

ENSAIO TEÓRICO

A RESPONSABILIDADE DOS LINGUISTAS NO COMBATE À MANIPULAÇÃO DA OPINIÃO PÚBLICA VIA INTERNET

Eleonora ALBANO  

Departamento de Linguística – Universidade Estadual
de Campinas (UNICAMP) | CNPq

RESUMO

Este artigo busca inspiração no ensaio de Chomsky (1967) sobre a responsabilidade dos intelectuais nas guerras imperialistas dos EUA para conclamar os linguistas a denunciar o uso das tecnologias de fala e linguagem nos mecanismos de controle da opinião pública do capitalismo atual. Para tanto, analisa a expansão e popularização dessa tecnologia como passo decisivo na consolidação do capitalismo digital do século XXI. Aponta, ainda, a sua mercantilização como fonte de mudanças no mercado de trabalho global conducentes à precarização em todos os campos, inclusive a ciência. Expõe, por fim, a exploração da mão de obra científica precarizada pelas gigantes da tecnologia digital como um dos principais meios da captura de dados pessoais via internet – um negócio ainda à espera de regulamentação.

ABSTRACT

This work seeks inspiration in Chomsky's 1967 essay on the responsibility of intellectuals in the US imperialist wars to urge linguists to denounce the use of speech and language technologies in the mechanisms of current capitalism for controlling public opinion. To this end, it analyzes the expansion and popularization of this technology as a decisive step for the consolidation of 21st century digital capitalism. It also points to its commodification as a source of changes in the global labor market



OPEN ACCESS

EDITADO POR

– Claudia Wanderley (UNICAMP)
– Anna Christina Bentes (UNICAMP)

REVISADO POR

– Roberto Camacho (UNESP)
– Rodrigo Borba (UFRJ)

DATAS

– Recebido: 09/04/2020
– Aceito: 05/10/2020
– Publicado: 11/11/2020

COMO CITAR

ALBANO, Eleonora (2020). A responsabilidade dos linguistas no combate à manipulação da opinião pública via Internet. *Cadernos de Linguística*, v. 1, n. 4, p. 01-17.

leading to loss of job security in all fields, including science. Finally, it describes the exploitation of precarious scientific labor by the giants of digital technology as one of the main means of capturing personal data via the internet, a business still awaiting regulation.

PALAVRAS-CHAVE

Tecnologias de Fala e Linguagem; Fraudes de Internet;
Vigilância de Internet.

KEYWORDS

Speech and Language Technology; Internet Fraud;
Internet Surveillance.

À GUIA DE INTRODUÇÃO: O ENSAIO DE CHOMSKY SOBRE A RESPONSABILIDADE DOS INTELECTUAIS

Ao apontar a responsabilidade de eminentes intelectuais estadunidenses em legitimar as invasões de Cuba e do Vietnã, Chomsky (1967) dirigia-se a um público bem informado que incluía muitos dos seus colegas acadêmicos, tanto das Ciências Humanas e Sociais como das Ciências Exatas e Naturais. Naquele momento, seria difícil imaginar que o conceito de ciência como empreendimento intelectual coletivo poderia mudar tão drasticamente em apenas algumas décadas.

Este artigo sugere que essa mudança se deve à gradual adesão de boa parte da comunidade científica a um dos dois “perfis” intelectuais extraídos por Chomsky da retórica de Irving Kristol, famoso intelectual¹ coetâneo alinhado à política externa dos EUA. Kristol definia os intelectuais alinhados como “especialistas ‘responsáveis e não ideológicos’, que ‘dão conselhos sobre questões táticas’”. Os não-alinhados, por sua vez, eram definidos como ‘tipos ideológicos irresponsáveis’, que ‘arengam sobre princípios e se afligem com questões morais e direitos humanos’” (CHOMSKY, 1967, p. 12, tradução minha). Chomsky demonstra claramente o viés ideológico dos dois “perfis” e chama intelectuais e cientistas à responsabilidade de defender os direitos humanos.

Não obstante, uma substancial massa pensante dos EUA e do mundo acabou por aderir inadvertidamente ao primeiro perfil. Isso decorre diretamente da crescente participação da ciência nos rumos do capitalismo. A sociedade do conhecimento contemporânea – inicialmente idealizada como fonte de bem-estar propiciado pelo acúmulo de “capital intelectual” – tornou-se o principal motor da economia mundial, sujeitando-se, portanto, à lógica da acumulação de lucro.

Sob a pressão dos mercados, a curiosidade dos cientistas foi subjugada e canalizada para uma indistinção crescente entre as ciências básicas e aplicadas. Nesse cenário, a especialização transformou muitos indivíduos cientificamente treinados em técnicos altamente qualificados, sem interesse nos fundamentos de seus campos. O modelo do cientista de sucesso deixou de ser o do desbravador de novos terrenos para ser o do tecnocrata dogmático e ambicioso.

O objetivo deste artigo é chamar a atenção dos linguistas, especialmente os brasileiros, para a responsabilidade da área em entender e combater a ameaça que essa situação representa para a autonomia e até para a sobrevivência do esforço intelectual humano.

A participação de linguistas – mesmo que involuntária – na produção de ferramentas indispensáveis às fraudes de internet – noticiosas (*fake news*) ou outras – constitui

¹ Ensaísta nova-iorquino, ex-trotskista, considerado pai do neoconservadorismo.

um bom motivo para refletirmos sobre as armadilhas que rondam a produção científica da área. O investimento em Ciências Humanas realizado pelo Brasil durante a expansão e democratização das universidades federais tornou-o um dos países mais aptos a abraçar essa discussão. Diferentemente dos países ricos, onde a maioria dos linguistas já tem um perfil tecnocrático, o Brasil ainda possui uma substancial massa crítica de humanistas atuando no campo.

A tarefa é urgente, na medida em que muitos linguistas, assim como os demais cientistas, não se dão conta de que o uso dos seus resultados para fins políticos é tão deletério para a classe quanto para a população em geral. Diferentemente dos cúmplices das guerras imperiais do século XX denunciados por Chomsky, os cientistas de hoje raramente vislumbram as implicações de médio e longo prazo da cultura acadêmica ao seu redor.

A liquefação de valores, costumes, empregos e instituições – conforme a sugestiva metáfora de Bauman (2007) – levou a comunidade científica a se render ao conformismo geral e adotar uma visão de ciência como empresa global estratificada, na qual quadros recrutados no exército industrial de reserva² da ciência agem como técnicos altamente qualificados, cumprindo ordens de uma elite de tecnocratas a serviço do todo-poderoso mercado.

Um aspecto pouco estudado do fenômeno é o papel das tecnologias de fala e linguagem nos mecanismos de controle da opinião pública do capitalismo atual. Para entender o uso das ferramentas linguísticas nesses mecanismos, que incluem as fraudes de texto e vídeo difundidas na internet, é preciso entender que a sua expansão e popularização responde a demandas do capitalismo do século XXI. É preciso também entender como a sua mercantilização precipitou as mudanças no mercado de trabalho global que levaram à precarização em todos os campos, inclusive a ciência. É preciso, ainda, entender como a exploração dessa mão de obra viabilizou certos mecanismos, ainda não regulados, de manipulação do público via internet nos últimos 20 anos.

1. O CAPITALISMO ATUAL E A COMUNICAÇÃO LINGUÍSTICA

Como apontam Davis e Walsh (2017), o neoliberalismo e a financeirização não são o mesmo fenômeno, embora convirjam em antagonizar o estado de bem-estar social. Enquanto o neoliberalismo busca a desregulação dos mercados, a financeirização busca a supremacia do mercado financeiro – facilitada, obviamente, pelas políticas neoliberais. Não cabe aqui discutir as origens dessa aliança hostil ao trabalho, já analisada por

² Termo marxista que designa a força de trabalho que excede a necessidade da produção, regulando o valor do salário.

competentes cientistas sociais (DASGUPTA, 2013; BARRADAS, 2019; entre outros). Cabe apenas indagar como o seu compromisso com a automação afeta a produção de conhecimento nas ciências da linguagem.

O fato mais importante a esse respeito é que, a partir da década de 1960, duas áreas da Ciência da Computação implicadas nesse novo capitalismo passaram a se abeberar diretamente da linguística formal. A primeira é o processamento de linguagem natural, dedicada à análise de grandes corpora envolvidos em sistemas tais como a tradução automática e a interação homem-máquina. A segunda é a tecnologia de fala, que desenvolve sistemas de síntese e reconhecimento em diferentes línguas para as mais variadas aplicações.

Essas áreas evidentemente se cruzam, na medida em que os conversores texto-fala e fala-texto usam analisadores gramaticais para segmentar, depreender e hierarquizar estruturas prosódicas, gramaticais e/ou discursivas, tanto do lado da fala como do lado do texto. Da mesma forma, os analisadores de conteúdo usam informações fonéticas e prosódicas para segmentar, depreender e hierarquizar estruturas gramaticais, semânticas e/ou discursivas em textos orais.

A interação entre as duas áreas se materializa nos recursos de síntese e reconhecimento de fala embutidos nos nossos computadores e telefones celulares. É fácil notar que os anúncios veiculados por esses dispositivos são disparados por algoritmos capazes de direcionar conteúdos a públicos específicos, como ocorre na comunicação fraudulenta na internet. As informações angariadas para tanto são coletadas pelo registro de cliques e/ou buscas escritas ou faladas em browsers. Isso permite detectar, classificar e armazenar os perfis dos usuários rastreados, tornando a sua privacidade vulnerável.

Passemos, então, às formas de regulação do trabalho científico e empresarial que abriram caminho às formas de controle em questão e às fraudes relacionadas. Cabe observar, a propósito, que, diferentemente das ciências duras e/ou exatas, as Ciências Humanas só recentemente se voltaram para o mercado de trabalho da tecnologia digital. Dentre elas, a Linguística ocupa lugar de destaque, por contribuir decisivamente para viabilizar a análise automática de corpora úteis a quaisquer projetos – científicos ou outros.

A indústria das telecomunicações, berço das tecnologias de fala e linguagem, foi uma das primeiras a absorver o excesso de mão de obra científica decorrente da saturação dos postos acadêmicos nos países ricos. Nas últimas décadas do século XX, alguns jovens cientistas emigraram em busca de oportunidades no exterior. Outros se empregaram na indústria dos seus países. Ao mesmo tempo, a globalização e a demanda de comunicação multi-língue na internet incrementaram as vagas científicas da indústria no mundo inteiro.

2. A CADEIA PRODUTIVA FRAGMENTADA DAS TECNOLOGIAS DE FALA E LINGUAGEM

As atuais tecnologias de fala e linguagem envolvem uma fragmentação radical do objeto linguístico, assim como das equipes envolvidas na sua análise. Os mais jovens do campo talvez sequer imaginem que o ambiente de trabalho das gerações anteriores era bem menos especializado que o seu. A fragmentação e a especialização surgiram na última década do século XX em resposta à demanda de um mercado cada vez mais dominado pelos gigantes da tecnologia digital.

Antes de serem cobiçadas pelo mercado, as máquinas falantes e/ou ouvintes instigaram por séculos a imaginação de físicos, matemáticos, engenheiros e fisiologistas – bem como de linguistas, filólogos e professores de línguas. Assim, os primeiros sistemas de síntese e reconhecimento de fala eram realistas, i.e., supunham uma análise dos processos de produção e percepção humanas em componentes regrados e antropomórficos, i.e., capazes de emular as funções – não necessariamente a forma – dos órgãos envolvidos.

Recentemente, porém, essas tecnologias foram incorporadas a sistemas operacionais que exigem cada vez mais rapidez e eficiência para atender às demandas dos usuários. Isso levou o campo inteiro a aderir rapidamente àquilo que os cientistas da computação chamam de força bruta³. Alguns exemplos ajudarão a mostrar como a disseminação desse estilo de computação desviou significativamente os rumos do campo dos interesses da pesquisa básica em produção e percepção da fala.

A primeira máquina falante de sucesso surgiu no século XVIII. O seu inventor foi um engenheiro austro-húngaro que se assinava Wolfgang Ritter von Kempelen ao escrever em alemão (DUDLEY e TARNOCZY, 1950). A sua concepção da física da fala era bastante avançada para a época. Uma palheta vibrátil e um fole flexível simulavam, respectivamente, a laringe e o trato vocal, reproduzindo as suas propriedades acústicas essenciais. As vibrações da palheta e as variações de pressão no fole, obtidas via controle manual de teclas e pedais, emulavam razoavelmente o que hoje se entende como as fontes sonoras e os filtros acústicos da fala (FANT, 1960).

Cabe notar que o realismo dessa análise é físico e não fisiológico. Pressupõe que é a forma das cavidades por onde passa o ar saído dos pulmões que modula os tons produzidos na laringe, bem como os ruídos produzidos por constrições no seu interior. Em contrapartida, os chamados sintetizadores articulatórios aplicam o mesmo princípio a modelos fisiologicamente realistas dos órgãos da fala. Isso exige informações precisas sobre a sua

³ Trata-se de uma forma de programação que, ao invés de analisar a tarefa, soluciona problemas por busca exaustiva graças a um alto poder de armazenamento e computação.

mecânica e aerodinâmica, o que torna a modelagem mais custosa, dada a dificuldade de estimar todos os parâmetros envolvidos. Historicamente, esses modelos têm sido usados apenas para investigar questões articulatórias.

O mais exitoso fruto do realismo físico deve-se a Klatt (1980) e se baseia diretamente na Teoria Acústica da Produção da Fala (FANT, op. cit.). Trata-se de um sintetizador capaz de gerar e interconectar vários tipos de fontes sonoras e filtros acústicos como os da fala. Um conjunto de geradores de tons e ruídos alimenta um conjunto de filtros passa-bandas em faixas de frequência diversas. Os filtros podem se conectar em cascata (i.e., sequencialmente) ou em paralelo. Grosso modo, o primeiro método funciona melhor com vogais e o segundo, com consoantes.

Uma das dificuldades desse tipo de sintetizador é que a qualidade da sua saída depende de uma modelagem minuciosa da estrutura fonético-acústica da língua-alvo. Embora a literatura de então não oferecesse respaldo suficiente, Klatt e sua equipe conduziram todos os estudos necessários para viabilizar a sua implementação em inglês, que alcançou resultados impressionantes. O software resultante obteve grande sucesso de mercado, até começar a ser preterido por formas de síntese economicamente mais vantajosas.

Note-se que esse projeto tecnológico tinha uma afinidade inalienável com a pesquisa básica. Em princípio, a parametrização de um sintetizador de Klatt em qualquer língua supõe um estudo profundo da sua estrutura fonético-acústica. Já não se pode dizer o mesmo da síntese concatenativa de hoje, que depende crucialmente da força bruta e da inteligência artificial.

Num passado recente, a síntese concatenativa requeria estudos prospectivos para avaliar o tamanho e o número das unidades a concatenar. A montagem de um banco de dados eficaz e econômico passava por equilibrar unidades curtas (i.e., um ou dois fones – chamados monofones e difones) e longas (i. e., vários fones – chamados polifones). Nesse primeiro momento, o método fomentava uma colaboração estreita entre linguistas e engenheiros na busca de soluções regradas para os problemas de seleção e concatenação das unidades.

As unidades longas ensejavam menos pontos de concatenação e, portanto, favoreciam a naturalidade, preservando a coarticulação no seu interior. Por outro lado, consumiam muito espaço de memória, por serem mais numerosas. Já as unidades curtas consumiam menos memória, mas dificultavam os procedimentos de coleta, etiquetagem e concatenação do sinal.

À primeira vista, parece extraordinário que, em pouco tempo, a força bruta tenha tornado desnecessárias as regras de seleção e concatenação de unidades. Isso não quer dizer que supere a engenhosidade das soluções regradas desenvolvidas antes. O seu êxito baseia-se num poder computacional gigantesco, que permite buscar

instantaneamente unidades concatenativas ótimas num enorme banco de dados de fala exaustivamente anotado.

Para entender esse estado de coisas, uma importante distinção se faz necessária. Enquanto a exaustividade da busca depende exclusivamente de técnicas de computação, a exaustividade da anotação do banco de dados depende crucialmente de um conjunto substancial de anotações humanas, em geral feitas por linguistas.

É que nelas reside a base para a expansão do banco. Só um corpus contendo suficiente anotação humana é capaz de incorporar novos dados por meio de inteligência artificial. Por exemplo, as informações dependentes de juízos pragmáticos e/ou de conhecimento de mundo são geralmente etiquetadas manualmente, ainda que possam, mais tarde, ser processadas por inteligência artificial.

Não é diferente a situação nos sistemas de reconhecimento de fala.

Durante algum tempo, a hegemonia do campo foi disputada entre modelos de base física e modelos de base estatística (RAO, 1996). Os primeiros continham regras para buscar correspondências entre propriedades acústicas segmentadas no sinal de entrada e unidades armazenadas num banco transcrito e etiquetado. Os segundos tratavam as unidades linguísticas como estados de um modelo oculto de markov (*hidden markov model*), modelando-as pela avaliação probabilística da sua adequação ao contexto. Essa técnica não se baseia em anotações fonético-acústicas e, sim, em parâmetros acústicos genéricos.

Muitos projetos combinavam as duas abordagens em sistemas híbridos. Porém, em qualquer caso, a otimização do desempenho do banco exigia um treinamento bastante longo. A tentativa de reduzi-lo levou à incorporação de modelos de unidades superiores (p. ex., a palavra, a frase). A partir daí, a inteligência artificial passou a contribuir significativamente para a implementação e integração dos componentes envolvidos.

Atualmente esses componentes costumam ser três, chamados de modelos acústico, de pronúncia e de língua (XIONG et al., 2018). O modelo acústico extrai fragmentos do espectro e/ou forma de onda, combinando-os e comparando-os a padrões armazenados até identificar um fone plausível. O modelo de pronúncia usa o mesmo método para testar a probabilidade de os fones identificados formarem palavras. O modelo de língua faz o mesmo para a probabilidade de as palavras identificadas formarem frases. A intervenção da inteligência artificial permitiu calcular os três modelos conjuntamente, rendendo melhores taxas de acerto e menores tempos de treinamento.

Em conjunto, essas mudanças no estado da arte das tecnologias de fala e linguagem provocaram uma guinada científica e econômica radical no campo. A competitividade das empresas passou a depender da sua capacidade de ampliar recursos computacionais e gerir uma política de recursos humanos baseada em terceirização.

Assim, os gigantes da internet e/ou da computação pessoal, tais como o Google, a Apple e a Microsoft, passaram a dominar o mercado, antes liderado pela indústria de

telecomunicações. É importante notar que os novos protagonistas se dedicam apenas à pesquisa de novos produtos, mas não ao seu desenvolvimento. Jovens e promissores especialistas, alocados a departamentos pequenos e estimulantes, exploram, testam e afinam respostas de ponta aos desafios do campo. Uma vez bem delineado o protótipo, o seu desenvolvimento é repassado a empresas terceirizadas para fins comerciais.

Trata-se de pequenas empresas, espalhadas pelo mundo, que se dedicam a executar partes especializadas do projeto da contratante. Assim, por exemplo, uma pode contratar linguistas temporariamente para alimentar um banco de dados com informações altamente especializadas sobre os segmentos fônicos, tais como transcrição fonética, segmentação e etiquetagem de pistas no sinal de fala. Outra pode tratar de delimitar, hierarquizar e classificar constituintes prosódicos. Outra, ainda, pode tratar de delimitar, hierarquizar e classificar constituintes gramaticais.

Em qualquer dessas atividades, os linguistas e demais cientistas envolvidos só conhecem o projeto maior da contratante em linhas gerais. Além disso, desconhecem quais outras empresas colaboram na alimentação dos bancos de dados. O mesmo ocorre com as tarefas de outras áreas. Por exemplo, diferentes encomendas de algoritmos podem ser distribuídas entre empresas diferentes, sediadas em cidades ou mesmo países diferentes.

Essa divisão de trabalho inoculou definitivamente o fordismo e o taylorismo⁴ na atividade científica. Como na indústria em geral, assume-se que a fragmentação e a especialização aumentam a eficiência, elevando as taxas de produtividade.

Sublinhe-se que essa prática relega a formação científica a um mero preparo técnico de alto nível. Assim, ainda que o contato com as ideias inovadoras da matriz crie ambientes sedutores e estimulantes, os desafios limitam-se à melhoria do desempenho dos sistemas em construção. Analogamente, o pensamento crítico limita-se à tentativa de superar resultados anteriores.

A falta de transparência dessa política laboral abre caminho ao uso de milhares de cientistas como inocentes úteis na fabricação de informação fraudulenta para circulação na *internet*.

Um caso exemplar é a manipulação da opinião pública pela empresa britânica *Cambridge Analytica* no Brexit e nas eleições presidenciais dos EUA em 2016.⁵ A firma, hoje nominalmente extinta⁶, usava mineração de dados e inteligência artificial para respaldar campanhas de marketing. O seu site oferecia serviços que iam desde montar a campanha

4 O fordismo fragmenta as tarefas e as executa por meio de linhas de montagem. O taylorismo busca a eficiência pelo foco, a disciplina e a especialização.

5 Rodrigo Borba sugeriu que a esta se adicionasse uma discussão de casos brasileiros. Seria, de fato, muito útil, mas falta documentação. O escândalo da *Cambridge Analytica* gerou reportagens, documentários, artigos e livros. A participação de cientistas na fraude eleitoral das eleições brasileiras de 2018 ainda precisa ser rastreada e documentada. Fica aqui registrado o desafio.

6 Na verdade, só o nome foi extinto. A firma deixou sequelas com outros nomes espalhadas pelo mundo.

propriamente dita a assessorar empresas interessadas em manipular dados para “mudar o comportamento do usuário”.

As fraudes tiveram participação de dois grupos de especialistas em tecnologia de fala e linguagem. O primeiro consiste nos participantes diretos da concepção do projeto, sob a batuta de Robert Mercer⁷, bilionário estadunidense que investia pesadamente em campanhas de extrema direita. O segundo consiste nas vítimas de violação de segurança das empresas terceirizadas pelo Facebook, a saber: os trabalhadores temporários envolvidos na análise e organização dos perfis de usuários.

Muitos linguistas, psicólogos, sociólogos, demógrafos e publicitários criam taxonomias de conteúdos de redes sociais sem desconfiar que os resultados podem ser usados para fins escusos tais como identificar votantes vulneráveis em eleições. Aparentemente, o Facebook usa essa análise para fins comerciais sem conflito com a lei⁸.

Não obstante, o modo vigente de distribuição do trabalho na cadeia produtiva praticamente inviabiliza o esclarecimento detalhado da fraude. Mesmo uma perícia a mando judicial tropeçaria na dificuldade de rastrear a origem das anotações. Um obstáculo é a sua própria trajetória fragmentária. Outro é a sua reorganização por meio de uma das mais notáveis ferramentas da inteligência artificial.

Trata-se de um novo tipo de aprendizagem de máquina, conhecido como aprendizagem profunda (ALOM et al., 2018). Os algoritmos envolvidos têm o poder de reorganizar e otimizar a própria saída a partir de uma entrada anotada em vários níveis, de início alimentada manualmente por especialistas. O tratamento “profundo” estende a anotação e faz emergir estruturas novas e insuspeitadas. A chave está em combinar a expertise de uma vasta e precisa anotação manual com a capacidade do algoritmo de construir e organizar redes de alta complexidade a partir de estatísticas das relações entre os conteúdos anotados.

A aprendizagem profunda é de grande valia para todas as áreas das ciências, básicas ou aplicadas. Pode servir a diversos fins de diagnóstico, pesquisa e prognóstico. Infelizmente, teve o mesmo destino de outras invenções históricas, i.e., caiu nas mãos de grupos mal-intencionados. Entretanto, seria simplista rotular esse fato como uma “conspiração do mal”. Na verdade, ele reflete a ideologia de um sistema político-econômico que dissemina o obscurantismo, desqualificando subrepticamente a ciência básica – sobretudo, as Humanidades e as Ciências Humanas e Sociais.

7 Trata-se de um cientista da computação respeitado, que enriqueceu graças às suas contribuições à linguística computacional.

8 As leis anteriores à era digital não previam casos de fraude por esse meio.

Esse pensamento simplista, encontrável tanto à direita como à esquerda, insiste em ignorar o fato de que a maioria das grandes descobertas tecnológicas se origina em conceitos e métodos nascidos da reflexão sobre questões de pesquisa básica.

Tal é o caso da aprendizagem profunda, derivada de outra notável família de algoritmos, inventada por um linguista no calor de uma controvérsia na área da sintaxe. Jeffrey Elman (1990), saudoso pioneiro da Ciência Cognitiva, concebeu as chamadas redes recorrentes⁹ no intuito de superar as dificuldades das redes neurais da época em lidar com séries e hierarquias de dados. O objetivo era tomar a sua capacidade de aprender estruturas sintáticas do inglês como evidência contra a tese gerativista do inatismo da gramática universal (CHOMSKY, 1986).

Embora a controvérsia tenha persistido à revelia do sucesso de Elman, cabe sublinhar que uma ferramenta capaz de mudar os rumos da computação surgiu de um debate epistemológico na Linguística. Aliás, não faltam exemplos análogos na história do campo. Basta lembrar que a gramática gerativa – admita-o ou não – criou formalismos fáceis de computar. De qualquer modo, o fato histórico mais emblemático para a presente discussão é que a Teoria Acústica da Produção da Fala (FANT, op. cit.), marco fundador da tecnologia de fala, se enraíza na busca de respaldo para a teoria dos traços distintivos (JAKOBSON; FANT; HALLE, 1953).

3. O PAPEL DAS TECNOLOGIAS DE FALA E LINGUAGEM NAS FRAUDES DE TEXTO E VÍDEO

A tecnologia de fala e linguagem tem uma afinidade natural e bastante benigna com alterações automáticas e semiautomáticas de arquivos de texto e vídeo. Por exemplo, vem sendo incorporada, cada vez mais, a equipamentos de auxílio ao ensino ou de comunicação aumentativa para portadores de necessidades especiais. Também está presente nos corretores ortográficos e gramaticais dos editores de texto, nos leitores de textos para cegos e em softwares de ensino de leitura labial a surdos. Quem não se lembra, a propósito, da engraçada voz do computador do notável Stephen Hawking?

Também na indústria do entretenimento a presença dessas tecnologias é marcante. Respondem, por exemplo, pela verossimilhança das expressões faciais dos personagens do cinema de animação. Mesmo com atores experientes, a dublagem seria ineficiente se o

⁹ Uma rede recorrente simples (ou de Elman) compõe-se de três camadas: entrada, saída e intermediária (ou oculta). Essa se conecta a um conjunto de unidades de contexto, que copiam os pesos das suas conexões ao longo do tempo, formando uma memória. A dinâmica dessa rede é capaz de lidar com séries temporais e mesmo com hierarquias simples, prestando-se à análise de alguns fenômenos sintáticos.

movimento da face discrepasse muito do esperado, apesar de a expectativa ser apenas vagamente consciente.

O necessário ajuste de naturalidade é fruto da pesquisa básica em síntese e reconhecimento de expressões faciais da fala (e.g., HASEGAWA et al., 2007). Voltada inicialmente para esclarecer o papel da visão na compreensão da linguagem oral (MASSARO, 2006), essa tecnologia não tardou a encontrar múltiplas aplicações clínicas e educacionais.

O processamento de linguagem natural, por sua vez, serve tanto para agrupar perfis semelhantes, tal como já visto, quanto para criar textos fraudados verossímeis. Além de análises semânticas, é capaz de produzir análises estilísticas extraindo traços característicos da escrita de um autor. Portanto, é uma tecnologia apta, em princípio, a auxiliar impostores a disfarçar a falsa autoria dos textos veiculados na *internet*.

Aparentemente, porém, a maioria dos infratores dispensa disfarces, porque aposta na credulidade dos leitores. Isso fica evidente na desinformação sobre a Covid-19 circulada diariamente no Twitter e no WhatsApp. Para descartá-la, bastaria o leitor comparar o material recebido ao conteúdo dos sites das instituições e/ou autoridades invocadas. No entanto, a maior parte do público sequer se dá conta da necessidade dessa precaução.

Para influenciar a vasta maioria que se deixa levar por um simples logo ou qualquer outro sinal banal de identificação, há hoje um recurso extremamente perigoso – porque convincente. Trata-se da chamada falsificação profunda, fabricada com arquivos de áudio e vídeo por meio da aprendizagem profunda. Consiste em transferir gestos e expressões faciais entre gravações em vídeo de pessoas diferentes.

O algoritmo sobrepõe um conteúdo de áudio e vídeo encenado por um ator a uma gravação compatível da vítima. A compatibilidade é fácil de obter: basta a encenação reproduzir a composição gráfica da cena original, i.e., closes, posição do corpo, enquadramento, etc. A transferência dos gestos se faz por computação gráfica, a partir de grades classificatórias já existentes para fins diversos, todos benignos.

Embora as suas raízes remontem à pesquisa sobre reconhecimento visual da fala ou da expressão corporal – ou, ainda, à pesquisa sobre reconhecimento de línguas de sinais (CHEOK; OMAR; JAWARD, 2019) –, as referidas grades já receberam aportes substanciais da indústria do entretenimento. Ultimamente, tem-se investido muito em traspor a movimentação de atores para personagens de animação, assim como em produzir efeitos especiais de movimentação em cena por meio de computação gráfica.

A falsificação mais difícil é a da voz, pois o uso da síntese concatenativa requer um corpus da voz da vítima de tamanho suficiente para extrair as unidades necessárias à composição da fala imputada. Porém, mesmo em caso positivo, o resultado esbarra na dificuldade de sintetizar uma prosódia verossímil. Qualquer modelo prosódico plausível integra efeitos gramaticais, peculiaridades individuais e traços estilísticos, tais como sotaque

regional e profissional. Obviamente, não existem corpora anotados das vozes a emular¹⁰. As bases de dados do campo costumam ser gravadas por locutores profissionais contratados por projetos científicos ou tecnológicos.

Apesar de se prestar à falsificação, a tecnologia de fala e linguagem gera, em princípio, mais benefícios que malefícios. Basta lembrar a contribuição da prótese vocal de Stephen Hawking para a discussão e disseminação das suas ideias científicas. O verdadeiro problema é que todos os malefícios das tecnologias concebidas para fins pacíficos têm origem ideológica: decorrem, na verdade, das ideias disseminadas pela máquina de propaganda de regimes políticos autoritários. Portanto, é impossível combater tais abusos¹¹ sem combater a mais deletéria arma desses regimes, a saber: o obscurantismo.

4. ARMAS E RESPONSABILIDADES DA LINGUÍSTICA NO COMBATE AO OBSCURANTISMO

Como se sabe, as formas de apropriação das tecnologias digitais pelo capitalismo neoliberal conduziram à globalização e à financeirização. Com o avanço do desemprego e da desigualdade, o medo e a ansiedade aumentaram e abriram caminho à ascensão de regimes fascistas em todo o mundo. Isso se agravou nas duas últimas décadas, devido à expansão das grandes plataformas da tecnologia digital. Como demonstra Zuboff (2019), nesse período elas subrepticamente erigiram a base daquilo que ficou conhecido como capitalismo de vigilância.

Trata-se da oferta de produtos e serviços gratuitos ou baratos pela internet “em troca” da captura, insuspeitada e, portanto, não consentida, de dados privados do usuário. Tudo começou com buscas, cliques, erros de digitação e contatos de e-mail. Logo, porém, passou a incluir fotos, gravações de áudio e vídeo, contatos em redes sociais, padrão de consumo, localização, trajetos, velocidade de locomoção, etc.

Constrangendo os usuários a aceitar contratos de privacidade vagos e obscuros, essas corporações faturam fortunas vendendo informações personalizadas aos seus parceiros comerciais. Oferecem, além disso, serviços de persuasão e modificação do comportamento, a fim de adequar anúncios e demais formas de publicidade aos estados de espírito minerados desses dados.

10 Um bom exemplo é o vídeo montado pelo ator, diretor e roteirista Jordan Peele para demonstrar o funcionamento da falsificação profunda e denunciar o seu risco. O “alvo” foi uma celebridade, Barack Obama, de quem não faltavam gravações de áudio e de vídeo. Apesar de didático, o resultado soa falso, justamente por causa da entoação, diretamente transferida da voz do ator para a do político. Confira-se em: <https://www.buzzfeed.com/br/craigsilverman/como-identificar-deepfake-video-obama-peelee>.

11 Bons exemplos associados à segunda guerra mundial são a aviação e a fissão nuclear.

O dispositivo que viabiliza essa pilhagem é o telefone celular. Além de registrar a rotina do usuário no seu sistema operacional, oferece uma enorme gama de aplicativos gratuitos capazes de colher indicadores de estados físicos e mentais. São exemplos: taxas de digitação e locomoção, padrão de alimentação e atividade física, comportamento na condução de veículos, índices de saúde e bem estar, etc.

O público entrega esses dados ao cair na sedução dos aplicativos disponíveis, superestimando a sua utilidade. Outra oferta sedutora são os jogos, geralmente gratuitos, que registram uma infinidade de dados sobre a atenção, o raciocínio e a agilidade dos jogadores. Cabe lembrar que esses meios de vigilância respondem por boa parte das patologias da atenção e concentração devidas ao uso intensivo da internet (PETRY e O'BRIEN, 2013). O embotamento progressivo do usuário rouba-lhe a lucidez, atraindo-o para distrações espúrias que subtraem a sua privacidade.

A sedução joga, ainda, com outro fator cada vez mais presente na sociedade atual: o medo da exclusão, cuja relação com o aumento da desigualdade é bem conhecida. Atualmente, ele ronda todos os estratos sociais, exceto, talvez, os 1% super-ricos. Nesse contexto, a propaganda maciça sobre a conectividade como meio de inclusão reforça, na verdade, o medo da exclusão. Essa tática conta com a cumplicidade dos bancos e instituições comerciais e governamentais, que validam transações via telefone celular, forçando os cidadãos a se render à vigilância.

Vê-se, portanto, que as condições para a manipulação da opinião pública já estavam dadas quando começaram a ser usadas explicitamente para fins políticos.

Depois do escândalo da *Cambridge Analytica*, a imprensa esclarecida passou a divulgar muitas denúncias a esse respeito. No entanto, parece que ninguém chegou, até agora, ao âmago da questão. A razão é que o obscurantismo embutido nas entranhas do sistema é difícil de detectar. Enquanto a imprensa se ocupa do terraplanismo e do movimento antivacinas, cujos riscos são óbvios, a internet segue reforçando o mito da máquina inteligente, ignorando a participação humana nessa inteligência.

Os quadros egressos do exército de reserva da ciência permanecem invisíveis, enquanto contribuem, com maior ou menor consciência, para viabilizar manipulações comportamentais deletérias à sociedade. As tarefas das equipes envolvidas na vigilância digital são em geral tão específicas que só vagamente apontam para o projeto maior das contratantes. Haja vista as confissões de remorso daqueles que se deram conta dos desvios éticos dos seus empregadores. Um exemplo bem conhecido é o do analista de dados Christopher Wylie, um dos principais vazadores do modus operandi da *Cambridge Analytica*.

O mito da máquina superinteligente, ameaçadora para o futuro da humanidade, exerce um papel salutar nas artes em geral (NATALE; BALLATORE, 2017). Já produziu importantes reflexões nas áreas cênica, plástica e literária. Não obstante, como é de esperar, a sua base factual é nula. A inteligência artificial não é uma concorrente da

inteligência humana. É um complemento poderoso que viabiliza o tratamento de grandes quantidades de dados e a sua organização em redes complexas, evidenciando relações inicialmente invisíveis para os pesquisadores.

A alimentação humana é mandatória em dois níveis do processamento de qualquer aprendizagem de máquina: os mais baixos e os mais altos. Os primeiros incluem tarefas tais como etiquetar dados sensíveis (i.e., visuais, auditivos, de movimento, etc.) de modo que façam sentido no quadro de referência da análise. Os segundos incluem tarefas tais como rotular os nós superiores da rede obtida à luz do referido quadro de referência.

Apesar da popularidade dos likes, emojis e demais ícones fáceis de processar automaticamente, qualquer hipótese minimamente refinada sobre o estado de espírito do usuário exige recurso a fontes cuja análise supõe bancos de dados específicos. Na verdade, as informações pessoais que interessam à clientela mais exigente são obtidas com base em resultados de análises especializadas de arquivos de texto, voz ou vídeo.

Ocorre que os níveis inferiores e superiores dos bancos de texto, voz e vídeo contêm necessariamente etiquetagem feita por um time composto de foneticistas, analistas de texto/discurso, psicólogos clínicos, especialistas em motricidade, entre outros. Não há leitura automática de textos, vozes ou faces que não remeta a taxonomias e/ou hierarquias elaboradas por intérpretes humanos com a necessária expertise.

O obscurantismo está em alta, mesmo entre os mais escolarizados, porque os ideólogos da pseudociência fascista, ao mesmo tempo em que cooptam cientistas humanos para servir ao capitalismo de vigilância, espalham boatos que desqualificam a pesquisa básica em Ciências Humanas.

Tais rumores repercutem tanto em setores reacionários como em setores progressistas da sociedade. Os reacionários acusam os humanistas de improdutividade, por desconhecerem o potencial das inovações epistemológicas para o fomento da ciência e da tecnologia. Os progressistas cobram desses estudiosos um ingênuo e impertinente imediatismo na aplicação das suas concepções de mundo e de homem às causas sociais.

Haja vista os acadêmicos que rotulam de paradoxal a posição de Chomsky (p. ex., LUKIN, 2013), por aliar ideias políticas de esquerda a teorias que defendem as bases genéticas das gramáticas das línguas naturais. Isso é negar um princípio básico e inalienável da ciência, a saber: a liberdade de escolha teórica e metodológica. Paradoxal é, na verdade, a contaminação de cientistas progressistas pelo utilitarismo difundido por colegas cooptados pelo capitalismo. Chomsky segue sendo um baluarte da ciência básica, assim como das causas sociais.

Todo o exposto sugere fortemente que a Linguística tem a obrigação de denunciar publicamente os avanços do obscurantismo e combater os seus efeitos não só entre os cientistas, mas também na população em geral.

Para tanto, precisa, urgentemente, historiar e divulgar exemplos notáveis de debate epistemológico sobre a linguagem que inspiraram e fomentaram inovações nas Ciências Exatas, Naturais e Tecnológicas. Precisa, ao mesmo tempo, denunciar o recrutamento do exército de reserva das Ciências Humanas pelas plataformas digitais para colaborar na fabricação de ferramentas destinadas a manipular o comportamento de indivíduos e grupos ou a opinião pública em geral. Precisa, por fim, esclarecer por que projetos dessa natureza, apesar de dependerem crucialmente de mão de obra científica, abusam do seu excedente para precarizá-la e aviltá-la.

A denúncia deve também ressaltar que a sobrevivência da ciência, tal como a conhecemos, depende da solidez das instituições onde é praticada. Conhecimentos de potencial utilidade para o bem público precisam de instituições públicas para germinar. Assim, a defesa da ciência está intrinsecamente ligada à defesa das universidades, institutos de pesquisa e agências de fomento públicas.

Caso não tenhamos sucesso em engrossar a resistência à vigilância conscientizando colegas e alunos dos riscos aqui expostos – e outros que fatalmente surgirão –, as próximas gerações de linguistas não passarão de uma manada de executores de atividades pseudo-científicas a serviço de uma ordem social que marcha inexoravelmente para o fascismo.

REFERÊNCIAS

- ALOM, M. Z. et al. *The history began from AlexNet: a comprehensive survey on deep learning approaches*. arXiv.org > cs > arXiv:1803.01164, 2018.
- BARRADAS, R. Financialization and neoliberalism and the fall in the labor share: A panel data econometric analysis for the European union countries. *Review of Radical Political Economics*, Vol. 51(3) 383–417, 2019
- BAUMAN, Z. *Liquid times: Living in an age of uncertainty*. Cambridge: Polity Press, 2007.
- CHEOK, M.; OMAR, Z.; JAWARD, M. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning & Cybernetics*. 10, 131–153, 2019.
- CHOMSKY, N. The responsibility of intellectuals. *The New York Review of Book Special Supplement*, 23 pp., 1967.
- CHOMSKY, N. *Knowledge of language: its nature, origin and use*. New York: Praeger, 1986.
- DASGUPTA, B. *Financialization, labour market flexibility, global crisis and new imperialism - A Marxist perspective*. 2013. ffhalshs-00840831ff 2013.
- DAVIS, A.; WALSH, C. Distinguishing Financialization from Neoliberalism. *Theory, Culture & Society*, Vol. 34(5–6) 27–51 2017.
- DUDLEY, H.; TARNOCZY, T. The speaking machine of Wolfgang of von Kempelen. *Journal of the Acoustical Society of America*, 22 (2): 151 – 166, 1950.
- ELMAN, J. Finding structure in time. *Cognitive Science*, v. 14.2, pp. 179-211,1990.
- FANT, G. *Acoustic theory of speech production*. Haia: Mouton, 1960.

HASEGAWA et al. Audiovisual speech recognition with articulator positions as hidden variables. *Proceedings ICPHS XVI*, paper 1719, 2007.

JAKOBSON, R.; FANT, G.; HALLE, M. *Preliminaries to Speech Analysis*. Cambridge: The MIT Press, 1953.

KLATT, D. Software for a cascade/parallel synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995, 1980.

LUKIN, A. Journalism, ideology and linguistics: The paradox of Chomsky's linguistic legacy and his 'propaganda model'. *Journalism*, 14(1) 96-110, 2012.

MASSARO, D. A computer-animated tutor for language learning: Research and applications. In Spencer, P. E.; Marshark, M. (Eds.), *Advances in the spoken language development of deaf and hard-of-hearing children*. New York, NY: Oxford University Press, pp. 212-243, 2006.

NATALE, S.; BALLATORE, A. Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*, Vol. 26(1) 3-18, 2017.

PETRY, N. M.; O'BRIEN, C. P. Internet gaming disorder and the DSM-5 [Editorial]. *Addiction*, 108(7), 1186-1187, 2013.

RAO, P. Is rule-based acoustic phonetic speech recognition a dead end? In FUJISAKI, H. (Ed.) *Recent Research Towards Advanced Man-Machine Interface Through Spoken Language*. Amsterdam: Elsevier, pp. 160-164, 1996.

XIONG, W. et al. The Microsoft 2017 conversational speech recognition system. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, pp. 5934-5938, 2018.

ZUBOFF, S. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs, 2019.