

SEARLE E OS DESAFIOS DA INTELIGÊNCIA ARTIFICIAL (IA) FORTE

Joelma Marques de Carvalho¹

ORCID iD: [0000-0001-6421-7864](https://orcid.org/0000-0001-6421-7864)

Resumo: O objetivo principal deste artigo é analisar quais são os desafios que a inteligência artificial (IA) forte deve superar conforme as críticas feitas por John Searle (1932*). As duas principais teses que Searle apresenta contra a IA forte são: (i) a semântica não é uma parte intrínseca da sintaxe e (ii) a sintaxe de uma linguagem não é uma propriedade intrínseca dos objetos físicos. Esses argumentos serão explicitados por meio de suas teorias da linguagem, da mente e da realidade social. No final, veremos que os argumentos searleanos são suficientes para demonstrar que um programa computacional qualquer não é capaz de gerar mentes como as nossas, porém isso não significa que a tese de que computadores digitais poderão um dia possuir estados mentais como propriedades emergentes seja impossível do ponto de vista lógico. No entanto, os desafios que os idealizadores da IA forte enfrentam são difíceis, visto que nossas mentes não possuem apenas uma estrutura sintática e formal, mas também um conteúdo semântico e a capacidade de compreender esse conteúdo.

Palavras-chave: Searle; IA; Naturalismo Biológico; Mente; Linguagem.

Abstract: The main objective of this article is to analyze which challenges strong artificial intelligence (AI) must overcome according to the criticisms made by John Searle (1932*). The two main theses that Searle presents against strong AI are: (i) semantics is not an intrinsic part of syntax and (ii) the syntax of a language is not an intrinsic property of physical objects. These arguments will be made explicit through his theories of language, mind, and social reality. In the end, we will see that the Searlean arguments are sufficient to demonstrate that any computer program is not capable of generating minds like ours, but this does not mean that the thesis that digital computers may one day possess mental states as emerging properties is impossible from a logical point of view. However, the challenges that the creators of strong AI face are difficult, since our minds do not only have a syntactic and formal structure, but also a semantic content and the ability to understand this content.

Keywords: Searle; IA; Biological Naturalism; Mind; Language.

1. Inteligência Artificial (IA)

Desde que Alan Turing (1912-1953) publicou o seu artigo “Computing Machinery and Intelligence”, em 1950, ele passou a ser considerado o pai da inteligência artificial (IA). Nesse artigo, Turing elabora um experimento mental, no qual um ser humano (a) interage com uma máquina (b). Porém, um observador externo (c) não é capaz de identificar qual deles, (a) ou (b), é um ser humano ou uma máquina. Nesse contexto, esse autor sugere que se uma máquina é capaz de passar nesse teste, então essa máquina é inteligente, posto que ela é capaz de pensar. Sendo assim, isso nos mostra que há fortes indícios de que a existência da inteligência artificial (IA) é possível.

¹ Doutora em Filosofia pela Ludwig-Maximilians-Universität (LMU) em Munique na Alemanha. Atualmente é Pesquisadora Sênior no Curso de Filosofia na Universidade de Salzburgo na Áustria com a bolsa Lise Meitner Stipendium. E-mail: joelma_marques@yahoo.com.br

Contudo, parte-se da ideia de que uma máquina é inteligente se ela apenas puder simular um comportamento inteligente. Por esse motivo, a inteligência artificial (IA) foi muito criticada desde o seu início e um desses críticos foi o filósofo John Searle. Porém, para compreendermos suas críticas devemos primeramente distinguir dois tipos de Inteligência Artificial (IA), a saber: (i) *IA forte* e (ii) *IA fraca*. Conforme a IA forte (i), um computador digital “qualquer” pode instanciar estados cognitivos, ou seja, ter uma mente. De acordo com a IA fraca ou cautelosa (ii), uma máquina ou um computador é apenas um instrumento fundamental para melhor entendermos como a mente humana funciona.

Searle não questiona a IA fraca (i), mas apenas a IA forte (ii). A rigor, ele defende duas teses, a saber: (a) a semântica não é uma parte intrínseca da sintaxe e (b) a sintaxe de uma linguagem não é uma propriedade intrínseca dos objetos físicos. A primeira tese (a) está contida em seu experimento mental do quarto chinês e será discutida no segundo, terceiro e quarto tópicos desse artigo. A segunda tese (b) foi defendida em seu livro “O Mistério da Consciência” (1998) e será abordada no quinto tópico.

2. O argumento do quarto chinês

Antes de examinarmos o experimento mental ou argumento searleano do quarto chinês, devemos inicialmente compreender os três níveis básicos da linguagem, a saber: (a) *sintaxe*, (b) *semântica* e (c) *pragmática*. A sintaxe (a) é a parte da linguagem que estabelece a relação formal entre os signos de uma determinada linguagem, isto é, ela trata das regras constitutivas da linguagem. Por exemplo, ao afirmarmos que a palavra “impressionante” contém “ss” e não “c”, estamos esclarecendo algo que faz parte da sintaxe da língua portuguesa. Em geral, regras gramaticais são regras sintáticas. A semântica (b) trata dos significados literais dos termos da linguagem. A pragmática (c) refere-se à relação entre os usuários da linguagem. Imaginemos que um indivíduo A questione a um indivíduo B se esse último sabe que horas são e o indivíduo B responda apenas: “sim”. Nesse caso, podemos afirmar que o sujeito B considerou apenas o nível semântico da linguagem e ignorou o nível pragmático da mesma, visto que, nesse contexto, espera-se que um falante responda não apenas se ele sabe ou não do horário, mas que nos diga exatamente que horas são.

Com base nessas informações, analisemos agora então o experimento do quarto chinês reapresentado por Searle em sua obra “Mentes, Cérebros e Programas” (1980). Imaginemos que um sujeito A nativo da língua inglesa encontra-se em um quarto fechado. Nesse quarto, ele

recebe três calhamaços, a saber: (i) um calhamaço com um texto em chinês, (ii) um calhamaço com regras em inglês, as quais lhe ensinam como fazer a correlação entre os textos (i) e (ii), e (iii) um calhamaço com regras em inglês que lhe orientam a fazer a correlação entre os textos (i), (ii) e (iii). Esse último calhamaço (iii) é entregue aos indivíduos externos, os quais são falantes nativos da língua chinesa.

Esses calhamaços são denominados por esses indivíduos respectivamente de: (i) roteiro ou programa, (ii) história ou perguntas e (iii) respostas às questões. Nessa situação, a partir do ponto de vista dos falantes chineses, o indivíduo A é capaz de compreender as perguntas (ii) em inglês postas por eles e responder essas questões através de (iii) de modo tão satisfatório que os chineses diriam que o indivíduo A é capaz de compreender também a língua chinesa, ou seja, aquilo que está contido no calhamaço (i). Contudo, Searle afirma que o sujeito A apenas manipula símbolos formais tal como um computador e, por isso, não é correto dizer que ele compreende a língua chinesa, mas apenas que ele se comporta como tal. Desse modo, esse autor pretende demonstrar que o domínio da sintaxe de uma linguagem não é suficiente para o domínio de sua semântica. Em outras palavras: Uma vez que um computador abarca apenas a sintaxe da linguagem, ele não é capaz de compreender os significados dos termos da linguagem.

A partir desse experimento mental, muitos concluíram que, sendo assim, um programa de computador que se comporta como sendo uma mente humana, não é uma mente, ou seja, um sistema que possui estados cognitivos. Isso significa que apesar de uma máquina ser capaz de passar no teste de Turing, não seria correto dizer que essa máquina é inteligente e que é capaz de pensar. Contudo, Searle afirma que seu experimento mental do quarto chinês não tenta demonstrar que uma máquina não é inteligente nem que talvez um dia será capaz de pensar.

Com base no quarto chinês, é possível perceber que esse autor rejeita a tese da IA forte de que os programas de um computador são idênticos às mentes humanas, uma vez que nossas mentes são capazes de compreender todos os níveis da linguagem, já os programas de um computador abarcam apenas um conjunto de regras sintáticas. Os argumentos que formam a base desse experimento mental são resumidos pelo próprio Searle (SEARLE, 1998, p. 38) do seguinte modo:

1. Programas são totalmente sintáticos.
 2. As mentes têm uma capacidade semântica.
 3. A sintaxe não é a mesma coisa que a semântica, nem é, por si só, suficiente para garantir um conteúdo semântico.
- Consequentemente, programas não são mentes.

Segundo esse autor, a conclusão de que programas computacionais não são idênticos às nossas mentes pode ser rejeitada caso uma das premissas desse argumento seja falsa. A premissa (1) refere-se à definição de “programa” aceito por Turing. A segunda premissa (2) afirma que nossas mentes são capazes de compreender os significados dos termos de uma língua. A terceira premissa (3) parte da ideia de que apenas a manipulação de símbolos não é suficiente para a aquisição da compreensão dos significados e referências das palavras e frases de uma língua. Por esses motivos, esse filósofo rejeita a tese da IA forte, mas aceita a tese da IA fraca.

Para Searle, não faz sentido afirmar que um computador compreende uma língua, pois a noção de “compreensão” está relacionada com a nossa capacidade de interpretação de significados dos termos ou sentenças num determinado contexto. Vejamos adiante em que consiste exatamente essa capacidade semântica segundo esse filósofo, ou seja, de que modo o significado e a referência das palavras são determinados por seus usuários e compreendidos pelos mesmos. Desse modo, entenderemos melhor a posição searleana diante da IA forte e como ela está interligada com outras teses defendidas por esse filósofo em suas teorias da linguagem, da mente e da realidade social.

3. Como ocorre a determinação e compreensão do significado?

O diálogo é um tipo de ação humana. Quando duas pessoas conversam, elas realizam geralmente diferentes atos de fala. Para explicitar como a linguagem humana funciona através desses atos de fala, Searle retoma algumas distinções feitas por J.L. Austin (1962) tais como: (i) *atos ilocucionários* e (ii) *atos perlocucionários*. Os atos ilocucionários (i) constituem a menor unidade de comunicação e devem ser realizados geralmente de modo intencional. Já os atos perlocucionários (ii) são os efeitos causados pelos atos ilocucionários (i) e não devem ser necessariamente intencionais. Por exemplo: Se um indivíduo A pergunta para o indivíduo B:

(1) “Você vai ao concerto com essa roupa?”

A sentença (1) é um ato ilocucionário e o efeito desse ato de fala poderia ser, por exemplo, convencer o indivíduo B a ir com outra roupa ou simplesmente deixá-lo com raiva. Esses efeitos podem existir independentemente se o indivíduo A através da pergunta (1) teve a intenção de causar tais efeitos ou não. As consequências decorrentes diante do ato ilocucionário

(i) constituem assim os atos perlocucionários (ii). De acordo com Searle (SEARLE, 2000, p.127): “O fato de os atos ilocucionários serem essencialmente intencionais, enquanto os atos perlocucionários poderem ou não ser intencionais, é uma consequência do fato de o ato ilocucionário ser a unidade de significado na comunicação”.

Na visão searleana, o significado é um tipo de intencionalidade. Por “intencionalidade” entende-se a capacidade que a mente possui de se dirigir ou representar fatos no mundo. Nesse âmbito, devemos diferenciar dois tipos de intencionalidade, a saber: (a) *intencionalidade intrínseca* e (b) *intencionalidade derivada*. A intencionalidade intrínseca (a) é uma propriedade de estados mentais. Essa característica é denominada “intrínseca”, posto que sua existência independe de nossas opiniões sobre a mesma. Por exemplo: O indivíduo A tem sede e deseja matar sua sede não com água comum, mas bebendo água com gás. Nesse caso, o estado mental do desejo de matar a sede que o sujeito A experencia tem como objeto intencional uma água com gás. A propriedade “ser com gás” é um aspecto intencional desse estado mental. Isso significa que nossos estados intencionais possuem condições de satisfação que são determinadas pelo próprio sujeito. Mas observem também que os estados intencionais de sentir sede e desejar saciar a sede existem independentemente de qualquer opinião que tenhamos acerca desses estados psicológicos e poderiam existir inclusive mesmo que o indivíduo A não fosse capaz de se comunicar por meio da linguagem. Em outras palavras: Alguns estados mentais intencionais podem ser fenômenos anteriores à linguagem.

A intencionalidade derivada (b) é uma propriedade da linguagem que provém da intencionalidade intrínseca dos falantes. Tal característica não é intrínseca, pois sua existência depende da opinião do falante e outros observadores. Isso significa que quando falamos ou escrevemos, nós impomos nossa *intencionalidade intrínseca* para as frases. Imaginemos agora que o indivíduo A vá até um quiosque e peça a um vendedor:

(2) “Dê-me uma garrafa de água com gás, por favor!”

A sentença (2) é um ato ilocucionário. Por meio desse proferimento, espera-se como efeito o ato perlocucionário da venda ou entrega de uma água com gás. O ato ilocucionário (2) possui uma intencionalidade derivada dos estados mentais do indivíduo A. Nesse contexto, Searle também distingue dois tipos de significado: (i) o *significado literal* e (ii) o *significado do falante*. O significado literal ou usual (i) das palavras é determinado através dos significados dos termos e pelo modo como tais palavras são organizadas, isto é, pelas convenções

linguísticas num determinado contexto. Contudo, o significado do falante (ii) é determinado pelas intenções do falante diante de certos limites. Tomemos como exemplo uma ironia. Imaginemos que o preço da água com gás seja tão caro que o indivíduo A afirme ao vendedor:

(3) “Esta água deve ser milagrosa”

Nesse contexto, o significado literal (i) da sentença (3) depende apenas das convenções linguísticas, mas aquilo que o sujeito A quer dizer através da sentença (3), isto é, o significado do falante (ii), é determinado por suas intenções. Nesse caso, percebe-se que o indivíduo A por meio do proferimento (3) teve a intenção de significar o contrário do significado literal ou usual (i) dessa sentença. No entanto, isso não significa que o sujeito A possa afirmar uma sentença qualquer e significar o que ele quiser, pois as convenções da linguagem determinam certos limites para esse tipo de significado. Mesmo assim, o significado do falante (ii) é o modo primário de significado e nos fornece o caminho para compreendermos o funcionamento da linguagem. Nas palavras de Searle:

A chave para a compreensão do significado é a seguinte: o significado é uma forma de intencionalidade derivada. A intencionalidade original ou intrínseca do pensamento do falante é transferida para palavras, frases, marcas, símbolos e assim por diante. Se pronunciadas de forma significativa, essas palavras, frases, marcas e símbolos passam a ter uma intencionalidade derivada dos pensamentos do falante. Eles não têm apenas um significado linguístico convencional, mas também um significado desejado pelo falante. A intencionalidade convencional das palavras e frases de uma língua pode ser usada por um falante para realizar um ato de fala. Quando um falante realiza um ato de fala, ele impõe sua intencionalidade àqueles símbolos (SEARLE, 2000, p. 131).

Um dos conceitos fundamentais para entendermos como essa transferência de intencionalidade da mente para a linguagem ocorre é, segundo Searle, as *condições de satisfação*. Como já foi dito: Os nossos estados mentais possuem condições de satisfação. De modo análogo, as sentenças possuem condições de verdade. Retomemos o exemplo da sentença (2). Se o vendedor entregasse uma água sem gás em vez de água com gás ao sujeito A, o indivíduo A poderia até matar sua sede, porém isso não satisfaria uma das condições de satisfação do seu desejo que é saciar a sede com água gaseificada. Isso significa que nós impomos condições de satisfação aos proferimentos que fazemos, os quais, por sua vez, já possuem condições de satisfação.

Um ato de fala possui diversas condições de satisfação, visto que, em geral, quando falamos temos diferentes tipos de intenção. Nos casos das sentenças (1), (2) e (3), os falantes tiveram primeiramente (i) a intenção de proferir essas sentenças e (ii) a intenção de que essas frases possuam determinadas condições de satisfação. Essa última é também denominada “intenção de significado”. Além dessas duas intenções, os falantes têm, em geral, (iii) a intenção de que seus proferimentos sejam compreendidos pelos ouvintes, ou seja, que eles possuem uma *intenção de comunicação*. Desse modo, para que a comunicação seja bem sucedida, é necessário que os ouvintes reconheçam esse conjunto de intenções.

Contudo, para Searle, a compreensão do significado e referência dos termos ou sentenças de uma linguagem envolvem ainda outros elementos que iremos abordar agora: a rede (*network*) e o pano de fundo (*background*). Nós não experienciamos um estado mental de modo isolado, pois eles fazem parte de uma rede de outros estados mentais. Por exemplo: O indivíduo A sente sede, possui o desejo de matar sua sede, tem a intenção de comprar uma garrafa de água, sente calor, tem a crença de que o preço da água é caro etc. Essa rede de estados mentais é relevante para a determinação de condições de satisfação dos estados mentais, pois as condições de satisfação de um estado intencional não são determinadas isoladamente, mas em relação à rede de estados intencionais, a qual todo estado mental faz parte. Por esse motivo, o reconhecimento de pelo menos uma parte da rede de estados intencionais é fundamental para a compreensão de atos ilocucionários de um falante. Retomemos o último exemplo da ironia. O indivíduo A fez uma ironia por meio da sentença (3) para expressar seu estado mental de indignação. Esse estado mental faz parte de uma rede de outros estados mentais. Ele pressupõe um conjunto de crenças tais como, por exemplo, a crença de que essa água custa menos em outros estabelecimentos, de que a existência de águas milagrosas é improvável etc.

Além da rede de estados intencionais, Searle defende que os elementos da rede de estados intencionais necessitam ainda de um pano de fundo (*background*) para que a rede funcione da forma como geralmente atua, posto que os elementos que fazem parte da rede não são auto-interpretativos (SEARLE, 1997, p. 251). Por “background” entende-se o conjunto de capacidades não representacionais, disposições, atitudes, formas de comportamento etc. Existem dois tipos de *background*: (i) *background profundo* e (ii) *background local*. O *background profundo* (i) é constituído pelas capacidades que possuímos em comum devido nossas características biológicas, como por exemplo, andar, comer, perceber etc. Já o *background local* (ii) é formado por nossos conhecimentos e capacidades ou práticas adquiridas

em uma determinada cultura tais como beber cerveja em um copo, comer com talheres, ir a um restaurante etc.

Searle apresenta vários argumentos a favor da existência do *background* e de sua ligação com a rede de estados intencionais que, a rigor, faz parte do pano de fundo, mas irei tratar aqui apenas alguns argumentos que são fundamentais para a compreensão de suas teses utilizadas para criticar a posição da IA forte de que programas computacionais são mentes. Em sua teoria da linguagem, ele defende que a rede e o *background* são necessários para a compreensão do significado literal das frases e de metáforas. Tomemos os seguintes exemplos utilizados pelo próprio Searle para explicitar sua posição (SEARLE, 1999, p.189):

(4) O presidente abriu a reunião

(5) A artilharia abriu fogo

De acordo com sua opinião, o conteúdo semântico do termo “abrir” nas frases (4) e (5) é o mesmo, mas as condições de satisfações dessa palavra são diferentes nesses casos, visto que abrir uma reunião não é o mesmo que abrir fogo. Isso significa que a compreensão do significado dos termos e sentenças exige mais do que a mera apreensão do significado literal de suas expressões. Analisemos ainda a seguinte sentença:

(6) O indivíduo A é amargo

No caso da frase (6) temos uma metáfora com a estrutura X é Y e podemos compreender essa metáfora mesmo se não existir uma semelhança entre X e Y. Em geral, sabemos que o termo “amargo” não significa nesse caso o contrário de “doce”, mas sim “rancoroso”. Esse tipo de conhecimento seria considerado por Searle como um saber prático que faz parte do *background* local e cultural de quem domina a língua portuguesa. Em outras palavras: Para Searle, a compreensão linguística não consiste apenas em apreender o significado literal ou conteúdo semântico dos termos e um conjunto de crenças pressupostas. Ela só funciona em virtude de um saber prático cultural e biológico que ele denomina “background” (Searle, 1999, p. 192).

Essa noção de *background* é uma expansão das noções wittgensteineanas de “quadro de referência” e “jogos de linguagem”. Para o segundo Wittgenstein (WITTGENSTEIN, 1996), uma palavra só faz sentido dentro de um jogo de linguagem e os jogos de linguagem só

funcionam dentro de um quadro de referência, ou seja, como sendo parte de nossas atividades sociais e culturais que foram desenvolvidas historicamente conforme leis físicas. Por exemplo: A sentença:

(7) A sanfona pesa 11 kg

A sentença (7) faz sentido dentro de um jogo de linguagem que existe porque objetos como um acordeão não mudam de peso o tempo todo. Em um mundo possível, no qual tais objetos mudassem de peso constantemente, o jogo de linguagem do peso não faria sentido. De modo análogo, algumas expressões podem fazer sentido numa determinada cultura, noutras não. Tomemos como exemplo a expressão alemã:

(8) Viel Glück!

A expressão (8) “viel Glück” significa “boa sorte” em português e geralmente pode ser proferida para um indivíduo que, por exemplo, irá fazer uma prova importante no Brasil. Contudo, essa expressão foi considerada como descabida por uma japonesa que conheci. Ela tinha uma prova de alemão que era fundamental para o seu doutorado na Alemanha, pois caso ela não tivesse êxito nessa prova, ela perderia sua bolsa de estudo e não poderia fazer seu doutorado. Porém, ao lhe desejar boa sorte um dia antes da prova, ela disse que não me compreendia. Daí eu traduzi a expressão (8) para o inglês e ela me explicou que sabia o significado literal dessa expressão alemã, mas que essa expressão não fazia sentido naquele contexto, visto que ninguém tem êxito em provas devido a sorte. Nesse caso, o que eu deveria lhe desejar era:

(9) Seja estudiosa!

Isso significa que a expressão (8) pode fazer sentido num *background* local e noutro não. Para Searle, cada indivíduo tem o seu *background*. Isso explica também porque algumas piadas podem ser difíceis de serem compreendidas por alguns indivíduos, visto que as piadas são geralmente bem culturais e exigem às vezes um *background* local que nem todos possuem.

Outro argumento a favor da existência do *background* é que para interpretarmos o significado de uma expressão, nós precisamos inicialmente do conhecimento de certas regras.

Contudo, com a prática linguística é possível que um indivíduo compreenda e possa fazer certos proferimentos mesmo sem ser capaz de explicar explicitamente as regras que ele está seguindo. Além disso, a interpretação de uma regra *x* não pode ser ela mesma representacional, pois caso contrário seria necessário termos uma outra regra *y* para interpretarmos a regra *x* que, por sua vez, para ser interpretada exigiria uma regra *z* e isso nos levaria a um regresso ao infinito. Nesse sentido, proferir um ato ilocucionário e compreender proferimentos são comportamentos análogos a esquiar ou dirigir um carro e pressupõem as capacidades e disposições de um *background* não representacional.

Desse modo, a afirmação de que um computador é capaz de compreender uma determinada linguagem faria sentido na visão searleana se os programas computacionais tivessem uma rede de estados intencionais e um *background* tanto local quanto profundo que foram constituídos biologicamente ao longo da história. Isso significa que a capacidade semântica que possuímos é, para Searle, uma capacidade biológica e mental. Não é a toa que, em sua visão, a filosofia da linguagem é compreendida como um ramo da filosofia da mente (SEARLE, 1999, p. 206). Vejamos agora mais detalhadamente sua tese de que programas não são mentes no experimento mental do quarto chinês através de sua noção “mente” a partir de teoria da mente.

4. O que significa dizer que programas não são mentes?

Para compreendermos a tese defendida no experimento searleano do quarto chinês de que programas não são mentes, devemos analisar também o que esse autor entende por “mente” em sua posição denominada de “naturalismo biológico”. Os termos “naturalismo” e “biológico” decorrem dos fatos de que a mente e suas propriedades são fenômenos biológicos e fazem parte do mundo natural tanto quanto a digestão, a mitose etc.

De acordo com Searle, a mente possui as seguintes características: (i) *consciência*, (ii) *subjetividade ou ontologia de primeira pessoa*, (iii) *intencionalidade intrínseca* e (iv) *causalidade mental*. A consciência (i) é uma propriedade física e biológica de seres humanos e animais, a qual é causada por processos cerebrais de nível mais elevado. Estados conscientes são essencialmente internos, qualitativos e subjetivos (SEARLE, 2000, p.45). Eles são internos, visto que ocorrem no cérebro e fazem parte de uma rede de outros estados mentais. Eles são qualitativos na medida em que há um modo de serem experienciados pelos sujeitos. Nesse sentido, há uma diferença qualitativa entre beber um vinho e ouvir uma música. Por fim, eles

são subjetivos ou possuem uma existência de primeira pessoa. A subjetividade ou ontologia de primeira pessoa (ii) consiste apenas no fato de que estados mentais conscientes só existem na medida em que são experienciados por um sujeito. Essa é uma das características principais da mente consciente em sua teoria do naturalismo biológico, pois é em virtude dela que a consciência não pode ser reduzida ontologicamente a processos cerebrais. A intencionalidade (iii) é a capacidade que a mente possui de representar ou de se dirigir a outros objetos e fatos do mundo. Mas vale observar que, para Searle, nem todos os estados conscientes são intencionais, como por exemplo, a ansiedade. A causalção mental (iv) é a capacidade que a mente possui de ser eficiente, isto é, de fazer com que outras coisas ou fatos ocorram (SEARLE, 1999, p. 177). Por exemplo: O sujeito A teve sede e foi até um quiosque comprar água. Nesse caso, sua ação de comprar água foi causada por seu estado mental da sede e da rede de estados mentais intencionais, a qual ele faz parte.

Para que um ser A tenha uma mente, não é necessário que A tenha todas essas propriedades ao mesmo tempo, mas para ser consciente, A necessariamente deve apresentar as propriedades de serem internos, subjetivos e qualitativos. Por esses motivos, afirmar que um programa qualquer não é uma mente consciente significa a partir da visão searleana: Em primeiro lugar, um programa não possui estados internos, qualitativos e subjetivos, os quais são causados por processos de nível superior. Em segundo lugar, ele não possui uma existência de primeira pessoa, ou seja, uma subjetividade ontológica. Em terceiro lugar, ele não apresenta uma intencionalidade intrínseca. Um computador possui um conjunto de regras que, por sua vez, constituem a sintaxe de uma linguagem, mas essa linguagem possui apenas uma intencionalidade derivada. Por último, um programa computacional qualquer não teria uma causalidade mental no mundo. Por exemplo: Se um computador jogasse xadrez com o indivíduo A e ganhasse o jogo, a ação de ganhar o jogo de xadrez só teria uma causalidade mental se ela fosse a realização das condições de satisfação de um conjunto de intenções que fazem parte de sua rede de estados intencionais e de seu *background*. Em outras palavras: Ele deveria inicialmente ter tido um conjunto de intenções como: (i) a intenção de jogar xadrez com o indivíduo A e (ii) ganhar do mesmo. Se ele não tem uma rede de estados intencionais nem um *background*, seja ele profundo ou local, isso não apresenta uma causalção mental no sentido searleano.

Contudo, devemos considerar aqui que Searle não nega a tese de que programas poderiam ter ou instanciar estados cognitivos enquanto propriedades emergentes, mas a tese de que a execução de um programa computacional “qualquer” já é suficiente para garantir a vida

mental ou constituir uma mente (SEARLE, 1997, p. 40). Analisemos agora o segundo argumento que esse autor apresenta contra a IA forte.

5. Por que a sintaxe não é uma propriedade intrínseca de objetos?

A tese de Searle de que a sintaxe de uma linguagem não é uma propriedade intrínseca dos objetos físicos pode ser melhor compreendida por meio de algumas teses e distinções que fazem parte de sua teoria da realidade social, especialmente desenvolvida em sua obra “The Construction of the Social Reality” (1997). A ideia principal dessa teoria é que nossas mentes são capazes de construir uma realidade social por meio da intencionalidade e da linguagem. Por exemplo: Um pedaço de papel não passa de um fato bruto, mas ao longo da história, alguns pedaços de papéis passaram a desempenhar determinadas funções como, por exemplo, ser dinheiro, certificado etc. Essa função denominada por Searle de “função de status” faz com que um fato bruto se torne um fato social ou institucional dentro de uma determinada comunidade. Desse modo, um pedaço de papel pode possuir um determinado valor e com ele podemos geralmente comprar outros objetos tais como uma garrafa de água ou uma casa. Imaginemos que o indivíduo A tenha entregue uma nota de cinco euros para o vendedor do quiosque B. A propriedade daquele pedaço de papel (fato bruto) de valer cinco euros não é uma propriedade intrínseca desse pedaço de papel, mas existe apenas em relação a observadores. Isso significa que o pedaço de papel possui uma propriedade extrínseca que é ser dinheiro. De modo análogo, os significados das palavras são características extrínsecas e só existem em relação a observadores. Ao contrário do dinheiro e significado das palavras ou frases, a mente consciente possui uma propriedade intrínseca, visto que ela existe independentemente do que outros possam pensar acerca da mesma.

Neste contexto, Searle destaca que esse argumento é diferente e mais profundo do que o argumento apresentado em seu experimento do quarto chinês (SEARLE, 1997, p.44). Enquanto esse argumento esclarece que a sintaxe não é uma característica intrínseca de objetos físicos, os argumentos do quarto chinês descrevem apenas que a sintaxe não garante a semântica da linguagem. Isso significa que apesar de um programa poder possuir uma sintaxe, essa sintaxe só existe em relação a observadores externos e a existência dessa sintaxe não garante o domínio da semântica da linguagem. Por esses motivos, devemos rejeitar a ideia de que “qualquer” programa computacional pode garantir a existência de uma mente, isto é, um conjunto de estados mentais conscientes e/ou intencionais.

6. Considerações finais

Após a explanação dos dois principais argumentos searleanos contra a tese da IA forte de que “qualquer” programa computacional é suficiente para gerar uma mente, podemos responder a questão principal desse artigo, a saber: Quais foram os desafios que Searle estabeleceu por meio de seus argumentos aos defensores de uma IA forte?

Nesse artigo, analisamos dois argumentos searleanos contra a tese da IA forte de que programas computacionais são mentes. De acordo com o primeiro argumento que é apresentado no experimento do quarto chinês, um programa computacional possui apenas uma sintaxe e essa sintaxe não é suficiente para a existência da semântica. A passagem da sintaxe de uma linguagem para a semântica ocorre devido ao fato de que seres conscientes possuem uma rede de estados mentais intencionais e um *background*, os quais determinam condições de satisfação aos seus enunciados. Usuários de uma linguagem transferem sua intencionalidade intrínseca para as palavras e enunciados e os conferem um significado (o significado do falante). Além disso, a compreensão de enunciados é, para Searle, uma capacidade biológica e mental. Desse modo, uma vez que um programa computacional só possui uma sintaxe e não uma semântica, ele não é idêntico a uma mente humana.

Nesse contexto, a questão que se levanta é: será que temos aqui uma circularidade no argumento do quarto chinês, posto que na segunda premissa ela já pressupõe que a capacidade semântica é mental e biológica? Em minha opinião, não. Ele apenas descreve a mente como sendo uma rede de estados intencionais interligados que funcionam sob um *background* e isso lhe permite ter uma capacidade semântica. A rigor, poderíamos utilizar aqui o *princípio de identidade dos indiscerníveis* de W. Leibniz (1646-1716) de que duas entidades x e y só são idênticas se todas as suas propriedades são as mesmas. Se a mente possui uma capacidade semântica e o programa computacional não possui essa característica, então essas duas entidades não são iguais. Com isso, ele lança vários desafios para os idealizadores da IA forte, a saber: O programa computacional deve possuir uma capacidade semântica para ser idêntico às nossas mentes. Nossa capacidade semântica é mental e biológica e pressupõe intencionalidade intrínseca e um *background*. Sendo assim, programas computacionais devem também possuir intencionalidade intrínseca e *background*. Contudo, devemos observar que mesmo que estas teses sejam verdadeiras, elas são apenas verdades contingentes. Sendo assim, ele não poderia afirmar que é impossível, do ponto de vista lógico, que no futuro tenhamos

programas que instanciem estados mentais. No entanto, ele consegue demonstrar que “qualquer” programa não é suficiente para gerar uma mente análoga ou idêntica à mente humana.

Conforme o segundo argumento de Searle, as frases ou “respostas” que um programa qualquer nos oferece possuem uma sintaxe, ou seja, um conjunto de regras que permitem montar uma palavra ou uma frase com sentido. O problema é que a sintaxe tem palavras com significados e sentidos somente em relação a nós usuários (conscientes) de uma linguagem. O significado não é uma característica intrínseca desses termos e enunciados. Além disso, a sua compreensão pressupõe o reconhecimento das intenções do falante, de uma parte da rede de estados mentais e capacidades de *background*. A rigor, “enunciados” feitos por programas que não instanciam estados mentais conscientes não poderiam ser considerados por Searle sequer como atos ilocucionários. Por esses motivos, não faria sentido tentarmos compreender que atos perlocucionários um programa computacional “pretende” obter por meio de seus proferimentos, mas apenas nos perguntar com qual objetivo ele foi programado.

Retomemos aqui o exemplo do jogo de xadrez, no qual o computador ganha do indivíduo A. Nesse caso, apesar do computador ser considerado por nós (usuários conscientes de uma linguagem) como o vencedor, não faria sentido afirmar que esse programa computacional compreende que ele ganhou esse jogo de xadrez nem que ele ficou triste ou alegre após sua vitória. Aquilo que denominamos “vitória” já faz parte de uma realidade social que construímos intersubjetivamente através da linguagem e de nossa intencionalidade intrínseca. De fato, seria no mínimo estranho parabenizá-lo seriamente por essa vitória, pois essa atitude intencional só faria sentido se o programa tivesse aprendido as regras do jogo de xadrez por meio de vários treinos com a intenção de ganhar esse jogo. Algumas pessoas nem mesmo diriam seriamente que ele “sabe” jogar xadrez, pois ele não aprendeu essa atividade, mas apenas “comportou-se” como lhe foi programado e não seria assim auto-suficiente. Conforme nossas experiências e pre-suposições de *background*, sabemos que comportar-se como sabendo algo, sendo inteligente ou alegre, não é o mesmo que saber algo, ser inteligente ou alegre. Em outras palavras: Não basta simular ter uma mente para ter uma mente, visto que parecer ter x ou comportar-se como sendo x não é o mesmo que ter ou ser x.

Em suma: um dos principais méritos dos argumentos searleanos contra a IA forte é o esclarecimento e estabelecimento de diversos desafios para esses idealizadores. Searle nos lembra que a nossa mente não é um fenômeno biológico com apenas uma estrutura formal, mas também com um conteúdo semântico. Esse conteúdo semântico foi provavelmente

desenvolvido por meio de nossa intencionalidade intrínseca e do *background*. Do ponto de vista puramente lógico, é possível que um dia tenhamos um computador digital que seja capaz de instanciar uma vida mental (com pensamentos, dores, raivas etc) parecida com a nossa. Contudo, para que isso seja empiricamente possível, é necessário que esse computador digital possua de fato uma vida mental com as propriedades naturais e biológicas que nossas mentes apresentam, e não que apenas que ele simule ter algumas dessas propriedades mentais. Além disso, nesse âmbito, não devemos ignorar o fato de que a lista de capacidades que nossas mentes desenvolveram ao longo da história para compreendermos os proferimentos linguísticos e a partir daí construirmos uma realidade social é extensa.

Referências

- AUSTIN, J. L. *How to do things with words*. Oxford: Oxford University Press, 1962.
- CARVALHO, J. M. Intencionalidade e consciência em Searle. *Revista de Filosofia Moderna e Contemporânea*, v. 4, n. 2, p. 54-66, 2016.
- CARVALHO, J. M. *Linguagem e intencionalidade*. Guarapuava: Apolodoro Edições Virtuais, 2019.
- DREYFUS, H. *What computers can't do*. New York: Harper & Row, 1972.
- NEGRU, T. Intentionality and Background: Searle and Dreyfus against Classical AI Theory. *Filosofia Unisinos* 14, 1, jan./abr., p. 18-34, 2013.
- SEARLE, J. R. *Os actos de fala*. Coimbra: Livraria Almedina, 1984.
- SEARLE, J. R. *Mente, cérebro e ciência*. Lisboa: Edições 70, 1984.
- SEARLE, J. R. *The Construction of Social Reality*. Greatos Britain: Penguin Books, 1996.
- SEARLE, J. R. *A redescoberta da mente*. São Paulo: Martins Fontes, 1997.
- SEARLE, J. R. *O mistério da consciência*. São Paulo: Editora Paz e Terra, 1998.
- SEARLE, J. R. *Intencionalidade: um ensaio de filosofia da mente*. Lisboa: Relógio D'água, 1999.

SEARLE, J. R. *Mente, linguagem e sociedade: filosofia no mundo real*. Rio de Janeiro: Editora Rocco, 2000.

SEARLE, J. R. *Liberdade e neurobiologia*. São Paulo: Editora UNESP, 2007.

SEARLE, J. R. *Consciência e linguagem*. Tradução de Plínio Junqueira Smith. São Paulo: Martins Fontes, 2010.

TURING, A. Computing Machinery and Intelligence. *Mind*, v. 59, p. 433-60, 1950.

WITTGENSTEIN, L. *Investigações filosóficas*. Coleção Os pensadores. São Paulo: Nova Cultural, 1996.