

Podem as máquinas falar?

Ronaldo Martins¹

Mírian dos Santos²

La più bella e sanguinosa conquista della machina è l'uomo.
(PITIGRILLI. *Il dito nel ventilatore*)

Celebra-se, neste ano de 2012, o centenário de nascimento de Alan Turing, matemático britânico que é por muitos considerado o pai das ciências da computação e, mais especificamente, da Inteligência Artificial. Em 1950, dois anos antes de ser condenado à castração química por ser homossexual, e quatro anos antes de seu suicídio, Turing fez publicar, na revista *Mind*, o influente ensaio "Computing Machinery and Intelligence", cujo objetivo primeiro, declarado já na sentença de abertura, era analisar se as máquinas pode(r)iam pensar. Nosso objetivo aqui será retirar, do texto de Turing, algumas contribuições para a análise de outra questão que, entendemos, especifica a interrogação original e revela-se hoje chave para que possamos entender o funcionamento da linguagem nesta nova realidade digital: podem as máquinas falar? No que se segue, percorremos o trajeto argumentativo do texto original de Turing e, analisamos, com mais profundidade, dois de seus argumentos a favor da possibilidade de uma inteligência mecânica: o problema da incompletude e a questão da aprendizagem de máquina. Por fim, reafirmamos, com o autor, que não há nenhum impedimento formal ou material para que as máquinas possam vir a emular habilidades linguísticas humanas.

Podem as máquinas pensar?

Para exorcizar os demônios que se escondiam e ainda se escondem na polissemia inevitável de cada uma das componentes da questão "podem as máquinas pensar?" — e cujo melhor estratagema é convencer-nos de que não existem — Turing propõe uma nova formulação para o problema:

¹ Professor do Programa de Pós-Graduação em Ciências da Linguagem da Universidade do Vale do Sapucaí – Univás.

² Professora do Programa de Pós-Graduação em Ciências da Linguagem da Universidade do Vale do Sapucaí – Univás.

A nova formulação do problema ['podem as máquinas pensar?'] pode ser descrita em termos de um jogo a que nós chamamos "jogo da imitação". É jogado por três pessoas: um homem (A), uma mulher (B) e um interrogador (C), que pode ser de qualquer dos sexos. O interrogador permanece num quarto, separado dos outros dois. O objetivo do jogo, para o interrogador, é determinar em relação aos outros dois, qual o homem e qual a mulher. É permitido ao interrogador fazer perguntas a A e B, tais como: Será que X poderia me dizer qual o comprimento de seu cabelo? [...] O objetivo do jogo para A é tentar induzir C a uma identificação errada. [...] O objetivo do jogo para a terceira jogadora (B) é ajudar o interrogador. Sua melhor estratégia será provavelmente dar respostas verdadeiras. Ela pode acrescentar frases como: "Eu sou a mulher, não escute a ele". Mas isso será inútil, porque o homem pode dar respostas semelhantes. Agora formulamos a questão: "O que acontecerá quando uma máquina ocupar o lugar de A nesse jogo?" Será que o interrogador decidirá erroneamente com a mesma frequência, quando o jogo é jogado dessa forma, do que quando o fazia ao tempo em que o jogo era jogado entre um homem e uma mulher? Estas questões substituem a pergunta original: "Podem as máquinas pensar?". (Turing, [1950], p. 50-51)

O jogo da imitação — que ficaria conhecido como "Teste de Turing" — vem superar o risco da vagueza metateórica que indisputavelmente contaminaria a discussão acerca da possibilidade de existência de uma inteligência artificial. Na impossibilidade de definir estruturalmente a inteligência humana, não restaria ao pesquisador senão definir o que é um comportamento inteligente. E, se não se pudesse mais distinguir, do ponto de vista de seu desempenho, a máquina e o homem, caberia afirmar que ao homem e à máquina, ainda que a partir de estruturas muito diferentes, poderiam ser consignadas as mesmas categorias comportamentais, entre elas o adjetivo "inteligente". O sucesso no Teste não significa, portanto, que homem funcionaria tal qual a máquina, ou que a máquina mimetizaria, ponto por ponto, os processos efetivamente verificáveis para os homens. A máquina seria tão-somente uma réplica (ou um modelo) que, como réplica, guardaria semelhanças de resultados, sem que se mantivesse qualquer compromisso com semelhanças de processos em relação ao original.

No entanto, Turing identificava nove grandes objeções à sua postulação. Seriam elas argumentos clássicos que desautorizariam não apenas a possibilidade de uma inteligência extra-humana (ou extracerebral), mas mesmo a ideia e a validade do teste que o autor propunha. Reportamos brevemente cada uma dessas objeções e a correspondente resposta de Turing.

A primeira, a objeção teológica, afirmaria, segundo ele, que o pensamento é uma função da alma humana imortal, dádiva de Deus. A objeção é descartada porque envolveria uma contradição: a impossibilidade de as máquinas pensarem implicaria uma séria restrição à onipotência do Todo-Poderoso.

A segunda objeção, por ele chamada a objeção das "cabeças na areia", é de caráter eminentemente ético: as consequências de máquinas pensantes seriam terríveis. Não é, portanto, uma objeção que possa ser refutada a partir do problema, mas da consideração de aspectos sociais que o autor, deliberadamente, evita.

A terceira objeção é a objeção matemática consubstanciada no teorema de Gödel. Em qualquer sistema lógico suficientemente poderoso podem-se formular enunciados que não são passíveis de prova ou refutação dentro do sistema, a menos que possivelmente o próprio sistema seja inconsistente. Em consequência, a máquina proposta por Turing, como sistema lógico, estaria fadada a uma incapacidade à qual o intelecto humano, exatamente por sua ilogicidade, não estaria sujeito. A refutação ao argumento é, a princípio, elusiva: "Nós mesmos frequentemente damos respostas erradas a perguntas, e assim não temos justificativa para estar muito satisfeitos com semelhante evidência de falibilidade por parte das máquinas" (p. 64). No entanto, dada sua importância para a análise da questão a que este trabalho se propõe, voltaremos a esta objeção na terceira seção deste texto.

A quarta objeção é descrita como o argumento da consciência: a única maneira de a pessoa estar segura de que a máquina pensa é ser ela a própria máquina e sentir-se pensando. Turing não refuta este argumento, mas denuncia o seu caráter solipsista e, conseqüentemente, a sua esterilidade. Se é de fato necessário ser e sentir para saber, então cada um de nós está absolutamente inseguro a respeito da inteligência alheia (porque obviamente nós não somos o outro e não experimentamos o que o outro experimenta).

A quinta objeção reúne os argumentos das várias incapacidades. Segundo o autor, o argumento consiste em afirmar que sempre haverá alguma coisa que a máquina não estará aparelhada para fazer (como, por exemplo, apaixonar-se, ou ser o tema de seus próprios pensamentos). Turing reduz essas objeções ou a um problema de memória da máquina (que certamente se expandirá), ou a um problema semântico a respeito do que seja "ser o tema de seus próprios pensamentos", por exemplo.

A sexta objeção vem de uma dissertação de Lady [Ada] Lovelace em que a autora declara que "a Máquina Analítica não tem nenhuma pretensão de criar o que quer que seja. Pode fazer *tudo quanto saibamos ordenar-lhe que faça*" (grifos da autora). O argumento esconde, segundo Turing, duas inverdades: a primeira, que o conceito de "novo" seja decididamente incontestável; a segunda, que a máquina não possa

aprender. Pela sua importância, também voltaremos a este argumento mais adiante, na quarta seção deste texto.

A sétima objeção remete à continuidade do sistema nervoso: a máquina discreta seria incapaz de imitar o comportamento do cérebro humano (reconhecidamente contínuo). A refutação da objeção, especialmente se considerada a proposta de Putnam (1995), que afirma a independência do *software* sobre o *hardware*, pode ser considerada ingênua: Turing simplesmente anuncia a possibilidade de uma máquina não discreta (o analisador diferencial).

A oitava objeção é o argumento da informalidade (imprevisibilidade) do comportamento humano. Na medida em que o comportamento da máquina é previsível, máquinas não poderiam ser confundidas com homens. A resposta de Turing se dá em duas direções: há leis de comportamento que, diferentemente de leis de conduta, são fortemente previsíveis (e constituem o objeto de estudo da Psicologia); e a previsibilidade do comportamento da máquina somente pode ser afirmada, em muitos casos, pelo seu próprio programador. Voltaremos também à questão da "imprevisibilidade" quando analisarmos, com mais detalhes, o teorema de Gödel, na terceira seção.

A nona e última objeção expõe o argumento da percepção extrassensorial. Turing rejeita a desautorização sumária do argumento e postula sua pertinência, mas acredita que a interferência de fatores de ordem extrassensorial poderia ser controlada na experiência proposta.

Como se pode perceber, a resposta final de Turing para a pergunta inicial é positiva: sim, as máquinas podem (vir a) pensar, porque não há objeções insuperáveis à reprodução artificial da inteligência humana. Essa hipótese, evidentemente, ainda não é pacífica. Mais de 50 anos após sua publicação, o Teste de Turing continua sendo objeto de intensa crítica, que tem hoje se concentrado em torno de três pontos principais: o fato de que o Teste não permite diferenciar entre "inteligência real" e "inteligência simulada", ou seja, entre "ser inteligente" e "parecer inteligente"³; o fato de que os resultados do Teste possuem validade muito limitada, porque dependem também da

³ Este, que é talvez o mais forte argumento contra o Teste de Turing, foi exposto por John Searle (1980) por meio do célebre exemplo da câmara chinesa, em que um tradutor, dispendo de uma tábua de correspondências, traduzia um texto sem "entender" efetivamente seu conteúdo. Para Searle, os computadores poderiam pensar se, e somente se, tivessem uma "mente".

"inteligência" do interrogador, que é muito variável⁴; e o fato de que o Teste não avalia efetivamente a "inteligência" de uma máquina, mas apenas sua capacidade de parecer humana⁵.

Observa-se, no entanto, que todas essas novas objeções reeditam o problema terminológico que, desde o início, Turing quis deliberadamente evitar ao propor a ideia de um "teste de imitação" para a avaliação da inteligência artificial. Queremos crer que a discussão poderia ser imediatamente interrompida se afirmássemos que o computador não tem nenhuma "mente" e não é efetivamente "inteligente", no sentido que estes termos podem alcançar no estudo da psicologia humana. Isso não impede, contudo, a constatação de que este computador – que não tem nenhuma "mente" e que não é "inteligente" – possa ser confundido, em um teste como o proposto, com alguém que está equipado com uma "mente" e a quem pode ser atribuído o adjetivo "inteligente". E este, em última análise, é todo o objetivo: desenvolver máquinas que possam provocar este tipo de equívoco. Em que medida esta "inteligência computacional" seria ainda uma inteligência e em que medida esta "mente computacional" estaria relacionada aos estados mentais do homem, embora possam ser questões extremamente pertinentes para os que investigam a psicologia humana, são absolutamente irrelevantes do ponto de vista do desenvolvedor e do usuário de sistemas computacionais.

O fato, porém, é que o Teste de Turing, mais de 50 anos após sua publicação, representa ainda o principal ponto de partida para a discussão e a avaliação de sistemas computacionais ditos "inteligentes", embora esses sistemas tenham abdicado, pelo menos por enquanto, da pretensão de uma inteligência mais geral, e venham se concentrando no desenvolvimento de máquinas que trabalham com objetivos muito mais modestos e específicos. À exceção do controvertido prêmio Loebner, que premia anualmente o melhor sistema de "chatterbox" (ou "máquina de conversar") capaz de "enganar" um humano, a pesquisa em Inteligência Artificial é hoje flagrantemente mais especializada: em lugar de reproduzir toda a inteligência humana, busca-se reproduzir determinados comportamentos inteligentes, como o de jogar xadrez, por exemplo. Ou o de executar determinadas tarefas linguísticas, como as de corrigir, resumir e traduzir

⁴ Interrogadores mais ingênuos ou mais cooperativos (como os que, por exemplo, já tendem a antropomorfizar seus dispositivos eletrônicos, com quem já conversam e em que identificam sentimentos e disposições tipicamente "humanos", por mais estúpidos que esses dispositivos possam ser) tendem a "aprovar", no Teste de Turing, mecanismos que seriam rejeitados por outros interrogadores, mais céticos ou mais experientes.

⁵ Segundo essa perspectiva, nem todo comportamento humano seria, necessariamente, "inteligente".

textos, que vêm sendo objeto sistemático de todo um campo de investigação científica: o assim chamado "Processamento Automático de Língua Natural", ou PLN. E é exatamente neste ponto que um corolário da questão mais geral passa a nos interessar e a nos incomodar: poderão as máquinas "falar"? Para que possamos examinar essa questão, e evitar, uma vez mais, toda a polissemia do que seja "falar", será necessário, como fez Turing, propor uma nova formulação para o problema, e considerar, com mais detalhe, dois dos impedimentos por ele reportados: o teorema de Gödel e o argumento de Lady Lovelace.

O jogo da competição linguística

Para que possamos endereçar a questão central deste texto será necessário, como fez Turing, escapar às ciladas terminológicas a que nos conduzem a generalidade e a vastidão de um verbo como "falar". Não pode ser nosso objetivo, aqui, tentar definir o que caracteriza o conhecimento ou o comportamento linguístico de um falante. Será mais pertinente propormos, novamente, um jogo, mas que não será, não agora, um jogo de imitação, dado que nosso objetivo será avaliar um modo ou uma forma específica de inteligência: o conhecimento e o uso da linguagem. O novo jogo será antes um jogo de competição linguística, jogado por três falantes: A, B e C. Ao avaliador C caberia propor uma série de tarefas linguísticas (como as de traduzir, resumir ou corrigir textos, por exemplo) e verificar quem, entre os competidores A e B, as realiza com maior desenvoltura, segundo os critérios normalmente aceitos pela comunidade de fala, e tomando por base apenas os resultados da produção (ou seja, sem que houvesse interação direta entre C e os competidores). Trata-se, em linhas gerais, de um velho jogo, conhecido de todos nós, e praticado sistematicamente em vários exames de seleção. A questão que nos interessa, como no caso de Turing, é a seguinte: o que acontecerá quando uma máquina ocupar o lugar de A nesse jogo? Será que, em uma avaliação às cegas, o competidor humano suplantar sempre o competidor mecânico? Estas questões, aqui como antes, substituem a pergunta original: podem as máquinas falar?

Observe-se que não se faz aqui a afirmação, como Turing não a fizera, de que as máquinas reproduzirão, ponto por ponto, o comportamento linguístico humano. Não se trata de observar se as máquinas falarão como os homens falamos. Trata-se apenas de verificar em que medida as máquinas seriam capazes de provocar o equívoco, e de

avaliarmos sua inteligência — agora "inteligência linguística" — por recurso à incidência desse equívoco. Não nos importa, aqui, se as máquinas de fato "entenderam" os textos que eram objeto de análise; se foram capazes de formular, a partir deles, os mesmos sentidos que um humano formularia; se foram ou não afetadas por eles, como um humano o seria. Interessa-nos apenas verificar se a máquina soube produzir, como resultado de uma tarefa linguística, um produto que pudesse ser confundido (e mais bem avaliado) do que o produzido por um humano. Como no teste comportamental de Turing, não nos interessam os "estados interiores" da máquina ou do humano, mas apenas o produto externo, público, mensurável de uma determinada atividade linguística. Nesta perspectiva, nossa análise se reduz ainda a um critério behaviorista para a inteligência humana, que já se provou inadequado para que se possam explicar inúmeros fenômenos linguísticos, como a própria aquisição da linguagem humana⁶. Mas nosso objetivo, uma vez mais, não é iluminar aspectos obscuros da fala humana. Trata-se tão-somente de verificar se falantes mecânicos podem ser confundidos com falantes humanos em uma competição que envolvesse o desempenho em tarefas linguísticas que envolvem, não apenas a manipulação do significante, mas também a manipulação do significado.

Se concordarmos, com Turing, que nenhuma das objeções elencadas constitui impedimento efetivo para que a máquina possa vir a ser bem sucedida neste novo jogo, poderíamos também concluir que, sim, as máquinas podem falar. Mas a questão, aqui como lá, não parece efetivamente ser "se" as máquinas podem ou não pensar (ou falar); a verdadeira questão, que se insinua no texto de Turing e que começa aqui a reclamar também alguma atenção, é "como" as máquinas poderiam vir a fazê-lo. Mas para que possamos avançar mais sobre esse ponto, será necessário que examinemos, com mais profundidade, dois dos impedimentos reportados por Turing: o teorema de Gödel e o argumento de Lady Lovelace.

Da incompletude

⁶ O célebre argumento sobre a pobreza dos estímulos primários, apresentado por Chomsky (1980), reafirma a necessidade de uma faculdade linguística inata – dos "estados mentais internos" ignorados pelas abordagens behavioristas – que habilite as crianças a induzir a gramática de uma língua a partir da experiência linguística muito precária e fragmentária do "motherese", a "língua da mãe" a que elas são normalmente expostas durante o processo de aquisição da linguagem.

Segundo o primeiro teorema da incompletude de Gödel [1931]⁷, os sistemas formais não podem ser ao mesmo "consistentes" e "completos": se são consistentes, são incompletos; se são completos, são inconsistentes. Não é possível flagrar, em um sistema formal, um estado de coisas em que cada uma delas possa ser exclusivamente definida por oposição às outras, ou seja, diferencialmente. Deverá haver sempre um conjunto de conceitos (os axiomas) que fazem referência a elementos extrassistêmicos (os postulados), ou, dito de outra forma, dentro de um sistema de relações há sempre elementos que devem ser definidos positivamente. O teorema prova a existência das chamadas "proposições indecidíveis", ou seja, que não podem ser provadas verdadeiras ou falsas em um dado sistema axiomático.

O teorema de Gödel, que representa a terceira objeção repertoriada por Turing, constituiria, pois, um sério impedimento à ideia de máquinas pensantes, que deveriam ser ao mesmo tempo "consistentes" e "completas", ou seja, autossuficientes, sob o risco de se tornarem dependentes do humano e, por extensão, incapazes de um processamento completamente automático, como o que esperamos das máquinas. Ainda que pudéssemos criar um sistema consistente que (re)produzisse os mesmos resultados produzidos pelos humanos, esse sistema seria incompleto e dependeria, em algum ponto, de informações definidas fora do sistema, ou seja, que foram nele inscritas pela inteligência humana, mas que não podem ser ali representadas. Em resumo: todo sistema formal contém, necessariamente, "falhas".

A principal implicação do teorema de Gödel é a ideia de que a inteligência humana não é completamente "formalizável", ou de que haverá sempre uma distância intransponível entre as línguas naturais e línguas artificiais desenhadas para mimetizá-las. De que as máquinas são, também elas, e apesar de toda a mitologia que se constituiu sobre elas, ferramentas imperfeitas.

A refutação de Turing, reportada na primeira seção deste texto, é curiosamente atual. Turing afirma que a inteligência humana, como uma eventual inteligência

⁷ O primeiro teorema de Gödel, também conhecido como teorema da incompletude ou da indecidibilidade, expresso em notação matemática, foi assim traduzido por Kleene: "Any effectively generated theory capable of expressing elementary arithmetic cannot be both consistent and complete. In particular, for any consistent, effectively generated formal theory that proves certain basic arithmetic truths, there is an arithmetical statement that is true, but not provable in the theory." (Kleene 1967, p. 250).

artificial, também é "imperfeita"⁸. Ou seja, de que o teorema de Gödel se aplicaria, não apenas aos sistemas formais, mas à própria estrutura do pensamento humano e, por que não dizer, da linguagem humana, que poderiam provar sua consistência se, e somente se, fossem inconsistentes.

Afirmar a "incompletude" do pensamento e da linguagem [do homem e da máquina] é reconhecer ali a presença de uma "falta": de uma "exterioridade", que não deve ser interpretada, porém, no sentido do "fora" da linguagem ou do pensamento. Como afirma Orlandi (1996, p. 31):

Essa exterioridade da qual falamos não tem a objetividade empírica do "fora" da linguagem. Ela é tomada tal como intervém na textualidade. É exterioridade discursiva (e não empírica). Ou seja, esta exterioridade é o interdiscurso, definido em sua objetividade material contraditória (M. Pêcheux, 1988): algo fala sempre antes, em outro lugar e independentemente, isto é, sob o domínio do complexo das formações ideológicas. Isso dá ao sujeito a sua realidade enquanto sistema de evidências e de significações experimentadas. Aí se dá o processo de constituição do discurso, na memória, no domínio dos dizeres já-ditos ou possíveis que garantem a formulação do dizer. É esse jogo entre a formulação e a constituição dos sentidos que produz o efeito de exterioridade, o sentido-lá.

A exterioridade é, pois, uma "anterioridade", um "antes. A impossibilidade de uma circularidade perfeita, na linguagem como no pensamento, é função do caráter histórico desses sistemas. Os sistemas estão em perpétuo movimento e não se pode isolá-los em um determinado estágio sem que não se faça referência aos estágios anteriores, de que eles são não apenas uma sequência, mas um produto, e com quem integram a mesma narrativa. A tese parmenidiana de que não existe a mudança, ou de que é possível abstrair dos recortes da realidade o seu aspecto dinâmico, "enganador" — o que legitimaria uma abordagem atemporal da língua e da inteligência — fracassa quando da tentativa de explicar aparentes anomalias sincrônicas que não deixam de ser, sobretudo, regularidades diacrônicas.

Não pode ser objetivo deste texto investigar em que medida essa "historicidade" funciona da mesma forma no homem e na máquina. Orlandi (2009) propõe que estamos

⁸ A ideia de imperfeição aparece aqui entre aspas porque é, também ela, "imperfeita", como o denuncia, por exemplo, Paul Henry (1992). Caberia falar em "pensamento imperfeito" ou "linguagem imperfeita" se, e apenas se, admitirmos a possibilidade de um "pensamento perfeito" ou de uma "linguagem perfeita", o que constitui, na verdade, um oxímoro, se aceitamos, como se faz aqui, que o pensamento e a linguagem são, por definição, incompletos ou inconsistentes.

diante aqui de duas formas de "memória" distintas em suas materialidades: a "memória histórica", característica do homem, que é limitada em sua extensão e marcada pela "falha", pelo esquecimento, mas que é, por isso mesmo, condição do "possível" e do "vir-a-ser", ou seja, da "produção", da "formulação", da "re-significação"; e a "memória metálica", característica da máquina, que seria aquela "que não falha", "ilimitada em sua extensão", mas que, exatamente pela ausência de falhas, do esquecimento, apenas "produz o mesmo em sua variação", a "fórmula", o "rearranjo".

Se admitirmos, porém, que a "falha" é também característica da máquina — ou seja, de que a memória da máquina, embora ilimitada em sua extensão, não é estática, mas dinâmica, e encontra uma formulação, sempre provisória, em uma língua que, embora lógico-matemática, não é, porém, desmemoriada, porque também ela incompleta, e regularmente autofecundada pelas interações do sistema —, se reconhecemos, enfim, que as máquinas também têm "história", representada em seus diferentes estados de memória, cumpriria admitir que, sim, as máquinas também poderiam ser bem sucedidas em uma tarefa linguística que procurasse emular, por exemplo, a "produtividade linguística" do falante humano.

Há pelo menos um exemplo notável dessa "incompletude" da máquina: o mecanismo de retroalimentação que caracteriza a chamada Web 2.0, a geração de serviços digitais "colaborativos", ou seja, que se distinguem pela incorporação continuada da opinião e de dados do usuário. Diferentemente da geração anterior, de sistemas invariáveis e impermeáveis às interações com os usuários, em que a informação fluía em único sentido, dos servidores para os terminais, os serviços atuais são bidirecionais e muito mais porosos: utilizam a história do usuário (ou da comunidade de usuários) para produzir resultados consideravelmente variáveis. Os atuais motores de busca, por exemplo, utilizam parâmetros dinâmicos (tais como o histórico de páginas visitadas, os dados geostacionários do usuário e a avaliação da comunidade de usuários) como critérios de busca, de forma que a mesma pesquisa, feita por usuários diferentes, ou em dias ou lugares diferentes, pode produzir resultados diferentes. Um segundo exemplo: os anúncios publicitários que pululam nas páginas da internet são quase sempre "customizados" (ou seja, "escolhidos" para cada usuário) a partir de dados normalmente coletados nas inúmeras redes sociais, como o Twitter e o Facebook. Como um último exemplo, citem-se os sistemas de tradução automática, como o Google Translate, que operam hoje com uma memória incremental,

cotidianamente atualizada pelo acréscimo de novos corpora, de novas bases de conhecimento, de novos glossários terminológicos e memórias de tradução, além, evidentemente, do feedback dos usuários. Em todos esses casos, a máquina já não é mais o lugar da mera repetição.

O fato é que o ciberespaço já não é hoje, definitivamente, o espaço do "mesmo", e sofre um agudo processo de dialeção e idioleção da informação que parece nos conduzir, em breve, para a singularização de nossa interação com a máquina: os resultados serão individualizados, construídos a partir de uma memória digital personalizada, embora parte de memórias mais gerais, e nem sempre armazenada apenas nos dispositivos do usuário.

A questão que se coloca é, pois, se essa memória, tão dinâmica e tão variada, não será capaz de produzir a "re-significação", tão caracteristicamente humana. Queremos crer que essa memória coincide ainda com a "memória discursiva", no sentido que este termo recebe em Pêcheux (1999):

A memória discursiva seria aquilo que, face a um texto que surge como acontecimento a ser lido, vem restabelecer os 'implícitos' (quer dizer, mais tecnicamente, os pré-construídos, elementos citados e relatados, discursos-transversos, etc.) de que sua leitura necessita: a condição do legível em relação ao próprio legível. (p. 52)

Apenas ela "significa", por exemplo, as diferenças entre os resultados de um mesmo motor de busca para uma mesma pesquisa, ou entre os diferentes anúncios publicitários exibidos na mesma página do Facebook de diferentes usuários, ou as diferentes traduções que se obtêm frequentemente para os mesmos textos. Mas compreendemos também que o funcionamento dessa que vai se tornando uma "hiper-memória", caracterizada por uma extensão incremental e ilimitada, e produto de uma progressiva particularização, estaria associado, não à falha do esquecimento, mas à permanente reorganização do sistema, pela incorporação de mais e novas informações. No entanto, não nos parece que estejamos ainda diante de uma outra memória. Trata-se, com certeza, de uma memória que não esquece, e que vai acumulando os "já-ditos"; mas esta acumulação não é a-histórica: o novo é "interpretado" à luz do que já está armazenado, e que é continuamente ressituaado dentro do sistema. Não se trata, pois, de mero empilhamento de enunciados. Ou seja, o não esquecimento, qual no caso de

Funes, "el memorioso", célebre personagem de Borges, não representa uma impossibilidade de re-significação, e nossas interações com a máquina o vêm provando cada vez mais frequentemente.

Poder-se-ia argumentar, talvez, que esta memória seria variável apenas em sua forma empírica, mas não em sua forma material. Ou seja, que o acúmulo de informações afetaria apenas a "organização" da memória da máquina, e jamais sua "ordem". Mas o fato é que as máquinas são hoje menos passivas do que podem parecer, principalmente quando dotadas, como indicado na próxima seção, de um algoritmo de aprendizagem, objeto da última e uma das mais notáveis partes do texto de Turing.

Da aprendizagem de máquina

A última questão que emerge da proposta de Turing diz respeito à aprendizagem e tem muito a ver com vários modelos propostos no âmbito da Inteligência Artificial, naquilo que vem sendo hoje referido como "aprendizagem de máquina". Sobre esse ponto, seria interessante partir de um fragmento de Simon:

Esta afirmativa de que os computadores podem fazer aquilo para que são programados é intuitivamente óbvia e intuitivamente verdadeira, mas não apoia nenhuma das implicações que são ordinariamente daí extraídas. Um ser humano pode pensar, aprender e criar porque o seu programa biológico assim o dotou, conjuntamente com a capacidade para as alterações que este programa produza pela interação com o ambiente e que o habilitem a pensar, aprender e criar. Se um computador pensa, aprende e cria é porque um programa o dotou de tais capacidades. Este programa não será repetitivo e estereotipado, mas fará com que seu comportamento seja condicionado pelo meio ambiente. Terá capacidade de extrair indícios do meio ambiente, indicando seu próprio progresso em relação aos objetivos. Terá de ser um programa que analise por algum meio seu próprio desempenho, diagnostique suas falhas e providencie mudanças para melhorar sua eficiência futura. (SIMON, [1960], p. 116)

A observação de Simon sintetiza o argumento de Turing contra Lady Lovelace e introduz uma dicotomia que deve aqui ser considerada. Na tentativa de replicar o comportamento cognitivo do homem, apresentam-se basicamente duas respostas: ou ensinar a máquina a pensar, ou ensinar a máquina a aprender a pensar. Lady Lovelace, em sua dissertação, considera apenas a primeira possibilidade e, contra ela, afirma que a representação do conhecimento do homem não seria suficiente para que se pudesse afirmar que as máquinas pensam, porque, em última instância, as máquinas estariam apenas a reproduzir aquilo para o que foram programadas. Os homens, segundo Lady

Lovelace, não seriam tão mecânicos, ou contariam com um grau considerável de imprevisibilidade — que explicaria, por exemplo, a questão da variação — que desautorizaria essa aproximação.

A refutação ao argumento de Lady Lovelace tem duas faces: a primeira, interna, já foi considerada nos comentários sobre as objeções ao teste do Turing e, pode ser definida, essencialmente, pela crítica à ideia de “imprevisibilidade” – “imprevisível” pode muito bem ser definido como “qualidade do que ainda não pôde ser previsto”, em lugar de “qualidade do que não poderá nunca ser previsto”. Não haveria, portanto, nenhuma espécie de impedimento original à proposta de uma máquina pensante. No entanto, o argumento de Lady Lovelace deve ser interpretado à luz de uma questão mais prática, ou seja, somente se poderá ensinar a máquina a pensar quando estiver extremamente claro o funcionamento da inteligência humana. A possibilidade de uma máquina que pensa estaria associada à previsibilidade total do comportamento humano, o que é ainda uma quimera. Em última análise, o que Lady Lovelace afirma é que a máquina será como um homem apenas quando o homem puder ser definido como uma máquina. Trata-se, evidentemente, de uma possibilidade – quem garantirá que não se estabelecerão modelos totais do comportamento humano? –, mas de uma possibilidade hoje improvável e muito remota. Lady Lovelace duvida, portanto, que se possa representar integralmente, satisfatoriamente, robustamente, todo o conhecimento do homem, condição *sine qua non* para uma máquina pensante.

O texto de Simon expõe, porém, outra refutação ao argumento de Lady Lovelace. Trata-se da segunda estratégia para fazer com que as máquinas pensem: ensinar a máquina a aprender a pensar. Essa segunda resposta, contra a primeira [ensinar a máquina a pensar], procura evitar exatamente o desenvolvimento de uma plataforma de trabalho que dependa excessivamente do conhecimento de mundo acumulado pelo homem ao longo de sua história. Desconfia-se, deste lado, que é impossível representar todo o conhecimento humano, porque esse conhecimento está em permanente transformação; e que essa síntese, ainda que fosse possível, seria uma síntese não-computacional (não-formalizável). É forçoso reconhecer que o argumento de Lady Lovelace, em função dessa nova posição, perde o vigor. Não seria necessário esperar que o homem pudesse ser definido como máquina para que as máquinas pudessem funcionar como homens. Bastaria que se estabelecesse um modelo de aprendizagem que, na máquina, replicaria o que se passa com o homem. Nas palavras de Turing:

No processo de tentar imitar a mente humana adulta, temos de refletir bastante sobre o processo que a levou até o ponto onde se encontra. Cumpre atentar para três componentes:

- (a) O estado inicial da mente, isto é, ao nascer;
- (b) A educação que recebeu;
- (c) Outras experiências, que não são as descritas como educação, a que foi submetida;

Em vez de tentar produzir um programa que simule a mente adulta, por que não tentar produzir um que simule a mente infantil? Se ele fosse então submetido à educação apropriada, ter-se-ia um cérebro adulto. Presumivelmente, o cérebro da criança é algo assim como um desses cadernos que se compram em papelaria. Pouco mecanismo e muitas folhas em branco. Nossa esperança é a de que haja tão pouco mecanismo no cérebro da criança que algo que se lhe assemelhe pode ser programado. (...)

Dividimos assim nosso problema em duas partes: o programa infantil e o processo de educação. Estas duas partes permanecem intimamente ligadas. Não podemos esperar encontrar uma boa máquina-criança logo na primeira tentativa. Deve-se experimentar ensinar uma máquina tal para ver como ela aprende. Pode-se então tentar ensinar outra para ver se sai melhor ou pior. Há uma conexão óbvia entre esse processo e a evolução, por via destas identificações:

Estrutura da máquina-criança = material hereditário

Mudanças na máquina-criança = mutações

Seleção natural = juízo do experimentador (TURING [1950], p. 77-78)

Há seguramente, também aqui, várias inconsistências. A primeira delas diz respeito à eleição dos primitivos, dos pontos de partida, à montagem da “estrutura da máquina-criança”. A questão não é nem um pouco trivial e remete ao debate, secular, sobre o que é inato e o que é adquirido no comportamento humano. Que o cérebro da criança possa ser definido “como um desses cadernos que se compram em papelaria” já representa uma assunção hoje extremamente perigosa, porque corre o risco de reproduzir uma postura excessivamente passiva para o sujeito, que tem se revelado, desde Piaget e Chomsky, inadequada, por exemplo, para explicar o desenvolvimento da linguagem pela criança.

A questão talvez pudesse ser contornada se insistíssemos sobre o papel ativo dos “mecanismos” em detrimento das “folhas em branco”, mas este tem sido o ponto indecível onde se confrontam inatistas, construtivistas e interacionistas.

Há um segundo problema que deriva da formalização das mutações responsáveis pelas mudanças na máquina-criança. A dificuldade aqui talvez se revele inferior porque há já disponíveis regras de aprendizagem para modelos matemáticos (a regra de Hebb, por exemplo) que podem ser facilmente implementadas computacionalmente (e de fato o foram nos modelos conexionistas, sob a forma do *perceptron*, cf. Rumelhart 1996).

Mas há ainda uma dificuldade nada ignorável de alcance: as regras conseguem mudar a relação entre primitivos, mas não conseguem mudar os primitivos. Em que medida o processo de aprendizagem humana poderia ser replicado a partir de modificações estritamente relacionais?

No entanto, parece evidente, hoje, que as máquinas podem "aprender", ainda que não o façam de forma equivalente aos humanos. A postura hoje hegemônica no campo do processamento automático das línguas naturais, em inúmeros domínios, é exatamente a da formalização e implementação de um modelo de processamento da linguagem capaz de adquirir conhecimento linguístico, principalmente por meio de inferência estatística, a partir de *corpora*. Nas últimas competições entre sistemas promovidas pelo National Institute of Standards and Technology (Open MT Evaluation, NIST 2009), vários dos sistemas de tradução automática mais bem colocados, embora não alcançassem ainda nem mesmo 50% da referência humana, operavam apenas com modelos de linguagem extraídos automaticamente de *corpus*, ou seja, que não incorporavam conhecimento metalinguístico explícito. Parece, evidente, pois, que é possível à máquina adquirir algum conhecimento linguístico a partir de suas próprias experiências linguísticas, ainda que os resultados — pelo menos por enquanto — não nos pareçam satisfatórios.

Conclusão

Entre a possibilidade conceitual e a existência material de uma máquina que fala existe uma distância que, pelo menos hoje, é considerável. Em parte, poder-se-ia dizer, porque a linguagem não é lógica e, conseqüentemente, pouco passível de ser replicada a partir de modelos artificiais invariavelmente amparados na ideia (binária) do tudo ou nada. A univocidade, limitação física pressuposta por qualquer implementação computacional da língua, impediria o tratamento da irrecusável equivocidade emergente dos dispositivos linguísticos, reduzidos, na máquina, à mera caricatura do real.

No entanto, a confortável recusa da possibilidade de formalização da linguagem é tão legítima quanto estéril. Se é indisputável que haverá, sempre, um intervalo intransponível entre homem e máquina, é também indiscutível que inúmeros dos comportamentos humanos ditos “inteligentes” têm sido satisfatoriamente replicados — embora não mimetizados — pela teoria computacional. O reconhecimento automático,

de caracteres, a síntese de fala, o reconhecimento de voz em modo de ditado, entre outros, já são normalmente "aprovados" em testes como o proposto .

Obviamente, há larga distância entre o reconhecimento de caracteres e o processamento de alto nível da linguagem. A complexidade, que lá já não é pequena, revela-se inextricável quando se pensa, por exemplo, na construção de analisadores morfológicos e sintáticos, e absolutamente inabordável quando da formalização de analisadores semânticos, pragmáticos e discursivos. O que não pode significar que a máquina seja intrinsecamente incapaz de processar a língua natural. A complexidade do fazer linguístico aponta apenas para a impossibilidade procedimental de a máquina falar (ou pensar) tal qual fazem os homens; não compreende, porém, sua capacidade — e esta a contribuição de Turing — de atingir alguns dos mesmos resultados de forma completamente diferente da realizada pelos humanos. É célebre, por exemplo, o argumento de Putnam (1975) para reiterar o dualismo psicofísico que pelo menos desde Descartes contamina o pensamento ocidental: assim como não se estabelecem restrições proibitivas ou relações determinísticas entre hardware e software, no caso da máquina, também não seria lícito aprisionar na neurologia (*res extensa*) a psicologia humana (*res cogitans*). O pensamento humano — e a inteligência — independeriam, portanto, do cérebro do homem, podendo ser sintetizados pela máquina.

É interessante ressaltar, porém, que a metáfora, aqui, pouco tem a ver com a concepção mecanicista da mente que inaugura a Inteligência Artificial. Não se trata, absolutamente, da mente como máquina, do cérebro como hardware, mas da mente e da máquina como meios diferentes para se atingirem os mesmos resultados. A ausência de qualquer especularidade entre a arquitetura da mente humana e a do computador repele, já de início, a possibilidade de qualquer paralelismo e aproximação entre desempenho humano e desempenho computacional.

Trata-se, na verdade, agora e sempre, do mesmo paradoxo: a possibilidade de existirem dois caminhos — o humano e o computacional — para se chegar ao mesmo resultado: o comportamento linguístico inteligente. Essa polêmica redefinição de inteligência como epifenômeno, que constitui o centro de gravidade do texto de Turing, instala a necessidade de se investigarem vias de acesso que não passem, necessariamente, pelo que se passa efetivamente com o falante quando fala.

Esse objetivo se sobrepõe, muitas vezes, ao programa de pesquisa da semântica formal, mas não se confunde, em princípio, com as abordagens clássicas. Em primeiro

lugar, porque o que lá é o ponto de partida — a possibilidade de formalização da linguagem — aqui é (incerto) ponto de chegada. Em segundo lugar, porque, embora a perspectiva psicológica já não constitua o principal viés de análise, as restrições de desempenho — relacionadas aqui à máquina — sempre conduzirão a pesquisar muito mais modelos de performance do que teorias da competência. E, por fim, porque a singularidade da tarefa proposta consiste, exatamente, em tentar evitar o desenvolvimento de plataformas fixas ou estacionárias de representação do conhecimento linguístico. Parte-se do pressuposto de que não se pode representar satisfatoriamente, robustamente, todo o conhecimento linguístico para a máquina e, somente então, fazê-la funcionar. Em primeiro lugar, porque não é certo que seja possível representar todo o conhecimento linguístico; além disso, porque este funcionamento estaria ontogeneticamente limitado ao conjunto de informações representadas, negligenciando a produtividade que é característica básica da linguagem.

Interessa, pois, pensar a formalização de um modelo de processamento da linguagem não-supervisionado capaz de fundar suas próprias orientações de pesquisa e universos categoriais. Interessa, não dotar a máquina de um aparato teórico pronto e acabado, mas investigar a possibilidade de investi-la de uma autonomia que a torne capaz de replicar o dinamismo e a instabilidade que são próprios da linguagem. Em última instância, interessa pesquisar a formulação de estratégias computacionalmente implementáveis de aquisição (e não de manipulação) da linguagem.

Não se trata, evidentemente, de um percurso rápido. O trajeto, percebe-se desde o ponto de partida, é sinuoso e escarpado. Como diria Turing, no encerramento de seu ensaio: "podemos avistar só um pequeno trecho do caminho à nossa frente". Porque ensinar a máquina a aprender a linguagem será, sobretudo, ensiná-la a organizar seu passado como memória, os já-ditos como discurso, o enunciado como enunciação. E apenas começamos a fazê-lo. Mas Turing, é importante lembrar, nos encoraja a ir adiante: "[podemos avistar só um pequeno trecho do caminho à nossa frente] mas ali já vemos muito do que precisa ser feito".

Referências

CHOMSKY, Noam. **Rules and representations**. Oxford: Basil Blackwell, 1980.

GÖDEL, Kurt. On formally undecidable propositions of Principia Mathematica and related systems. In: FEFERMAN, S. et al. **Gödel's Collected Works**. Oxford: Oxford University Press, 1990[1931].

HENRY, Paul. **A ferramenta imperfeita: língua, sujeito e discurso**. Campinas: Unicamp, 1992.

KLEENE, Stephen C. **Mathematical Logic**. Dover: 1967.

NIST. NIST. 2009. Open Machine Translation Evaluation (MT09). **Official Release of Results**. Disponível em <http://www.itl.nist.gov/iad/mig/tests/mt/2009/ResultsRelease/>.

ORLANDI, Eni. Exterioridade e ideologia. **Cadernos de Estudos Linguísticos**, Campinas, (30): 27-33, Jan./Jun. 1996.

_____. **O que é linguística**. São Paulo: Brasiliense, 2009.

PÊCHEUX, Michel. **Papel da Memória**. Campinas, SP: Pontes, 1999

PUTNAM, Hilary. Minds and machines. In **Mind, Language and Reality**; Philosophical Papers — volume 2. Cambridge, MA: Cambridge University Press, 1995. pp. 362-385.

RUMELHART, David E. The architecture of mind: A connectionist approach. In: POSNER, Michel I. (Ed.). **Foundations of cognitive science**. Cambridge, MA: Bradford/MIT Press, 1996. pp. 133-159.

SEARLE, John. "**Minds, Brains and Programs**", Behavioral and Brain Sciences 3 (3): 417-457, 1980.

TURING, Alan. Computadores e inteligência. In: EPSTEIN, Isaac (org.). **Cibernética e Comunicação**. São Paulo: Cultrix, 1973[1950].

Para citar essa obra:

MARTINS, Ronaldo; SANTOS, Mírian. Podem as máquinas falar?. In. DIAS, Cristiane. **Formas de mobilidade no espaço e-urbano: sentido e materialidade digital** [online]. Série e-urbano. Vol. 2, 2013, Consultada no Portal Labeurb – <http://www.labeurb.unicamp.br/livroEurbano/> Laboratório de Estudos Urbanos – LABEURB/Núcleo de Desenvolvimento da Criatividade – NUDECRI, Universidade Estadual de Campinas – UNICAMP.

Endereço:

LABEURB - LABORATÓRIO DE ESTUDOS URBANOS

UNICAMP/COGEN / NUDECRI

CAIXA POSTAL 6166

Campinas/SP - Brasil

CEP 13083-892

Fone/ Fax: (19) 3521-7900

www.labeurb.unicamp.br/contato