça o mesmo uso da linguagem que nós fazemos. Segundo ele, isso é suficiente para estabelecer a linguagem como *marco* de humanidade.

Isso, contudo, leva-nos a um outro problema, cujas elaboração e solução serão tratadas na seção seguinte a respeito dos *Testes de Linguagem*.

1.3 Linguagem e intencionalidade: a perspectiva contemporânea de John Searle

Para finalizar nossa exposição de uma tradição de modo de pensar acerca da linguagem e de uma suposta *natureza humana*, trataremos, nesta subseção, da proposta do filósofo contemporâneo estadunidense John Searle, porque, embora pareça desassociar a linguagem, enquanto objeto, de qualquer aspecto ontológico daquele que fala, reintroduz o problema do *marco*, que problematizaremos na seção seguinte. Nesta subseção, a linguagem pode ser compreendida em suas dimensões sintática (puramente formal/simbólica) e semântica (intencional e acerca do conteúdo).

⁸ A racionalidade é propriedade intrínseca essencial da *res cogitans* e isso é, sem dúvida, um aspecto especial do humano. Com base nisso, um opositor hipotético poderia alegar: "Isso não parece correto, pois se a palavra revela a alma (*res cogitans*), não seria a alma o marco do humano? Não estamos aqui tomando o veículo pela mensagem veiculada?" A ele responderíamos: segundo Descartes, expressar o pensamento por meio da linguagem é uma capacidade exclusiva da substância pensante, logo, a linguagem funciona como *marco* do humano. Como reconhecer uma *res cogitans* como nós? Pelo uso da linguagem e da comunicação. É válido destacar que o *marco*, tal qual o definimos, é um indicador ontológico da presença de uma causa oculta inacessível, mas determinada por meio de efeitos específicos observados. Assim, para o filósofo francês, a linguagem é o único veículo possível para a mensagem veiculada – e nisto consiste a sua identificação como *marco* da *res cogitans*. Agradecemos ao avaliador anônimo pela observação.



De acordo com Searle (2002, p. VII), a Filosofia da Linguagem é um ramo da Filosofia da Mente. Isso será fundamento da explicação de uma das características mais fundamentais da mente, a intencionalidade. Mas é válido ressaltar que não analisaremos nem reconstruiremos detalhadamente a teoria de Searle sobre a intencionalidade, apenas apresentaremos seu modo de raciocinar.

Assim, segundo o filósofo estadunidense,

[...] a capacidade dos atos de fala para representar objetos e estados de coisas no mundo é uma extensão das capacidades mais biologicamente fundamentais da mente (ou do cérebro) para relacionar o organismo ao mundo por meio de estados mentais [...]. Uma vez que os atos de fala são um tipo de ação humana e uma vez que a capacidade da fala para representar objetos e estados de coisas faz parte de uma capacidade mais geral da mente para relacionar o organismo ao mundo, qualquer explicação completa da fala e da linguagem exige uma explicação de como a mente/cérebro relaciona o organismo à realidade (Searle, 2002, pp. VII-VIII).

Ou seja, o mecanismo de funcionamento da linguagem somente pode ser completamente explicado se explicitarmos a ponte entre o nosso mundo privado dos estados mentais e aquele público da realidade compartilhada. Esta ponte é estabelecida por meio da intencionalidade.

A intencionalidade pode ser compreendida como direcionalidade e como representação. Sob a primeira perspectiva, Searle diz que "Intencionalidade é aquela propriedade de muitos estados e eventos mentais pela qual estes são dirigidos para, ou acerca de, objetos e estados de coisas no mundo" (Searle, 2002, p. 1). Se um estado *D* é intencional, existem respostas para as perguntas "a que se refere *D*? Em que consiste *D*? O que é um *D* tal que?" Por isso, o sentido geral da palavra é melhor expresso por seu termo em inglês, *aboutness* ("sobreidade"). Crenças, temores, esperanças e desejos são intencionais, todavia, nem todo estado mental o é. As sensações, por exemplo, não se referem a algo. Igualmente, a intencionalidade não coincide com a consciência. De acordo com Searle, apesar de a frase de Edmund Husserl (2006) "toda consciência é consciência de algo" sugerir o oposto, algum tipo de identidade entre intencionalidade e consciência, isso não acontece, porque o "de" neste caso é diferente daquele da intencionalidade propriamente dita. Nos estados mentais intencionais, há uma distinção entre o estado em si e aquilo a que ele está direcionado. Para ilustrar, poderíamos responder "ansiedade" para quando perguntados "está com sensação de que?" Isso é muito diferente de dizer "tenho medo de cobras". Este é um estado intencional; aquele, não. Porém, com o intuito de estabelecer a ponte entre a subjetividade e o mundo externo, Searle se concentra em outro ponto: desenvolver a intencionalidade como representação.

Isso é feito por meio de uma comparação com o modelo dos atos de fala, porque "os estados Intencionais representam objetos e estados de coisas no mesmo sentido de 'representar' em que os atos de fala representam objetos e estados de coisas" (Searle, 2002, p. 5). Não é nosso propósito destrinchar

a teoria de John Searle. Para o desenvolvimento de nossa argumentação, basta-nos entender que ele não sugere que a intencionalidade seja essencial e necessariamente linguística, mas o contrário: "a linguagem é derivada da intencionalidade, e não o oposto" (Searle, 2002, p. 8).

Vale salientar que o sentido de "representação" empregado por Searle nada tem de ontológico. Afirmar que uma crença representa suas condições de satisfação é o mesmo que dizer que um enunciado representa suas condições de satisfação. Proferir que uma crença constitui uma representação nada mais é do que exprimir que ela possui um conteúdo proposicional e um modo psicológico; que tal conteúdo determina um conjunto de condições de satisfação sob certos aspectos; que tal modo psicológico determina a direção de adequação do seu conteúdo proposicional (Searle, 2002, p. 16). Isso tudo, alega Searle, é explicado pela teoria dos atos de fala.

Diante disso, já que a representação de uma sentença deriva da intencionalidade da mente, esta é mais primitiva e fundamenta a sua teoria linguística. Nesse sentido, a Filosofia da Linguagem é um ramo da Filosofia da Mente e, consoante Searle, ambas dependem da ontologia do mental. Logo, podemos compreender que a linguagem, examinada enquanto objeto, não possui, intrinsecamente, intencionalidade. Considere um texto descritivo de um cenário, um mapa, ou um livro: nada disso possui intencionalidade intrínseca, nós, seres humanos, é que atribuímos aos símbolos desses objetos, à sua linguagem, a propriedade de ser sobre ou de representar alguma outra coisa. Por isso, a linguagem, enquanto objeto, desprovida de sentido (semântica), nada mais é do que uma série de símbolos ordenados conforme uma regra (sintaxe) e não possui qualquer capacidade, por si só, de denotar algo da *natureza humana*. Ao que tudo indica, a partir desse momento, a linguagem não é mais um *marco* ontológico do humano, esse papel é realizado pela intencionalidade. Esse pensamento, como veremos, será fundamental para o *Argumento do Quarto Chinês*, proposto por John Searle.

Contudo, como explicaremos na seção seguinte, essa proposta não supera a distância existente entre um efeito observado e sua causa oculta inacessível: como saber se algo que se comunica conosco possui intencionalidade intrínseca? Afinal de contas, somente posso observar a desenvoltura de um interlocutor em jogos de linguagem, mas não determinar o que, por meio dessa interação, causa seu comportamento linguístico.

2 OS TESTES DE LINGUAGEM: TURING, DESCARTES E SEARLE

No século XX, Alan Turing (1950) enunciou a questão "máquinas podem pensar?" Embora essa pergunta evoque problemas filosóficos, teóricos, científicos e até mesmo técnicos, seu proponente a considerou sem significado para merecer discussão. Ao invés disso, ele nos apresenta o *Jogo da*



Imitação. Esse experimento mental envolve três participantes: um interrogador de qualquer gênero sexual e dois interlocutores, de sexos opostos. O objetivo do primeiro é distinguir entre quem é homem e quem é mulher. O jogo deve ocorrer às cegas, de modo que o interrogador tem acesso a informações exclusivamente pelos dados de uma conversa digitada com seus interlocutores. Sabendo disso, agora devemos supor que um deles é substituído por uma máquina, um computador teórico (a princípio) adequadamente programado. Neste caso, "o interrogador decidirá errado com a mesma frequência quando o jogo for jogado dessa forma como quando o jogo for jogado entre um homem e uma mulher?" (Turing, 1950, p. 433). Dessa maneira, no lugar de "máquinas podem pensar?", devemos indagar "máquinas podem passar no *Jogo da Imitação*?"

Na época em que o elaborou, Turing esperava que, no futuro, com os avanços técnicos e tecnológicos, da engenharia e da ciência da computação, seria possível uma máquina capaz de enganar o interrogador. Caso isso acontecesse, não seria estranho que falássemos de máquinas pensando sem a expectativa de sermos contestados, porque elas seriam indistinguíveis linguisticamente de nós seres humanos. Teríamos aí uma Inteligência Artificial (IA).

Entretanto, em que sentido "*uma máquina pode obter uma taxa de sucesso razoável no Jogo da Imitação*?" substitui "*uma máquina pode pensar*"? Antes de responder diretamente a isso, é preciso destacar que, para nossos propósitos, os gêneros dos participantes do jogo são irrelevantes, pois, em consonância com a leitura de Copeland (2000), Piccinini (2000) e Moor (2001), o que Turing realmente tinha em mente era um teste no qual uma máquina (computador, programa, sistema) deve fingir, por meio da linguagem, ser um ser humano bem o suficiente para fazer com que o interrogador se engane ao determinar quem é humano e quem é máquina. Qualquer teste que possa ser compreendido nesses termos gerais, chamamos, neste artigo, *Teste de Linguagem*.

Feita essa consideração, e esclarecendo que, para nossos objetivos não importa a taxa de sucesso nos testes, podemos dizer que há duas interpretações possíveis da relação entre as duas perguntas feitas por Turing: (1) se uma máquina passar no jogo, então ela será inteligente; e (2) se uma máquina passar no jogo, então será razoável que a consideremos inteligente. Em ambas as concepções, há uma noção intuitiva de que um determinado uso da linguagem está associado à inteligência, tal como nos pensamentos filosóficos expostos na seção anterior. Todavia, em (1), temos uma conclusão com implicações ontológicas: não é o fato de o interrogador dizer que a máquina é uma pessoa que faz dela inteligente, mas é algo nela (em seu projeto, construção ou programação) que lhe concede a inteligência para obter sucesso no jogo. Porque é inteligente, a máquina passou no teste. Já em (2), a inteligência da máquina é reconhecida, percebida, verificada ou corroborada pelo sucesso no teste, o que não consiste em implicação ontológica, e sim epistêmica: se a máquina é ou não um ser inteligente antes de tudo não faz sentido, porque a conclusão que se obtém neste caso é pela via epistêmica da avaliação de

seu desempenho no jogo. Dessa forma, podemos dizer que os *Testes de Linguagem* têm: uma interpretação ontológica, segundo a qual o teste dá informações suficientes para inferir características intrínsecas de inteligência da máquina; e outra epistêmica, na qual não ocorre atribuição de propriedades inerentes, apenas se reconhece a capacidade da máquina de realizar uma atividade, sem implicações metafísicas com causas ocultas.

Como lidar, portanto, com uma IA bem sucedida? Qual interpretação é mais adequada? É válido aqui relembrar e seguir o raciocínio de Descartes na resposta dada a *como saber se os animais pensam?*: "nada disto pode descobrir senão quando, examinando as operações dos animais, remonta dos efeitos às causas" (1973, p. 194). Se fizermos conforme o filósofo francês sugere e tivermos como premissa seu sistema metafísico, devemos, dos efeitos, realizar inferências sobre as causas, para concluir, corretamente, que num corpo de animal ou de uma máquina, não existe alma (nem mente, razão ou inteligência). Entretanto, isso só é verdadeiro num mundo em que uma máquina não é capaz de se comunicar satisfatoriamente, nem mesmo como um "homem bruto", ou seja, no cenário em que máquinas falham no Teste de Turing. Caso contrário, quando um autômato for bem sucedido, seremos obrigados a concluir que: o princípio causador ou deve ser o mesmo para nós e a máquina e, portanto, máquinas pensam como nós; ou este princípio não pode ser conhecido, nem no caso da máquina nem no de outras pessoas.

Neste ponto, é proveitoso fazermos um pequeno desvio, pois tais questões levantadas pela interpretação *Teste de Turing* podem ser identificadas também na leitura de Descartes. Nas *Meditações Metafísicas*, ele enuncia aquilo reconhecido tradicionalmente como *Problema das Outras Mentes* (Avramides, 2023), mediante a pergunta: "que vejo desta janela, senão chapéus e casacos que podem cobrir espectros ou homens fictícios que se movem apenas por molas?" (Descartes, 1973, p. 105). Nesse texto ele está preocupado com a possibilidade da existência de seres humanos mecânicos, sem espírito (mente), que se comportem como nós ao ponto de enganá-lo. Apesar de, nesta obra, Descartes responder à própria indagação graças à sua simples capacidade de julgar para distinguir as máquinas de pessoas reais, no *Discurso do Método*, ele propõe a solução baseada na natureza distinta do ser humano como algo pensante: um *Teste de Linguagem*. Considerando, novamente a possibilidade de haver máquinas capazes de imitar nossos corpos e ações, ele defende que poderíamos distingui-las entre seres humanos reais graças à capacidade destes "de arranjar em conjunto diversas palavras, e de compô-las num discurso pelo qual façam entender seus pensamentos" (Descartes, 1973, p. 69). Contudo, como explicamos no parágrafo anterior, essa solução somente funciona caso não haja máquinas que performam nossos "jogos de linguagem". Para Descartes, quiçá devido ao seu contexto histórico, social, cultural e conceitual, fosse difícil imaginar tal cenário, mas, para Turing, não – quanto mais para nós, contemporâneos do *chatGPT*.



Dito isso, perguntamos novamente: qual interpretação é mais adequada? Adotar a interpretação ontológica do resultado positivo do Teste de Turing significa estabelecer uma relação essencial entre linguagem e pensamento, linguagem e razão, linguagem e a qualidade do que é ser humano, em alinhamento com a metafísica cartesiana. Assim, num mundo em que uma máquina possa passar no *"Teste de Descartes"*, ela pensa e deve ser apropriadamente chamada de Inteligência Artificial, porque o princípio causador do seu sucesso deve ser o mesmo que nos permite nosso uso da linguagem: a mente. Segundo a interpretação ontológica, passar no teste é suficiente para tal conclusão, porque a linguagem é *marco* do humano. Em tal cenário, opor-se a essa conclusão leva ao *Problema das Outras Mentes*, porque negar que outra mente possa ser reconhecida, epistemicamente, mediante um princípio metafísico que funciona como *marco*, resulta na impossibilidade de se reconhecer como humana qualquer coisa que se comunique como nós, mesmo outros seres humanos.

Talvez, por vislumbrar embaraços metafísicos desse tipo, Turing tenha orientado que, ao invés de questionar "*máquinas podem pensar?*", devêssemos perguntar "*uma máquina pode obter uma taxa de sucesso razoável no Jogo da Imitação?*" Isso parece nos indicar uma interpretação com menor carga metafísica, pois se refere mais à questão epistêmica: quais fatores corroboram para o interrogador dizer que conversa com uma pessoa, quando, na verdade, dialoga com uma máquina? Segundo por esta via da interpretação epistêmica, é possível que o resultado positivo de uma IA no *Teste de Turing* não seja suficiente para dizer que uma máquina é inteligente, embora seja razoável, porque ela realiza uma imitação perfeita.

Isso tudo nos remete ao artigo do filósofo estadunidense contemporâneo, John Searle, *Minds, brains and programs* (1980), no qual ele publica, pela primeira vez, seu famoso *Argumento do Quarto Chinês*.

A ideia central por trás do argumento de Searle é mostrar que "a IA no sentido forte deve ser falsa, uma vez que um agente humano poderia instanciar um programa e, mesmo assim, não ter estados mentais" (1980, p. 417). Mediante seu experimento mental, o filósofo pretende demonstrar que uma mera implementação algorítmica computacional não é, por si só, suficiente para a consciência ou a intencionalidade, já que ninguém que apenas manipule símbolos desconhecidos de outra língua diria, a respeito de si mesmo, que comprehende verdadeiramente o idioma, ainda que assim o parecesse para os observadores. Ademais, ele finaliza com a defesa de que é necessária a existência de uma base causal física orgânica capaz de produzir nossos fenômenos mentais.

No início de seu artigo, Searle apresenta uma distinção entre os sentidos fraco e forte de IA. Segundo ele, o primeiro se refere ao uso instrumental das teorias computacionais: elas são meios para promover o estudo e a reflexão acerca da mente, da cognição e da consciência, mas não tratam dessas entidades em si. E isso, diz-nos Searle, permite-nos ser mais rigorosos e precisos do que antes. Já no



sentido forte de IA: "o computador não é apenas um instrumento para o estudo da mente" (1980, p. 417). Para este filósofo, de acordo com os adeptos da IA forte, computadores *adequadamente programados* têm estados cognitivos. "*Adequadamente programados*" se refere a uma situação na qual uma máquina pode ser programada para relacionar um conjunto de *inputs* (variáveis de entrada de um sistema) a um de *outputs* (variáveis de saída de um sistema), exibindo um comportamento perfeitamente similar ao humano.

Além disso, de acordo com Searle (1980, p. 418), os partidários da IA forte não diriam que um programa como o de Roger Schank (1977) apenas simula uma habilidade humana, e sim que: (1) a máquina comprehende a história e fornece respostas às questões; (2) o que a máquina e seu programa fazem explica a habilidade humana para entender histórias e responder questões sobre elas.

Entretanto, John Searle considera que essas asserções são insustentáveis, pois a mera manipulação simbólica, ainda que cause um comportamento complexo e igual ao nosso, não será suficiente para implicar em compreensão, um estado intencional.

No *Argumento do Quarto Chinês*, um não falante de chinês seria capaz de relacionar uma série de símbolos cujo significado desconhece graças a uma série de regras no seu idioma nativo do tipo "se receber tal ou tais símbolos, então devolva tal ou tais símbolos". Assim, à medida que a pessoa trancada no quarto aperfeiçoasse sua manipulação simbólica e os programadores melhorassem seus algoritmos, do ponto de vista externo ao quarto, as respostas dadas por ela seriam indistinguíveis daquelas fornecidas por falantes nativos de chinês.

À vista disso, alguém que raciocine como os partidários da IA forte deveria dizer que o quarto, ou qualquer coisa que esteja escondida por essa "caixa-preta", comprehende chinês. Contudo, Searle se coloca no interior do quarto, na posição de quem manipula símbolos, e nos sugere fazer o mesmo. Sob essa perspectiva diríamos que compreendemos chinês? Consoante ele, não, porque obtivemos as respostas por manipulações formais, sem qualquer significação. Nós, a respeito do chinês, teríamos nos comportado como um computador, executando operações computacionais com base em elementos formalmente especificados. Por conseguinte, conclui Searle, "para os propósitos do idioma chinês, eu sou simplesmente uma instanciação de um programa de computador" (1980, pp. 417-18).

Sabendo disso, Searle nos recomenda a análise das duas asserções feitas pelos defensores da IA forte.

Acerca da primeira ele afirma: "parece óbvio [...] que eu não comprehendo uma palavra das histórias em chinês" (Searle, 1980, p. 418). Mesmo com a relação perfeita entre *inputs* e *outputs*, promovida por um programa adequado, não importa o quanto o filósofo pareça saber chinês, ele não entende nada do que fala. O paralelismo estabelecido pela estratégia argumentativa é claro:



Pelas mesmas razões, o computador de Schank não comprehende nada das histórias, sejam elas em chinês, em inglês, ou em qualquer outro idioma. No caso do idioma chinês, eu desempenho o papel do computador, e nos casos onde não desempenho tal papel, o computador não faz nada além do que eu poderia fazer, ou seja, em ambas as situações não há compreensão (Searle, 1980, p. 418).

Sobre a segunda ele diz: "podemos verificar que o computador e seu programa não fornecem as condições suficientes para a compreensão"; e se pergunta: "será que ele fornece uma condição necessária ou uma contribuição significativa para a compreensão?" (Searle, 1980, p. 418). Ora, se, somarmos a tese (1) da IA forte ao funcionalismo, poderíamos dizer que a compreensão, definida em termos funcionais, é a mesma coisa para nós e para um computador executando um programa, ou seja, uma relação entre *inputs* e *outputs*. Porém, em um caso – quando falamos nosso idioma – nós parecemos ter uma noção muito clara a respeito da nossa compreensão, e no outro – operando com símbolos desconhecidos – parecemos ter uma noção muito clara acerca do nosso estado de não compreensão, ou melhor, de que não ocorre qualquer ideia em torno do significado original dos símbolos. Logo, Searle diz que, apesar de não poder demonstrar a falsidade de (2), tudo leva a crer que ela o é. Com isso, pode-se notar que Searle responde diretamente à pergunta original de Turing, dizendo: as máquinas não podem pensar.

Finalmente, agora, depois das explicações dos *Testes de Linguagem* de Turing, Descartes e Searle, podemos esclarecer algo que é motivo de discordâncias em torno de todas essas propostas: uma associação entre a linguagem e a alma (mente) humana, que leva àquela entre os testes e o *Problema das Outras Mentes*. Trata-se de uma confusão entre os aspectos metafísico e epistêmico das configurações dos *Testes de Linguagem* já prenunciado por Descartes: se somos interrogadores tentando distinguir entre os participantes quem é máquina e quem é humano, o resultado do teste é em função de questões epistêmicas; mas se estamos tentando convencer outra pessoa de que somos seres humanos, nós sabemos que somos tais criaturas, e esta é uma afirmação de teor ontológico.

O que isso significa para os *Testes de Linguagem*? Como veremos, independentemente da resposta, temos consequências filosóficas relevantes acerca daquilo que consideramos como tão certo a nosso próprio respeito.

3 "NÃO SOU UM ROBÔ": O PROBLEMA FILOSÓFICO PRENUNCIADO PELOS TESTES DE LINGUAGEM

Para corroborar nossa hipótese, deve-se destacar que Turing (1950) prevê possíveis objeções à sua abordagem, dentre as quais a *Teológica* (pp. 443-444) e a *da Consciência* (pp. 445-447). A primeira,



é dita por Turing da seguinte maneira: "pensar é uma função da alma imortal do homem. Deus deu uma alma imortal a cada homem e mulher, mas não a nenhum outro animal ou máquina. Portanto, nenhum animal ou máquina pode pensar" (1950, p. 443). Já a segunda, ele atribui a Jefferson (1949) por seu manifesto: "não até que uma máquina possa escrever um soneto ou compor um concerto por causa de pensamentos e emoções sentidas, e não pela sorte de alguns símbolos, poderíamos concordar que máquina é igual a cérebro – isto é, não apenas escrever, mas saber que o escreveu." Com a mesma finalidade, deve-se ressaltar que Searle recebe o *Problema das Outras Mentes* como uma objeção ao seu *Argumento do Quarto Chinês*: "se atribuímos cognição a outras pessoas [unicamente por seu comportamento], devemos em princípio atribuí-la também a computadores" (1980, p. 421).

Curiosamente, o argumento de Searle tem um efeito inesperado: desfaz os pressupostos das objeções feitas a Turing. Se o argumento do filósofo estadunidense é válido e pungente, o funcionalismo, a IA forte e toda tese que se apoie em programas de IAs simbólicas, como o *SHRDLU* (Winograd, 1973) e o *ELIZA* (Weizenbaum, 1966), ou em modelos de linguagem de larga escala (*large language models - LLMs*), como o *ChatGPT*, qualquer simulação de fenômenos mentais humanos baseada na máquina de Turing deve ser falsa⁹. E, além disso, quer dizer que *Testes de Linguagem* não são suficientes nem essenciais a uma conclusão ontológica acerca da mente daquele que é testado, porque a linguagem, enquanto objeto acessível ao exame alheio, não remete, intrinsecamente, ao que é qualidade de ser humano. Dessa maneira, a linguagem pura (sintaxe) não é *marco* de humanidade, como sugerem as réplicas a Turing¹⁰.

Ademais, ambas as objeções, a *Teológica* e a *da Consciência*, são muito similares àquilo que pode ser extraído do raciocínio metafísico de Descartes. Para ele, já que o ser humano possui alma, pensar é função da alma e o pensamento é expresso pela linguagem, então: máquinas e animais não podem pensar, nem produzir textos (escritos, musicais, imagéticos). Mas esse argumento, de caráter ontológico, para ser aplicado à realidade, depende da verificação da capacidade de pensamento, feita por meio da linguagem, procedimento este que, por sua vez, é epistêmico. Neste caso, ainda que haja uma diferença, como explica Searle (1980, pp. 421-422) na resposta à *Objecção das Outras Mentes*, entre *como saber que outras pessoas (ou se outras coisas) têm estados cognitivos* e *o que estamos lhes atribuindo ao dizer que têm estados cognitivos*, a arquitetura dos *Testes de Linguagem* mistura as perspectivas metafísica e epistêmica.

⁹Aplicamos as conclusões de Searle sobre IA forte e fraca também às IAs baseadas em redes neurais profundas como *ChatGPT* e outros, porque seu argumento não perde validade em tais casos. No entanto, é importante destacar que a arquitetura das IAs da qual Searle debate no influente artigo de 1980 correspondem sobretudo às IAs simbólicas.

¹⁰ Isso está em consonância com a resposta de Searle (1980) à objeção *The brain simulator reply*.



Essa mistura tem duas implicações possíveis: (1) o solipsismo, porque somente habilita cada um a falar de si; ou (2) a possibilidade real da IA forte tanto quanto de nós sermos como computadores, porque estabelece uma simetria entre 1^a e 3^a pessoa. Vejamos como isso ocorre.

Considerando o *Quarto Chinês*, como um exemplo de *Teste de Linguagem*, temos duas situações: uma externa, de observador/interlocutor (análoga àquela de interrogador no de Turing); e outra interna, de manipulador dos símbolos desconhecidos (análoga a de participante do *Teste de Turing*). Na primeira, as dimensões metafísica e epistemológica são mais claramente distinguíveis, afinal de contas, é evidente que, por uma série de motivos que entram em jogo, objetivos (que se referem à observação do modo de se comunicar, presença ou ausência de "gírias", regionalismos, erros ortográficos) e subjetivos (que se referem aos aspectos particulares que constituem a subjetividade do interlocutor/interrogador, as experiências pessoais, seus sentimentos e as particularidades de sua história de vida), nós, nesta posição, podemos nos enganar. É possível, do lado de fora, dizer erroneamente que dentro do *Quarto Chinês* há um falante de chinês, tanto quanto podemos nos enganar ao dizer quem é humano e quem é máquina no *Teste de Turing*. Em tal configuração, tem-se uma conclusão de cunho epistêmico e nada se pode dizer ontologicamente acerca do que está no interior da outra sala ou quais as causas reais daquelas expressões linguísticas, pois acessamos unicamente os efeitos visíveis, como já havia apontado Descartes. Essa conclusão é apresentada também por Searle. Por outro lado, no cenário em que somos nós manipulando os símbolos, temos acesso epistêmico privilegiado e imediato àquilo que é causa do comportamento linguístico observado de fora: nossa pura manipulação simbólica (que sabemos correr sem qualquer compreensão). Ou, no caso do *Teste de Turing*, sabemos que nós somos o ser humano, e não máquinas. Essa condição epistêmica é considerada ontológica, tanto por Descartes (*cogito ergo sum*) como por Searle, em sua definição da ontologia de primeira pessoa (Searle, 1997; 2007). Assim, saber "o que sou" é uma certeza máxima, a conjuntura em que o epistêmico (o que sei de mim) e o ontológico (o que sou) se confundem.

O objetivo dos *Testes de Linguagem*, entretanto, não é estabelecer, metafisicamente, quando nós mesmos, em exame de auto reflexão, estamos conscientes, intencionais e raciocinamos acerca dos conteúdos daquilo que expressamos por meio da linguagem. Nas *Meditações*, Descartes sugeriu uma maneira de chegar a essa conclusão. Com os *Testes de Linguagem*, por outro lado, aquilo que desejamos nos perguntar, seja "as máquinas podem pensar?", seja "máquinas podem passar no jogo da imitação?", é a respeito de outra coisa, algo diferente de nós mesmos, algo que só pode ser inferido a partir dos efeitos observados.

Dito isso, podemos explicitar a confusão produzida pela configuração dos *Testes de Linguagem*: não podemos aplicar os mesmos critérios que usamos para falar de nós mesmos (1^a pessoa)



ao julgamento acerca de outras coisas (3^a pessoa). Por isso, os *Testes de Linguagem* evocam uma dimensão na qual o *Problema das Outras Mentes* é absolutamente relevante e inevitável.

Searle (1980) parece discordar, porque não julga que a *Objeção das Outras Mentes* seja válida à sua proposta. Todavia, embora nos pareça que seu argumento seja apenas sobre a manipulação simbólica e a semântica, devemos recordar que esta é conferida à linguagem dos seres humanos pela intencionalidade, cujo fundamento, de acordo com Searle (1980, p. 417), está na mente e em sua respectiva base causal, o sistema nervoso. É compreensível que, se desejamos definir a mente, o fazemos de maneira teórica, mediante afirmações metafísicas, por exemplo. Mas, o que faríamos ao encontrar uma nova criatura, que age tal como se tivesse estados intencionais, inclusive exibindo um comportamento linguístico similar ao nosso? Será plausível que a pergunta "este ser pensa?" seja respondida por uma vivissecção ou ressonâncias magnéticas? Se, em algum momento, fizermos este tipo de questão, precisaremos de um *teste de intencionalidade*. E, ao que tudo indica, os problemas dos *Testes de Linguagem* igualmente se mantêm, pois a intencionalidade searleana é tão acessível quanto a alma cartesiana.

O filósofo estadunidense poderia rejeitar nosso apontamento pelo mesmo motivo que rejeitou o *Problema das Outras Mentes*: há uma diferença entre *como saber que outras pessoas (ou se outras coisas) têm intencionalidade* e *o que estamos lhes atribuindo ao dizer que têm intencionalidade*. Apesar disso ser verdade, existe um obstáculo não percebido por ele: a definição daquilo que atribuímos a outra pessoa (ou máquina, ou animal, ou coisa), ou seja, da intencionalidade, da mente ou do pensamento, é dada a partir de uma via de suposta máxima certeza em que os aspectos epistêmicos e metafísico se confundem. Em 1^a pessoa, "o que sei de mim" se confunde com "o que sou". Searle está correto em dizer que os *Testes de Linguagem* não são condições suficientes para uma conclusão ontológica acerca da mente daquele que é testado, porém pelos motivos errados, pois ele somente pode falar acerca de si próprio em uma condição na qual seus estados mentais lhe são transparentes. Quando saímos do *Quarto*, estamos novamente na condição de meros observadores, momento em que nossas afirmações sobre algo submetido a um teste são, principalmente, epistêmicas, pois a mente ou a intencionalidade alheias não podem ser determinadas por vias seguras. Não se segue do fato de que "sei o que sou" que "eu sei quando algo é como eu", porque existe uma assimetria característica dos *Testes de Linguagem* que não pode ser superada pela mera passagem de uma perspectiva à outra. Aqui temos um aceno ao solipsismo.

Nossa resposta ao filósofo é ratificada pela tréplica de Turing à *Objeção da Consciência*, porque ela pode ser traduzida nos termos da intencionalidade que Searle (1980) defende. Segundo o matemático e cientista da computação esse modo de pensar parece ser, antes de tudo:



[...] uma negação da validade do nosso teste. De acordo com a forma mais extrema desta visão, a única maneira pela qual alguém poderia ter certeza de que uma máquina pensa é ser a máquina e sentir a si mesmo pensando. Alguém poderia então descrever esses sentimentos para o mundo, mas é claro que ninguém estaria justificado em tomar conhecimento. Da mesma forma de acordo com esta visão, a única maneira de saber que um homem pensa é ser aquele homem em particular (Turing, 1950, p. 446).

E o solipsismo é, de certo modo, uma outra versão, ainda mais radical, do *Problema das Outras Mentes*.

Uma maneira de evitar essa consequência indesejada é aceitar que possamos fazer afirmações de valor ontológico acerca de outras coisas com base em nossa auto reflexão e naquilo que podemos observar de seu comportamento, mesmo que suas causas estejam ocultas. Nesse sentido, poder-se-ia sustentar a mistura entre as perspectivas ontológica e epistêmica. Porém, isso seria como dizer que analogias não são apenas metáforas, mas têm valor epistemológico-ontológico, pois nos permite dizer, em sentido forte, que "*algo que observo e considero se comportar como eu, é uma coisa de mesmo tipo*". Adotar essa postura, por um lado, tem uma grande consequência para a Inteligência Artificial, porque seria suficiente que uma máquina ou um programa realizasse funções que consideramos típicas da "inteligência humana" (se comunicar como nós, por exemplo) para concluirmos corretamente: sim, máquinas podem pensar, falar ou ter intencionalidade. Mas, por outro lado, temos também uma grave conclusão: fica cada vez menos claro o que significa ser humano - é ser racional, ser capaz de expressar os pensamentos, parecer ter intencionalidade? Se a máquina pode ser inteligente, não é óbvio por qual motivo não podemos operar segundo programas.

Não obstante, um opositor hipotético poderia sustentar que é possível defender a fusão entre as vias metafísica epistêmica na arquitetura dos *Testes de Linguagem*, porque (1) a via epistêmica é suficiente para revelar a estrutura metafísica/ontológica do sistema e (2) a via epistêmica não é suficiente, mas como não temos outra, só o que temos são "predições" baseadas em comportamentos¹¹. Contudo, mostramos que (1) é falso, pois o resultado do teste não é suficiente para revelar a estrutura ontológica do sistema (aqui concordamos com Searle). Além disso, apesar de (2) ser uma proposta com menos comprometimento metafísico, leva-nos a um dos cenários montados por nossas conclusões: ou ao *Problema das Outras Mentes* ou ao solipsismo, porque não saberíamos e nem poderíamos dizer quando algo tem mente ou não; ou ao estabelecimento de uma simetria entre o processo do pensamento e a execução de algoritmos por computadores adequadamente programados.

Se esse posicionamento não agradar, portanto, frente ao risco do solipsismo, resta uma opção: assumir que o *logos*, a razão, a capacidade de comunicação, o sucesso em *Testes de Linguagem* e a intencionalidade, bem como outros *marcos* tão imediata e intuitivamente aceitos como de humanidade,

¹¹ Agradecemos ao avaliador anônimo por esta observação.



na verdade, não o são. De um jeito ou de outro, encarar um *Captcha (Completely Automated Public Turing test to tell Computers and Humans Apart* - Teste de Turing público completamente automatizado para diferenciar computadores de humanos) é angustiante, pois é mais obscuro do que imaginávamos dizer se sou ou "não sou um robô".

CONSIDERAÇÕES FINAIS

Na *Introdução* deste artigo, anunciamos que mostrariíamos que uma certa competência linguística, identificada pelo desempenho em um *Teste de Linguagem*, ainda não seria suficiente para o reconhecimento de uma inteligência como uma propriedade intrínseca. Vejamos explicitamente como isso foi alcançado no decorrer do nosso trabalho.

Na *Primeira Seção*, mostramos que, para os antigos, por exemplo, a linguagem se relaciona intimamente com *logos* na definição de 'o que é ser humano'. Mas René Descartes vai além, pois, diante da dúvida metódica, antes de invocar a existência de Deus como princípio metafísico para o conhecimento, há somente uma certeza máxima: "*eu sou uma coisa pensante*". Como vimos, em seu dualismo, pensar é uma atividade da alma (mente), jamais do corpo. Mas o que garante que outras pessoas não sejam autômatos? O fato de máquinas e animais não serem capazes de expressar seus pensamentos, porque isso é, para ele, *marco* da alma humana. Assim, segundo Descartes, a linguagem é uma capacidade inerente e exclusiva do ser humano derivada de sua natureza particular. Por último, expusemos a proposta de John Searle, para quem a linguagem, enquanto objeto, não possui nada de intrinsecamente especial, logo, não é, ela mesma, *marco* de humanidade. A intencionalidade, na sua visão, desempenha esse papel. No entanto, o problema de Descartes pode ser transposto para sua abordagem, pois não é claro como podemos inferir que a exibição de qualquer capacidade ou função consiste num *marco* de intencionalidade. Em suma, há um problema sério em estabelecer *marcos* da *natureza humana* por meio da observação do comportamento linguístico, porque ou consideramos *a priori* que nada mais pode possuí-lo, ou não podemos saber quando ele é intencional ou não.

Na *Segunda Seção*, baseados na forma geral no *Jogo da Imitação* de Alan Turing, apresentamos a noção de *Testes de Linguagem*. Com isso, fomos capazes de identificar outros testes do mesmo tipo em Descartes e em Searle. Ademais, explicamos que há duas interpretações possíveis no caso de uma máquina obter sucesso em um *Teste de Linguagem*: uma ontológica, que atribui ao participante uma propriedade intrínseca (inteligência, intencionalidade, racionalidade); e outra epistêmica, que não se implica com nada além do fato de reconhecer uma capacidade comportamental análoga à do ser humano na máquina. Essa distinção fica mais evidente quando analisamos o *Argumento do Quarto*



Chinês. Aceitamos o que defende John Searle, os *Testes de Linguagem* não são suficientes nem essenciais a uma conclusão ontológica acerca da mente daquele que é testado, porque a linguagem, enquanto objeto acessível ao exame alheio, não remete, intrinsecamente, ao que é qualidade de ser humano. Porém, aqui temos um percalço: o único cenário em que isso ocorre é quando fazemos afirmações acerca de nós mesmos em nossos pensamentos. A máxima certeza, tanto para Searle como para Descartes, é alcançada no plano em que os aspectos ontológico e epistêmico estão misturados, quando "*o que sou*" se confunde com "*o que sei de mim*". A confusão ocorre ao se passar de uma interpretação a outra como se suas validades tivessem o mesmo grau. Isso leva, inevitavelmente, ao *Problema das Outras Mentes*.

Na *Terceira Seção*, mostramos que as conclusões que fazemos a nosso respeito são de qualidade muito diferente daquela que fazemos sobre outras. No primeiro caso, temos acesso das razões ou das causas pelas quais se fala ou se faz determinadas coisas, pois, ao que tudo indica, temos acesso epistêmico direto à nossa consciência, o que nos permite estabelecer conclusões ontológicas fortes. Já no segundo, não é uma consequência lógica que, da observação de um comportamento ou desempenho em *Teste de Linguagem*, haja ali uma entidade com determinadas propriedades intrínsecas. Nesta situação, entretanto, a conclusão negativa, de que não há ali uma entidade com determinadas propriedades intrínsecas, também não é possível. Por isso John Searle está correto: *Testes de Linguagem* não são suficientes para dizer que uma máquina é uma Inteligência Artificial no sentido forte. Mas também não são suficientes para dizer qualquer outra coisa além de "eu sei quando penso, comprehendo ou tenho estados intencionais", porque não nos dá acesso à base ontológica que fundamenta o comportamento linguístico.

Todavia, a conclusão a que chegamos é uma inclinação ao solipsismo, o qual se configura como um caso mais radical do *Problema das Outras Mentes*. Se quisermos evitá-la, temos de nos comprometer com a interpretação ontológica dos *Testes de Linguagem* e, por conseguinte, com a possibilidade de existência da Inteligência Artificial no sentido forte. Caso essa postura não agrade, teremos de assumir que não comprehendemos claramente o que são prerrogativas humanas, sejam eles intencionalidade, pensamento ou sucesso no *Teste de Turing*.

Assim, ao refletir sobre os *Testes de Linguagem* no tocante à Inteligência Artificial, além da conclusão negativa que apresentamos acerca deles, fomos levados a uma outra, essencialmente filosófica e contraintuitiva: talvez não saibamos com máxima certeza o que somos. No nosso mundo, onde as IAs têm grandes capacidades generativas de texto (*ChatGPT*), de imagens (*DALL-E*, *Midjourney*), de vídeo (*Sora*), bem como de áudio e música (*Suno*), é cada vez mais difícil dizer quando qualquer um desses objetos é produzido por uma máquina. Nesse contexto, como vimos, é um erro insistir nas estratégias para fazer distinção ontológica entre máquina e ser humano com base apenas no



conteúdo que fornecem, porque não há qualquer teste para *marco* de humanidade confiável baseado nisso. Alguém pode rejeitar nossa conclusão, alegando que podemos dizer quando uma imagem é feita por IA, já que basta procurar por padrões repetidos ou mãos com muitos dedos para se dizer que é um produto gerado por programa. Contudo, não podemos deduzir que será impossível um computador, adequadamente programado e construído, produzir conteúdos perfeitamente qualificáveis com algum "selo de qualidade humana."

Dessa forma, retornamos à nossa conclusão derivada do nosso argumento acerca da insuficiência dos *Testes de Linguagem*: é possível uma Inteligência Artificial tanto quanto sermos máquinas executando programas. E, se isso nos incomodar, podemos suspendê-la, mas temos de assumir que aquilo que tradicional e intuitivamente acreditamos e temos por tão certo do ser humano, está profundamente abalado. Nossa reflexão sobre a IA nos diz que não sabemos o que é preciso para dizer "não sou um robô".

Artigo recebido em: 14/09/2024

Artigo aceito em: 14/01/2025

Artigo publicado em: 31/03/2025



REFERÊNCIAS

- ARISTÓTELES. *Política*. Tradução e notas de Antonio de Castro Caeiro. Edição bilíngue. São Paulo: Martins Fontes, 2001.
- AVRAMIDES, A. The sceptic, the outsider, and other minds. *Topoi*, Springer Verlag, v. 42, n. 1, p. 175-186, 2023. DOI: <https://doi.org/10.1007/s11245-022-09827-9>.
- AYGÜN, Ömer. Human logos in Aristotle. In: KIRKLAND, Sean D.; SANDAY, Eric (orgs.). *A companion to ancient philosophy*. Evanston, Illinois: Northwestern University Press, 2018. (Rereading ancient philosophy), pp. 177-191.
- BRENTANO, F. *Psychology from an Empirical Standpoint*. London: Taylor & Francis Ltd - Routledge, 2014. Originally published in 1874.
- BRINGSJORD, S.; GOVINDARAJULU, N. S. Artificial Intelligence. In: ZALTA, E. N.; NODELMAN, U. (Ed.). *The Stanford Encyclopedia of Philosophy*. Summer 2024. [S.l.]: Metaphysics Research Lab, Stanford University, 2024.
- COPELAND, J. The Turing test. *Minds and Machines*, v. 10, n. 4, p. 519-539, 2000. DOI: <https://doi.org/10.1023/A:1011285919106>.
- CRANE, Tim. *Contemporary issues in the philosophy of mind*. Cambridge: Cambridge University Press, 1998.
- DESCARTES, René. Os Pensadores: *Discurso do método; Meditações; Objeções e respostas; As Paixões da Alma; Cartas*. Tradução de J. Guinsburg e Bento Prado Júnior. São Paulo: Abril Cultural, 1973.
- HÜLSZ, Enrique. Heraclitus on logos: language, rationality and the real. In: SIDER, David; OBBINK, Dirk (ed.). *Doctrine and doxography: studies on Heraclitus and Pythagoras*. Berlin; Boston: De Gruyter, 2013, pp. 281-302.
- HUSSERL, E. *Ideias para uma Fenomenologia Pura e uma Filosofia Fenomenológica*: Introdução Geral à Fenomenologia Pura. Revista e ampliada. São Paulo: Ideias & Letras, 2006.
- HUSSEY, Edward. Epistemology and meaning in Heraclitus. In: SCHOFIELD, Malcolm; NUSSBAUM, Martha Craven (ed.). *Language and logos: studies in ancient Greek philosophy presented to G. E. L. Owen*. Cambridge: Cambridge University Press, 1982, pp. 33-61.
- JEFFERSON, G. The mind of mechanical man. *British Medical Journal*, v. 1, n. 4616, p. 1105-1121, 1949. Disponível em: <http://www.jstor.org/stable/25372573>. Acesso em: 07 jan. 2025.
- JOHNSTONE, Mark A. On 'Logos' in Heraclitus. In: INWOOD, Brad (ed.). *Oxford studies in ancient philosophy*. V. 47. Oxford: Oxford Academic, 2014, pp. 1-29. Online edn, 17 nov. 2016. DOI: <https://doi.org/10.1093/acprof:oso/9780198722717.003.0001>.
- MOOR, James H. The status and future of the Turing test. *Minds and Machines*, v. 11, n. 1, p. 77-93, fev. 2001. DOI: <https://doi.org/10.1023/A:1011218925467>.
- PERNU, Tuomas K. The five marks of the mental. *Frontiers in Psychology*, v. 8, 07 jul. 2017. DOI: <https://doi.org/10.3389/fpsyg.2017.01084>.
- PICCININI, Gualtiero. Turing's rules for the imitation game. *Minds and Machines*, v. 10, p. 573-582, nov. 2000. DOI: <https://doi.org/10.1023/A:1011246220923>.
- PLATÃO. *Teeteto e Crátilo*. Tradução de Carlos Alberto Nunes. Belém: Universidade Federal do Pará, 1988.
- RUSSELL, S.J.; NORVIG, P. *Artificial intelligence: a modern approach*. Malaysia: Pearson Education Limited, 2016.



- SARTRE, Jean-Paul. *The psychology of the imagination*. London: Routledge, 1995. Originalmente publicado em 1940.
- SCHANK, R. C.; ABELSON, R. P. *Scripts, Plans, Goals, and Understanding*. Hillsdale, N.J.: Lawrence Erlbaum Press, 1977.
- SEARLE, J. R. Minds, brains, and programs. *Behavioral and Brain Sciences*, v. 3, n. 3, p. 417-424, 1980. DOI: <https://doi.org/10.1017/S0140525X00005756>.
- SEARLE, J. R. Reductionism and the irreducibility of consciousness. In: FLANAGAN, O.J.; BLOCK, N.; GUZELDERE, G. (ed.). *The Nature of Consciousness*. [S.l.]: MIT Press, 1997, pp. 69-80.
- SEARLE, J. R. Dualism revisited. *Journal of Physiology*, v. 101, n. 4-6, p. 169-178, 2007.
- SEARLE, J. R. *Intencionalidade*. São Paulo: Editora Martins Fontes, 2002.
- TROTT, Adriel M. Logos and the political nature of *anthrōpos* in Aristotle's *Politics*. *Polis*, v. 27, n. 2, p. 292-307, 2010. Disponível em: <https://ssrn.com/abstract=3369901>. Acesso em: 07 jan. 2025.
- TURING, A. M. Computing machinery and intelligence. *Mind, New Series*, v. 59, n. 236, p. 433-460, 1950. DOI: <https://doi.org/10.1093/mind/LIX.236.433>.
- WEIZENBAUM, J. Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, v. 9, p. 36-45, 1966. DOI: <https://doi.org/10.1145/365153.365168>.
- WEST, Martin L. *Early Greek philosophy and the Orient*. Oxford: Clarendon Press, 1971.
- WINOGRAD, T. A procedural model of language understanding. In: SCHANK, R.; COLBY, K. (Ed.). *Computer Models of Thought and Language*. San Francisco: W. H. Freeman, 1973, pp. 152-186.

