

## **LETRAMENTO DE DADOS E SUAS POSSIBILIDADES PARA A EDUCAÇÃO CIENTÍFICA EM LÍNGUA/LINGUAGEM**

### *DATA LITERACY AND ITS POSSIBILITIES FOR LANGUAGE SCIENCE EDUCATION*

Rodrigo Esteves de Lima-Lopes  0000-0003-3681-1553

Programa Pós-Graduação em Linguística Aplicada.

Instituto de Estudos da Linguagem - Universidade Estadual de Campinas

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

rl1307@unicamp.br

*DOI:* <https://doi.org/10.5281/zenodo.8021649>

*Recebido em 15 de fevereiro de 2023*

*Aceito em 08 de maio de 2023*

We need a broader definition of literacy itself, one that includes all literate practices, regardless of medium (Jay Lemke, 2006)

**Resumo:** Este artigo tem por objetivo refletir sobre as possíveis contribuições que um modelo baseado no Letramento de Dados pode trazer para o Letramento Científico no contexto das ciências da linguagem, especialmente a Linguística Aplicada. Discute-se, inicialmente, a necessidade e justificativa de um projeto desta natureza, o qual reconhece o crescente papel que os dados têm alcançado em nossa sociedade. Parte-se, então, para uma discussão sobre os processos hegemônicos na tecnologia, com especial foco no extrativismo de dados e capitalismo de vigilância. Segue-se uma discussão sobre o letramento de dados, com especial atenção às possibilidades de resistência por ele oferecido. Reflete-se, por fim, sobre as transformações em termos de paradigma que seriam trazidos pela aplicação de um modelo de Letramento de Dados ao contexto da LA.

**Palavras-chave:** Letramento Científico, Letramento de Dados, Internet e Cidadania

**Abstract:** This article reflects on the possible contributions that a model based on Data Literacy brings to Scientific Literacy in the context of language sciences, especially Applied Linguistics. Initially, the article points out the reasons why a project of this nature would be relevant to AL, recognising the growing importance that data has achieved in society. Next, the hegemonic processes in technology, with a particular focus on data extractivism and surveillance capitalism is discussed. A problematisation on data literacy follows, with attention to the possibilities of resistance. Finally, a reflection on the paradigm transformations which a Data Literacy model would be applied to the context of AL is presented.

**Keywords:** Data Literacy, Scientific Literacy, Internet and Citizenship



## 1. Introdução

Tenho por objetivo refletir sobre as possíveis contribuições que uma abordagem baseada no *Letramento de Dados* (BHARGAVA, 2019; GUTIÉRREZ, 2019) pode trazer para o desenvolvimento de modelos de *Letramento Científico* no contexto dos Estudos da Linguagem (MOTTA-ROTH, 2011; SILVA, 2021). Neste artigo compartilho a definição de Letramento Científico proposta pela Organização para a Cooperação e Desenvolvimento Econômico (OECD, 2019), para quem uma pessoa cientificamente letrada deveria ser capaz de engajar-se em discussões e reflexões sobre ciência e seus resultados de forma crítica. Assim, um cidadão cientificamente letrado seria capaz de discutir e refletir de forma fundamentada sobre o conhecimento e tecnologia e ciência, algo que demandaria três habilidades principais: 1) reconhecer explicações sobre diversos fenômenos naturais e tecnológicos (que incluiria os linguísticos);<sup>1</sup> 2) avaliar e descrever investigações científicas, propondo formas e abordagens das pesquisas ou fenômenos; e 3) interpretar dados em uma variedade de representações (OECD, 2019, p. 15).

A motivação para a reflexão que proponho neste trabalho nasceu durante minha participação no *I Encontro – Letramento Científico na Educação Básica*, promovido pela Universidade Federal de Campina Grande, em agosto de 2022, ocasião na qual tive a oportunidade de discutir criticamente o papel da tecnologia no contexto escolar, integrando suas possibilidades de pesquisa na formação de nossos alunos. A importância da questão reside, especialmente, no papel que os *Big Data*, ou Dados Massivos em uma tradução livre, têm desempenhado em diversas áreas da sociedade, sendo largamente utilizados no mapeamento de perfis de consumidores, manipulando questões mercadológicas, criação de políticas públicas, observação de tendências político-partidárias, influenciando investimentos em pesquisa, além de nossas percepções sobre ciência e seu papel na sociedade.

Nos últimos trinta anos, deste a chamada Revolução Digital (KNELL, 2021), a produção de dados massivos tem se tornado uma realidade. Tais dados, de acordo com Kitchin (2014) devem ser caracterizados não apenas pelo seu volume, mas também por outras características que lhe são determinantes. A primeira seria a *sincronicidade*, uma vez que tais dados são produzidos em tempo real, ou seja, são fruto de processos interacionais no momento exato em ocorrem. Tais dados são *diversos*, representando diferentes facetas dos processos sociais, e exaustivos, à medida em que, na maioria dos casos, representam a totalidade da produção interacional pesquisada. Os sistemas de interação digitais também permitem que estes dados sejam *delicados* em suas representações, possibilitando uma indexação particularizada. Por fim, eles são

<sup>1</sup> Neste trabalho, utilizamos a definição de tecnologia oferecida por Flusser (2007b), para quem ela representa tudo aquilo que estende nossa capacidade humana de interação com o mundo. Algo a nós externo e manufaturado de forma não-natural, mas que passa a ser parte integrante de nossa realidade. Flusser (2007b) parte da ideia de alavanca para construir seu argumento: ela, inicialmente, estenderia a capacidade do operário, sendo que, com o seu uso e construção técnico-social dele decorrente, ela se tornaria um elemento necessário para a realização destas tarefas. Haveria assim, uma fusão entre o braço do operário e da alavanca, a ponto de não haver mais diferença entre eles. A língua para Flusser (2007b, a) caminha pela mesma seara: uma simbologia artificial que estende nosso mundo interior e codifica o universo exterior, criando simbologias que passam a funcionar como elementos representacionais e definidores de nosso próprio mundo; a língua funde-se com homem e lhe dá caminhos de interpretação do mundo. Consequentemente, tais representações, artificiais e manufaturadas pela língua, passam a ser o código-fonte (no sentido de oferecer pré-condição de existência) de outras formas de linguagem, incluindo a programação de computadores, a produção audiovisual entre outras. Não haveria o vídeo ou o podcast, destarte, se a escrita não estivesse ali como tecnologia de suporte, como é a alavanca para o operário.



*relacionais*, pois os diferentes conjuntos de dados podem ser relacionados no processo e análise, algo que lhes confere um grau elevado de *flexibilidade*.

Kitchin (2016) e boyd e Crawford (2012) levantam questões importantes sobre este fenômeno. Para o primeiro, há uma explosão de trabalhos e projetos de financiamento para a temática, que acaba por ser tornar um dos temas favoritos por diversas agências de fomento. Em muitos casos, tais trabalhos sobrepõem os resultados dessas análises baseadas em grandes quantidades de dados à teoria mobilizada para interpretá-los. Já boyd e Crawford (2012) criticam seu pretenso poder de generalização, uma vez que eles representam apenas o universo de usuários dos serviços que os geram o que, por conseguinte, invalidaria a pretensão de encará-los como uma *polaroide* da sociedade.

Tradicionalmente, há a percepção de que os Big Data são fruto das *mídias sociais*, tais como Facebook, Twitter entre diversas outras. Ainda que tais plataformas desempenhem um papel importante neste contexto – especialmente por realizar um mapeamento algorítmico das preferências de seus usuários –, elas não são as únicas responsáveis por este multifacetado fenômeno (RIEDER, 2020). Todas as ações que realizamos e plataformas que acessamos nos mais diversos aparelhos digitais contribuem para a geração desses dados. Exemplos seriam os relatórios de mobilidade que a Google (2022) tem oferecido desde o início da pandemia de COVID-19. A empresa utiliza os sistemas de GPS nativamente instalados na plataforma Android para mapear a mobilidade urbana, oferecendo relatórios anônimos para a comunidade. Estes podem ser cruzados com dados públicos sobre a infecção da doença – como os oferecidos pelo Governo do Estado de São Paulo (2022) – criando interpretações sobre o avanço da doença. O mapeamento de nosso comportamento nos níveis pessoal, profissional e de consumo em motores de busca também desempenha um importante papel (ZUBOFF, 2019). Segundo Zuboff (2019), tais sítios criam perfis de seus usuários, construindo um mercado que sugestiona a compra de produtos em listas de oferecimento cada vez mais personalizadas. Os dados coletados são resumidos apenas àquilo que digitamos, mas também ao que falamos, uma vez que dados de voz podem ser coletados por diversos aparelhos, como seria o caso das TVs inteligentes (LEE, 2015).

Mais recentemente, o oferecimento de serviços públicos (FILGUEIRAS; FLÁVIO; PALOTTI, 2019) e privados passaram por um processo de plataformação.<sup>2</sup> Além da coleta de dados de navegação e ações de usuários em seus sistemas, ações comuns às plataformas – com relevantes implicações no direito ao sigilo destas informações –, há de se pensar que elas excluem tanto indivíduos que não possuem o letramento tecnológico e de dados necessários para realização das tarefas, como aqueles que não dispõem de condições materiais para aquisição e manutenção de hardware compatível com estas plataformas, incluindo as conexões necessárias. Exemplos deste processo seriam o acesso ao auxílio Brasil (BRASIL, 2020), um sistema de transferência de renda, e o agendamento de serviços do Instituto Nacional do Seguro Social (INSS) que, desde 2016, vem transferindo seus agendamentos e realização de serviços de triagem para aplicativos (BRASIL, 2016b).

Há sistemas como o GOVBR,<sup>3</sup> que congrega serviços públicos em um único aplicativo. O GOVBR exige informações como número de documentos pessoais e dados bancários e obriga seus usuários utilizar sistemas de reconhecimento facial. Em

<sup>2</sup> A plataformação poderia ser definida como a migração de importantes aspectos de nossa vida cotidiana para as plataformas digitais (POELL; NIEBORG; VAN DIJCK, 2019). Tal processo tem efeitos sociais, legais, políticos e econômicos, não podendo ser observado por uma única perspectiva.

<sup>3</sup> Disponível em <https://www.gov.br/governodigital/pt-br/conta-gov-br>. Acesso em 30/12/2022.



outros termos, além de centralizar quase todos os serviços federais em um único banco de dados – e por conseguinte perfil de usuário –, o aplicativo/sistema colhe informações pessoais e associadas à fisionomia. Tais informações passam a estar disponíveis dentro da esfera governamental, permitindo monitoramentos e cruzamentos diversos. Um exemplo de como o Estado pode utilizar tais informações sobre seus cidadãos ocorreu durante as eleições de 2022. O Tribunal Regional Eleitoral de São Paulo (TRE-SP, 2022) a partir de uma resolução do Tribunal Superior Eleitoral (BRASIL, 2021) – que se apoia na Lei Geral de Proteção de Dados (BRASIL, 2019) – utilizou os dados biométricos do Departamento Nacional de Trânsito para validação de identidade dos eleitores nas seções de votação do estado. Algo que aparentemente pode ser pensado como uma facilitação para o eleitor, que não precisa se cadastrar junto à Justiça Eleitoral, na verdade, abre caminho para o cruzamento de dados entre diferentes autarquias públicas, algo que poderia ter consequências políticas importantes.

Por fim, a pandemia de COVID-19 acelerou processos de virtualização escolar (LIMA-LOPES; CÂMARA; OLIVEIRA, 2021). Tal migração ocorreu, muitas vezes, em plataformas proprietárias<sup>4</sup> que tiveram acesso a toda sorte de dados. A riqueza destes dados inclui programas e cronogramas de curso, as mensagens – aluno/professor, professor/aluno; professores/professores –, documentos administrativos (por vezes sigilosos), gravações de aulas e reuniões, toda agenda institucional, dados de navegação – onde se clica e por quanto tempo lá se fica –, entre diversos outros. O que cada uma dessas plataformas pode fazer com estes dados, depende dos acordos de utilização fechados em cada contratação.<sup>5</sup>

No âmbito privado, vale destacar a crescente migração de serviços bancários (YEE-LOONG CHONG et al., 2010) – que já possuem uma longa história de monitoramento de clientes e suas operações, financeiras – e das vendas de varejo – implicando o mapeamento do comportamento de compra no varejo, especialmente em lojas virtuais – como alguns exemplos importantes. Os resultados desta virtualização passariam pela criação de modelos que tornam explícitos os padrões de navegação (HUYNH et al., 2020), a modelação de tendências de compras em comunidades virtuais (OLBRICH; HOLSING, 2011), levantamento de intenções de compra para o desenvolvimento de estratégias centradas nos hábitos de navegação e padrão de aquisição (CHEN; SAIN; GUO, 2012), além do vazamento de dados (CHAKRABORTY et al., 2016).

Algumas plataformas privadas, como mídias sociais, e públicas, além de universidades que trabalham com modelos de ciência aberta (FECHER; FRIESIKE, 2014), viabilizam o acesso a alguns destes dados. O Twitter (2023), por exemplo, oferece a possibilidade de raspagem de postagens, oferecendo ao usuário comum uma

---

<sup>4</sup> Software proprietário pode ser definido como aquele que utiliza formatos e protocolos próprios. Tais protocolos não são abertos à comunidade (o que inclui o seu código) e seu uso se dá por meio de pagamento de licenças em diversos formatos e valores, sendo que sua manutenção depende da empresa que o publica. Ele está em oposição ao que se chama *Free/Libre and Open Source Software (FLOSS)*, no qual o código é aberto e disponibilizado a comunidade. Seria leviano dizer que os FLOSS não visam lucratividade em sua publicação, mas o formato aberto permite aos diversos utilizadores personalizarem ou fazer escolhas sobre a natureza daquilo que utilizam (WHEELER, 2006).

<sup>5</sup> Diferentes aplicativos proprietários, especialmente mídias sociais e plataformas diversas, possuem dados de usuários que podem ser utilizados de acordo com os termos aceitos em ato de contratação. Esses dados passam a ser propriedade da plataforma. Um exemplo seria a mobilidade de usuários de telefones inteligentes (Android e iOS), a discussão sobre os dados de mobilidade durante a COVID-19, logo acima, exemplifica a granularidade destes dados. Como experiência pessoal, o autor deste artigo solicita ao Google (desenvolvedora do sistema Android) dados sobre seu deslocamento mensal, de forma observar como se dá o monitoramento de seu deslocamento.



cota de dados – limitados em quantidade e temporalidade – e ao usuário acadêmico (professores e pesquisadores) uma quota de 10 milhões de tuítes por mês. O Google (2023) oferece acesso a alguns dados por meio de cadastros avaliados de acordo com perfil do requisitante.

No contexto da administração pública brasileira, o acesso aos dados governamentais deriva do processo de abertura que teve início durante a última década, resultando de uma demanda popular pela transparência que se dá a partir da redemocratização em 1985. Criaram-se, assim, mecanismos para obtenção de dados governamentais garantidos pela *Lei de Acesso à Informação* (BRASIL, 2011), regulamentada por dois decretos (BRASIL, 2012, 2016a). À vista disso, quase todas as autarquias governamentais possuem sítios específicos para divulgação de seus dados. Os modelos de ciência aberta (FECHER; FRIESIKE, 2014) tornam públicos dados de pesquisa, de forma que elas possam ser replicadas. Apesar de haver algumas restrições sobre tais dados – especialmente em pesquisas que lidam com seres humanos ou dados sigilosos –, a possibilidade de abertura ajuda a transformar alguns paradigmas científicos. Entre eles está o seu uso para replicabilidade de pesquisa (MEIE, 1995), – algo deveras incomum nos estudos da linguagem – ou para o ensino, contexto no qual a replicabilidade seria o caminho para compreensão e interpretação linguística. Esta última, seria responsável pela criação de materiais didáticos abertos que também romperiam perspectivas culturais elitistas e centralizadoras (LIMA-LOPES, 2022).

Contudo, navegar por uma sociedade dadificada está longe de ser trivial, uma vez que os dados oferecidos pelas mídias sociais são apenas uma ponta do iceberg informacional disponível para estas corporações (ZUBOFF, 2019). Os dados massivos são por elas utilizados livremente na construção de um capitalismo de vigilância (SEGURA; WAISBORD, 2019),<sup>6</sup> perpetuando estruturas sociais racistas (NOBLE, 2018) que agora se baseiam no comportamento algoritimizado de seus usuários. Por outro lado, as regulamentações sobre os dados públicos no Brasil parecem incentivar uma transparência passiva (JARDIM, 2012), uma vez que tanto a disponibilização quanto o acesso a tais dados não são facilitadas ao cidadão. Tal passividade exclui as possibilidades de transformação que uma efetiva exposição aos dados traria (GUTIÉRREZ, 2019). Já os dados disponibilizados em contexto de ciência aberta podem ser organizados e apresentados em plataformas que seguem um modelo acadêmico, distanciando-os da sociedade (LIMA-LOPES, 2022).

De fato, tanto na esfera privada como na pública, o acesso aos dados – que são diversos em sua forma e em sua apresentação – depende de ações de letramento relacionadas à aquisição, geração, processamento e visualização. Os processos de geração e coleta em larga escala têm causado um grande impacto nas ciências exatas e da terra, apesar de ter seu futuro ainda incerto na área de ciências humanas (KITCHIN, 2014, 2016). Isso porque, apesar de áreas como Ciências Sociais e Ciências Políticas já estarem incluindo tais dados em suas análises, algumas áreas como a Linguística e a Linguística Aplicada (doravante LA), por exemplo, ainda são resistentes a sua validação em pesquisa. Precisamos estabelecer uma visão crítica dos processos tecnológicos

<sup>6</sup> Partindo de autores como Segura e Waisbord (2019), Zuboff (2019) e José Van Dijck (DIGILABOUR, 2019), o Capitalismo de Vigilância é uma forma de capitalismo que tem como fonte de lucro os dados que os usuários fornecem gratuitamente e de livre e espontânea vontade às empresas de tecnologia, incluindo todos os serviços da então chamada web 2.0, motores de busca, navegação, mídias sociais, mensageiros instantâneos (como Telegram e WhatsApp) entre diversos outros. Os dados são matéria-prima de mapeamento pessoal, que, quando refinados, tentam prever nosso comportamento de consumo. A ideia é que estes dados possam ser utilizados na sugestão de novos produtos (consumo e ideológicos) aumentando a chance de sucesso e engajamento do que se vende.



dentro do contexto da LA, tomando consciência de que o conhecimento de áreas, muitas vezes renegadas em nosso campo de atuação, pode trazer contribuições importantes para a compreensão da linguagem e seu papel na contemporaneidade. A meu ver, tal fenômeno é resultado de uma tradição qualitativista que rejeita ações de pesquisa que não sigam um modelo interpretativista restrito em termos metodológicos.

Como já discutido em Lima-Lopes (2022), uma reflexão sobre os dados e seu papel abre importantes portas para o contexto da LA. Naquela ocasião, o foco principal foi estabelecer uma discussão epistemológica associando modelos de Ciência Aberta aos potenciais de democratização dos Recursos Educacionais Abertos (REA). Aqui se busca discutir como o acesso aos dados e a introdução das práticas de letramento necessárias para sua análise e processamento podem auxiliar na construção de um modelo de letramento científico no contexto das ciências da linguagem. Busca-se, ainda que de forma inicial, refletir como tal letramento poderia ajudar na conscientização sobre o uso de nossos dados pelas diversas plataformas e sobre necessidade construir novas narrativas baseadas em visões não hegemônicas. No caso da primeira, caminhar-se-ia pela reflexão e autocontrole daquilo que decidimos conscientemente ceder a estas plataformas. Na segunda, utilizar-se-iam os dados cedidos por algumas destas plataformas e pelo serviço público em comparação a dados gerados localmente, criando formas de representação da realidade que não refletem, apenas, visões hegemonicamente dominantes.

## 2. Dados como extrativismo

A compreensão crítica de nossa relação com a tecnologia é essencial. O processamento computacional é parte da realidade social contemporânea e responsável por estabelecer e manter nossas relações, influenciando nossa percepção da realidade nos níveis político, profissional e pessoal (BRIDLE, 2019). Sua ubiquidade leva a uma percepção dos processos tecnológicos como mágicos, opacos para a maioria dos indivíduos e relativamente claros para aqueles que deles participam (BRIDLE, 2019). Razão esta que torna urgente o desmonte de mitos criados durante as primeiras décadas de presença das tecnologias digitais; talvez o mais importante seja o da neutralidade e universalidade epistemológicas dos processos tecnológicos.

Morozov (2018) e Hui (2020) observam que a realidade tecnológica vivida no contexto pós segunda guerra, está profundamente ligada às empresas de tecnologia localizadas no *Vale do Silício*. Hui (2020) reflete sobre os processos de colonização tecnológica e, por conseguinte, da constituição social de uma semiose da sincronização ideológica a partir de uma perspectiva oriental, ao passo que Morozov (2018), por seu turno, apoia-se também em uma percepção particular, a russa.<sup>7</sup> Assim, para Hui (2020), o que chamamos tecnologia não é algo universal: diferentes sociedades e seus diversos contextos políticos poderiam desenvolver seus próprios processos. Assim o teria sido antes da hegemonia europeia sobre Ásia, continentes com motivações tecno-sociais bastante diversas. A pretensa universalização serviria, logo, como uma ferramenta política de dominação, colonização e atomização. Ela seria responsável por desagregar os grupos em individualidades cada vez mais capilares, que na particularidade de sua expressão facilitariam a supremacia de um pensamento tecnológico de base europeia e estadunidense.

---

<sup>7</sup> É claro que seria aceitável pensar que tanto a China como a Rússia, seriam potências hemogênicas concorrentes ao estabelecido no contexto europeu ou estadunidense. Não obstante, vale pensar que os contextos de produção e resistência tecnológica defendida por Hui (2020) e Morozov (2011, 2018) pode ser um caminho crítico reflexivo para o estabelecimento de um pensamento local.



Além de impor conhecimentos e práticas diversas (HUI, 2020), as tecnologias representariam uma expressão da masculinidade (WAJCMAN, 2000), raça (NOBLE, 2018) e identidade corporal (BANKS, 2018) algo que excluiria mulheres, membros da comunidade transexual e atores não-brancos de seu protagonismo. Consequentemente, criar-se-ia uma sincronia tecnológica resultando na convergência em direção a um único tempo generalizante, uma simples operação de vetores, na qual o conhecimento tecnológico teria apenas uma via:

(...) [a] universalização de epistemologias particulares e, através de meios tecnoeconômicos, a elevação de uma visão de mundo regional ao status de metafísica supostamente global (HUI, 2020, loc. 174).

Então, se por um lado, a tese da compreensão tecnológica como um universo antropológico definido pela simples extensão da memória e superação de nossas características gregárias e fisiológicas visa à constituição de uma cosmopolítica tecnológica colonizante, por outro, se pensarmos que a tecnologia pode ser definida como antropoliticamente não-universal, poderíamos constituir uma multiplicidade de histórias e cosmotécnicas, todas se apoiando processos semióticos e sociais diversos. O *Vale do Silício* estabelece uma retórica baseada na ideia de revolução e transformação que tem seu *locus* nos espaços tecnológicos de interação. Nesse sentido, é válido encarar uma visão crítica desses sistemas teria importância para o desenvolvimento de processos comunicacionais, infraestruturas e aprendizado que sirvam aos interesses locais de produção.

Ao tratarmos a tecnologia contemporânea simplesmente como digital e não como uma instância política, estamos favorecendo a alguns setores aos quais se aferem protagonismo no gerenciamento e consolidação de dados. A dadificação da sociedade — definida como o fenômeno de quantificação da experiência humana por meio de seus hábitos digitais, muitas vezes, com fins econômicos (MEJIAS; COULDREY, 2019) —, é uma forma de monetização do comportamento humano em favor de uma política social baseada na adesão opcional (até certo ponto, claro) às plataformas que significam uma entrega voluntária das liberdades a favor de um monitoramento (MOROZOV, 2018). Tal monitoramento torna o comportamento humano em um produto monetizável (ZUBOFF, 2019), no qual nosso comportamento social, outrora parte de nossa vida cotidiana, passam a possuir valor com base em seu potencial de mercantilização.

No enquadramento sugerido por Morozov (2011), as diferentes plataformas tecno-sociais são responsáveis por um cerceamento prévio de nossas relações discursivas e ativistas em contextos nos quais o processo de transformação fosse considerado financeiramente relevante. Em outras palavras, o autor afirma que tais plataformas proporcionam um pacote completo de ação social: ao mesmo tempo que oferecem possibilidades de manifestação de uma política e cultura hegemônicas, ele torna disponível a nós os meios de rebeldia contra esta mesma cultura. Ao trazermos o pensamento de Morozov (2011) para a realidade da LA, podemos observar que as metodologias propiciadas pelas Humanidades Digitais poderão auxiliar não apenas no levantamento, como também em sua análise crítica, especialmente quanto ao ativismo de dados.

Para Singh (2019), a chamada Web 2.0 não cumpre os ideais de universalização e democracia previstos por teóricos como Jenkins (2006), para quem a criatividade incentivada pelo discurso dessas plataformas resultaria em uma cultura mais democrática, maior conscientização e participação política. Tal falha, como coloca Singh (2019), tem em sua gênese a construção de espaços disponibilizados de forma falsamente gratuita, uma vez que sua utilização se dá pela entrega de dados pessoais.



Tal fato teria, em consonância com Morozov (2011), consequências importantes, uma vez que tais plataformas passam a ser simultaneamente o lugar de agressão e reação, uma pseudo-liberdade incentivada pelas possibilidades de anonimato. Resultar-se-ia, assim, não apenas na aceleração, amplificação e intensificação, mas também no estabelecimento de monopólios informacionais e comunicacionais nos quais o discurso do acesso e da usabilidade ganha cores de cultura participatória. Tal discurso caracteriza a criatividade como intrinsecamente conectada a uma ideia neoliberal, na qual busca-se produzir condições para monetização e ações empreendedoras (SINGH, 2019). À vista disso, nossa ágora deixa de ser a praça pública e passa a ser composta por espaços que, apesar de abertos ao acesso, são privados, aferem lucro e possuem expressão política.

Barrett e Kreiss (2019) analisam as transformações pelas quais a plataforma Facebook passa em sua relação com a política internacional, relacionando tais transformações às pressões econômicas provenientes de investimentos de publicidade. Os autores criam o termo *platform transience* (*transciência* de plataforma) para descrever como as plataformas modificam de forma rápida, radical, contínua e adaptativa em favor da captura de investimentos relacionados a campanhas políticas e ideológicas. A *transciência* seria uma forma de manter a falta de transparência, evitando que tais empresas sejam encaradas e responsabilizadas por suas ações. Como coloca a pesquisadora holandesa José van Dijck em entrevista ao blog DigiLabour:

As plataformas não são constructos neutros ou livres de valor. Elas v[ê]m com normas e valores específicos inscritos em suas arquiteturas. (...) A dataficação e a mercantilização estão enraizadas nos princípios neoliberais do capitalismo dos Estados Unidos. Os dados são recursos transformados em mercadorias pelo processamento algorítmico. No entanto, o problema é que os processos sociais e econômicos estão ocultos nos algoritmos, nos modelos de negócios e nos fluxos de dados que não estão abertos ao controle democrático. Assim, a ideologia do neoliberalismo define a arquitetura da nossa sociedade conectiva (DIGILABOUR, 2019).

As cores de neutralidade e universalidade favorecem ações, que mais das vezes, poderiam ser classificadas como extrativistas. Tal extrativismo utilizaria nossos dados pessoais para construir uma infraestrutura e a ubiquidade tecnológica que transforma indivíduos em *commodities* dadificados, de forma a reproduzir a inequalidades a medida que novas formas de capitalismo digital se constituem (SEGURA; WAISBORD, 2019).

Se partirmos do argumento colocado por Segura e Waisbord (2019), seria correto observar que o poder econômico destas plataformas, localizadas no Vale do Silício, além de impor sua epistemologia tecnológica (HUI, 2020; MOROZOV, 2018), também devem realizar processos extrativistas muito similar a ações colonizatórias do século XIX. A diferença é que tal processo se dá pela obtenção de dados que, apesar de serem tratados discursivamente de maneira muito similar à extração de recursos naturais livremente disponíveis (COULDREY; MEJIAS, 2019), são pessoais em sua gênese. Dessarte,

[d]e forma semelhante ao colonialismo, o capitalismo de dados nega às populações a humanidade, transformando-as em recursos e sujeitos criadores de valor. No entanto, não emprega a força como o fez o colonialismo. O capitalismo de dados usa técnicas insidiosas e dúbias para extrair valor, mas suas táticas não são apoiadas por exércitos permanentes e escravidão (SEGURA; WAISBORD, 2019, p. 416, Tradução minha).<sup>8</sup>

---

<sup>8</sup> Do original: "Similar to colonialism, data capitalism denies populations their humanity by turning them resources and value-creating subjects. It does not, however, employ force the way colonialism did. Data



### 3. Letramento de Dados e suas possibilidades de resistência técnico-espistemológica

Mas seria possível resistir ou contrapor-se a este processo que autoras como Mann e Daly (2019) denominam de *colonialismo de dados*?

Segundo Gutiérrez (2019), a resistência à coleta massiva, indiscriminada tem se tornado uma importante ferramenta de oposição à manipulação massiva, oferecendo uma plataforma de resistência política capaz de gerar diferentes diagnósticos sociais e narrativas que podem atacar esta hegemonia. Infelizmente, a falta de programas que incentivem a utilização e observação crítica leva a maioria das pessoas a ter um papel de produtores inconscientes de dados, sendo vitimados por sistemas de vigilância estatal e privada. Para Segura e Waisbord (2019), o contexto dadificado em que vivemos também permitiria um conjunto de ações coletivas que objetivam desafiar a lógica tócnico-corporativa representada pelo capitalismo de vigilância. Tais ações caminham de forma a oferecer alternativas ao engajamento tradicionalmente presente em nosso contexto online. Para os autores, haveria quatro principais áreas nas quais essa resistência poderia existir.

A primeira poderia ser definida como a intervenção de maneira a possibilitar o acesso da população à conexão, contribuindo para a potencial inclusão de grupos menos favorecidos no debate tecnológico. Embora tal acesso não garanta o uso de qualidade da informação, no contexto brasileiro, os centros comunitários (STRAUBHAAR; SPENCE, 2007) e os telefones celulares, ou dispositivos locativos, (NEMER, 2018) têm tido um importante papel. A segunda enfatizaria a possibilidade de acesso à bens culturais pelas redes, destacando-se aqui ONGs, artistas e coletivos que disponibilizam seu acervo de forma aberta.

A terceira se refere ao acesso à informação pela disponibilização de dados públicos compilados. Um exemplo deste tipo de ativismo é a geração de dados realizada pelo Grupo Gay da Bahia,<sup>9</sup> que há anos compila por conta própria dados de violência contra a população LGBTQ+.<sup>10</sup> Ainda merecem destaque iniciativas que não apenas disponibilizam seus dados, mas capacitam pessoas a navegar por eles de forma a possibilitar transformações sociais. Entre tais iniciativas, destacam-se a PretaLab<sup>11</sup> – uma ONG que oferece formação gratuita em análise de dados para mulheres negras –, RLadies<sup>12</sup> – uma iniciativa que prevê a instrumentalização de mulheres na linguagem R –, a PyLadies<sup>13</sup> – um programa de mentoria que incentiva e prepara mulheres para a utilização de linguagem Python – e a Escola de Dados<sup>14</sup> – uma ONG que capacita pessoas e instituições para o trabalho com dados visando a mudança social. A última se materializaria pelo direito de privacidade na rede; movimento que, em nosso país, levou à criação da lei geral de proteção de dados (BRASIL, 2019).

Se pensarmos que os dados podem gerar alguma forma de existência, há alguns importantes exemplos. Gutiérrez (2020) reflete como a utilização de sistema de geolocalização pode contribuir para a construção de mapas coletivos. Tais mapas podem ser responsáveis por criar representações sobre localidades que diferem das formas

---

capitalism uses insidious and duplicitous techniques to extract value, but its tactics are not backed up by standing armies and slavery”(SEGURA; WAISBORD, 2019, p. 416).

<sup>9</sup> Disponível em <https://grupogaydabahia.com.br/>, acesso em 21/07/2020.

<sup>10</sup> O Movimento Gay da Bahia faz um levantamento inicial na imprensa sobre os crimes de ódio contra a população LGBTQ+ e, posteriormente, visita delegacias para obter os Boletins de Ocorrência e validar as informações.

<sup>11</sup> Disponível em <https://www.pretalab.com/>, acesso em 21/07/2020.

<sup>12</sup> Disponível em <https://rladies.org/>, acesso em 21/07/2020.

<sup>13</sup> Disponível em <https://pyladies.com/>, acesso em 21/07/2020.

<sup>14</sup> Disponível em <https://escoladedados.org/>, acesso em 21/07/2020.



hegemônicas, à medida que permite novos modelos informacionais e afetivos do espaço social.<sup>15</sup> Na perspectiva teórica, Gutiérrez (2021) e D'Ignazi e Bhargava (2020) refletem sobre os dados a partir de uma perspectiva feminista. A primeira discute como os algoritmos aplicados às produções audiovisuais possuem um viés preconceituoso em relação às mulheres. Seu estudo também reflete na falta de clareza de como tais algoritmos tomam decisões baseadas em gênero, oferecendo uma agenda de pesquisa para trabalhos futuros. D'Ignazi e Bhargava (2020) discutem como são construídas narrativas de dados utilizando tecnologias digitais e não digitais. Seu foco principal na pluralidade de representações e interpretações possíveis a partir de uma abordagem feminista à representação visual de dados em contextos educacionais.

Como já colocado anteriormente, navegar pelo universo dadificado está longe de ser trivial, especialmente porque a utilização de dados requer não apenas um, mas uma série de letramentos e habilidades que se entrecruzam. Tais habilidades são essenciais para que possamos criar contextos que favoreçam ações de resistência como as que discutimos. Apesar do termo *Letramento de Dados* ser amplamente utilizado para definir o resultado de tal convergência, acredito haver a necessidade de definição mais clara deste conceito.

Calzada Prado e Marzal (2013) partem da necessidade de sistematizar o termo *Letramento de Dados* motivados definição pouco precisa e pela importância que o termo ganha dentro do movimento de ciência aberta. A falta de precisão, para estes, nasce da necessidade de adaptar este termo aos diversos contextos de pesquisa. Calzada Prado e Marzal (2013), então, definem o conceito de *Letramento de Dados* a partir de cinco habilidades. Entre elas estaria a habilidade de obter dados dentro dos diversos tipos possíveis e fontes possíveis, julgando-as de acordo com sua aplicabilidade para o questionamento empreendido. A compreensão do que são dados é outro fator determinante. Tal conhecimento passaria não apenas pelos tipos e formatos, mas também pelo seu papel determinante como entidade transformadora da sociedade.

A terceira habilidade estaria relacionada com a capacidade de ler, avaliar e interpretar dados. Inicialmente, a leitura estaria ligada ao conhecimento sobre as convenções de apresentação, determinantes para a apreensão de sua estrutura e conteúdo. A avaliação passa pela análise crítica dos dados, observando se estes respondem aos questionamentos de pesquisa (CALZADA PRADO; MARZAL, 2013). Embora não mencionado pelos autores, aqui também caberia a avaliação ética de suas fontes e conteúdo. O gerenciamento exige a habilidade de salvar e conservar os dados em formatos legíveis pelo maior número de programas e linguagens, incluindo metadados consistentes. Ao meu ver, tal concepção passa pela desmitificação de diversos formatos fechados de processadores de texto e planilhas de dados. Por fim, o uso levaria a uma reflexão sobre a produção de representações que tragam uma síntese daquilo que se questiona, algo que inclui uma ponderação ética.

Wolff e colaboradores (2016) também refletem sobre esta questão. Para os autores, a sociedade em que vivemos está cada vez mais dependente de processos dadificados que interferem diretamente em nosso cotidiano, tornando necessário o aprendizado de componentes relacionados ao tratamento e interpretação de dados na escola. Uma vez que as pesquisadoras também possuem dificuldades em definir o termo, eles partem de do conceito de letramento estatístico como forma de examinar criticamente as necessidades básicas do navegar pelos dados. Eles analisam, então, como o *Letramento de Dados* é definido em diferentes abordagens, partindo sempre da

<sup>15</sup> No trabalho de Gutiérrez (2020) lugares de interesse a um grupo e que não estão mapeados por empresas de navegação são inseridos em mapas produzidos localmente.



prática analítica como norte. Em sua conclusão, apontam para a necessidade da compreensão dos dados como instrumento de entendimento de nosso mundo por meio de uma metodologia baseada no questionamento.

Caminhando por esta seara com apoio do trabalho de Bhargava (2019), é possível identificar cinco tendências gerais que parecem se completar. A primeira estaria relacionada à capacidade de se *adquirir* os dados, ou seja, o processo de coletar tais dados a partir de fontes já existentes ou gerá-los a partir de observações do mundo real. Esse tipo de habilidade está intimamente relacionada à busca, compreensão e extração desses dados, seja de fontes online (BHARGAVA, 2019) ou offline.

A segunda incluiria processos de análise de dados, aplicação de operações algorítmicas de análise por meio de programas de computadores e linguagens específicos. Em muitos casos, apesar de a análise de dados massivos enfatizar uma abordagem qualitativa, tal análise ocorre por meio de lentes quantitativas, levando a necessidade de utilização de métodos mistos (JOHNSON; ONWUEGBUZIE; TURNER, 2007). Um terceiro componente teria seu foco na representação, ou seja, transformar um conjunto bruto de dados em algo estruturado e não numérico. A representação é um passo importante para a compreensão dos dados, especialmente, na observação de padrões e na construção de narrativas. Obviamente, a representação é um fator relevante para a criação de argumentos baseados em dados, de forma reforçar ou desmistificar perspectivas em favor de transformações. Por fim, dever-se-ia considerar a necessidade de compreender a forma como os dados são obtidos, construídos e disponibilizados (BHARGAVA, 2019). Talvez estas duas últimas características sejam o ponto central, uma vez que ele oferece a possibilidade de refletir criticamente sobre o recorte de realidade representado pelos dados, incluindo questões éticas sobre sua obtenção, anonimidade e tratamento.

Carmi e colaboradores (2020) refletem sobre o papel dos dados em processos de letramento, especialmente no âmbito do atual contexto de má-informação e desinformação. As autoras identificam três principais falhas nos programas de letramento tecnológicos tradicionais: 1) a necessidade de se incentivar uma perspectiva além do individual; 2) a falta de criticidade em relação ao ecossistema digital e 3) falta de ações que incentivem a proatividade em meios digitais. Apesar dos desafios que a inserção dos dados poderia trazer para o contexto de ensino, as autoras afirmam que este poderia ser um importante caminho para discussão sobre as distorções informacionais. Carlson e colaboradores (2011) também trabalham com a questão de letramento. Seu foco está em compreender as necessidades de manipulação e análise de dados por grupos acadêmicos. Por esta seara também caminham Estevão e Strauhs (2020) que discutem a questão do *Letramento de Dados* a partir de experiência de ciência aberta na área de ciências sociais. Seus resultados apontam para a necessidade de enfatizar diferentes aspectos das pesquisas com dados, inserindo práticas pouco comuns às ciências humanas (SILVA, 2021).

Wolff, Montaner e Kortuem (2016) refletem sobre a inserção dos dados no Currículo Escolar do Reino Unido. Para os autores, ações práticas sobre o letramento de dados poderia partir de algumas atividades práticas. Deve-se incentivar a construção de uma narrativa, ou novas narrativas, a partir dos dados. Os autores incentivam em especial a produção, utilizando dados que os alunos podem coletar em seu contexto, engajando o aluno que passaria a desenvolver seu olhar crítico sobre o seu contexto. É nesse momento que os autores também sugerem que sejam formuladas perguntas, que servirão como base para a pesquisa.



**Figura 1:** Ciclo do letramento de dados (Adaptado de WOLFF; MONTANER; KORTUEM, 2016, p. 63)  
**Fonte:** Autor

A figura 1, acima, traz este ciclo traduzido em imagem. O primeiro passo seria estabelecer um problema a ser pesquisado, ou uma visão a ser contestada, algo importante para que os objetivos reais de uma análise baseada em dados possam ser estabelecidos. Segue-se, então, o estabelecimento de um plano de pesquisa. Deve-se eleger quais são as variáveis a serem estudadas, de forma observar se elas contribuem efetivamente para a resposta estabelecida em minha questão. Próximo passo seria colher estes dados, organizá-los e limpá-los, este último procedimento aplicado quando trabalhamos com dados obtidos de fontes externas. A análise e visualização busca traduzir os dados já tabulados em formas visuais, as quais podem facilitar a compreensão dos padrões encontrados. Por fim, estes dados são interpretados, e deles um novo problema surge.

#### 4. Letramento de Dados e alguns possíveis impactos na Linguística Aplicada

Em síntese, o *Letramento de Dados* pressuporia o levantamento, tratamento e representação e interpretação dos dados, que pode ou não ocorrer por meio de ferramentas computacionais. A perspectiva crítica é um pressuposto neste caminho, uma vez que a pretensa neutralidade da tecnologia e dos dados é desvelada em favor da construção de conhecimentos transformativos e de resistência. Ao mesmo tempo, ao partirmos para uma perspectiva de letamentos que tenha a pesquisa como um princípio norteador, estamos incentivando a reflexão não estanque sobre a linguagem em seus diferentes contextos. Se por um lado, tal postura contradiz a tradicional ideia de que a língua é algo aprendível em totalidade, mas fruto de uma construção de saberes constantes, por outro, ela incentiva processos de letramento científico dentro do contexto do aprendizado de línguas. Ou seja, por um processo de análise e comparação sistemáticos, questões de linguagem poderiam ser aprendidas a partir de ações de pesquisa. Inicialmente, acredito que esta perspectiva traz em si elementos importantes sobre letramento científico e sua aplicabilidade para as ciências da linguagem, uma vez que ela poderia nos ajudar a construir experiências de aprendizagem baseadas na

construção de conhecimentos locais que, comparados em uma rede de dados abertos, poderiam contribuir para a formação ampla sobre questões de linguagem. Esta visão sobre como a pesquisa pode ser um importante aliado ao ensino de línguas é compartilhada por outros pesquisadores, especialmente Silva (2021). O autor desenvolve um importante trabalho de releitura do conceito de educação científica presente nas ciências exatas e da terra de forma a demonstrar como práticas de estudo e pesquisa aplicados podem contribuir para o letramento crítico, garantindo um espaço democrático e culturalmente relevante para o desenvolvimento de práticas de reflexão e aprendizado.

A meu ver, a Linguística Aplicada possui muito a contribuir para essa perspectiva, por ser uma ciência voltada para a reflexão sobre uma temática socialmente relevante, a linguagem efetivamente em uso. A versatilidade da LA está compreender uma perspectiva plural as diferentes formas de linguagem, sendo que a língua pode ser observada como um sistema de representações (HALLIDAY, 1978). Um sistema de signos motivados (KRESS, 1993), que tem em seu estudo uma prática transformadora (HALLIDAY, 1999).

De acordo com Aiello (2020) a principal contribuição de Sociossemiótica para a compreensão dos processos de análise e representação dos dados está em seu papel social. Como coloca a autora, tal abordagem surge no âmbito da LA como uma síntese da semiótica estrutural e dos trabalhos de Michael Halliday (1978). Nesta abordagem, os signos são caracterizados como desenvolvidos a partir da necessidade de performar ações específicas (KRESS, 1993; MARTIN, 2016). Da mesma forma que outras abordagens presentes na área, ela está preocupada com a arquitetura dos textos, especialmente com a relação intrínseca e inseparável entre as diversas formas convencionalizadas de significação (AIELLO, 2020). Concordo com Tygel e Kirsch (2016) que realizam uma importante pesquisa ao compararem como metodologias de *Letramento de Dados* podem contribuir para a formação do indivíduo crítico em nossa contemporaneidade. Os autores partem do trabalho de Paulo Freire (2011, 2014), observando como os processos de compreensão de nossa sociedade poderiam passar pela construção de nossa própria representação crítica e dadificada.

Pode-se, à vista disso, pensar os dados como um modo de linguagem que interage e se inter-relaciona com outros modos, como a visualização de dados e a linguagem escrita, por exemplo. Seria importante também partir de uma discussão que inclua práticas de leitura e análise ainda pouco incentivadas na LA, como o uso extensivo de computadores para funções que além da escrita tradicional e entretenimento. Ações estas poderiam resultar em uma melhor interação entre a prática de aprendizado da do discurso científico, algo que possibilitaria duas transformações essenciais. Na primeira, buscar-se-ia compreender de forma crítica os processos sociais que subjazem os processos de construção do conhecimento científico, algo que poderia resultar ou ser incentivado pelo contato direto com os dados, sua organização e origem. A segunda, seria levar a uma reflexão também crítica sobre nosso papel como produtores de dados, incluindo nossa relação com serviços públicos e privados.

Desse modo, precisamos compreender que os dados devem ser encarados como uma forma de linguagem e representação. O processo investigativo é central, uma vez que ele nos ajudará a trazer à tona as contradições típicas do processo de compreensão e produção de dados. Logo as representações podem existir a partir do tratamento e geração de dados que seguem parâmetros relevantes para os contextos nos quais os processos de letramento se dão, resistindo as já convencionalizadas e dominantes.

Dessarte, se pensarmos em um programa de *Letramento de dados* voltado para a Linguística Aplicada, necessitaremos ampliar nossas reflexões em alguns pontos que acredito ainda serem poucos discutidos:

- **Papel das tecnologias e das mídias**

É necessário revisitar o papel das tecnologias (saber técnico) e das mídias (expressão cultural da tecnologia) (MANOVICH, 2005) em nosso fazer pedagógico. Inicialmente, compreendendo que elas refletem parâmetros ideológicos, considerando que o simples acesso não gera uma revolução crítica e participativa. Isto posto, precisamos olhar com mais atenção aos processos de informação e sua propagação na rede.

- **Dados e metadados**

O trabalho com dados é um pressuposto nesta nova perspectiva crítica. Eles podem ser tanto dados já existentes em outros contextos de aprendizagem e de pesquisa – falo especificamente de bases de dados disponíveis por modelos de ciência aberta ou por dados abertos –, como também aqueles produzidos localmente. Além disso, o conceito de metadados (ou dados sobre os dados) ainda é pouco discutido em ciências da linguagem. Os metadados (GREENBERG, 2009) têm um papel importante em pesquisas que partem dos *Big Data*, visto que eles podem trazer relevantes informações sobre o contexto de produção. Cabe também destacar que a utilização de dados gerados localmente podem não apenas trazer importantes contribuições para a consciência sobre diferenças e semelhanças entre outras variantes – especialmente a norma culta –, mas também criam a possibilidade de registros históricos ou geracionais, além das claras oportunidades de construção de identidade.

- **Programação e novas linguagens**

O trabalho com dados parte da utilização de novas linguagens na escola. Tais linguagens podem ser tanto sistemas de programação e código aberto, como Python (PYTHON, 2023) e R (R CORE TEAM, 2021), ou mesmo com planilhas de cálculo, como é o caso de ferramentas abertas como o LibreOffice, ou mesmo programas proprietários que estejam disponíveis. Entre os exemplos de trabalhos nesta área, pode-se citar o trabalho de Oliveira (2019), que desenvolve fórmulas para a utilização de planilhas de dados no ensino baseado em corpora.

- **Recursos Educacionais Abertos (REA)**

Como já colocado em Lima-Lopes (2022), o trabalho com os REA aplicados à pesquisa científica como método de ensino de línguas traz importantes contribuições. A mais importante talvez seja a possibilidade de discutir questões de linguagem e identidade por meio de reflexões que partem de realidades diferentes daquelas presentes nos manuais fechados. Isso certamente não é apenas uma reação aos conteúdos homogenizantes, como também uma forma de resistência contra um modelo capitalista que vê professores como aplicadores de manuais pré-formatados e alunos como consumidores de versões simplificadas deste mesmo manual.

- **Pesquisa e sua importância**

Pensar na pesquisa como elemento norteador dos diferentes caminhos de ensino e aprendizagem também são mister. Isto implicaria agregar dados em diferentes formatos em nossa prática cotidiana. Naturalmente que tais dados devem valorizar a produção de conhecimentos locais a partir da comparação com outros resultados e outras bases de dados disponíveis.

Um impacto importante seria a necessidade da presença de componentes de pesquisa no cotidiano de sala de aula. Claro que tal perspectiva também recai sobre o professor, que passa a ter necessidade de transitar por metodologias de pesquisa. A inserção da pesquisa científica como base da formação em linguagem levaria a natural comparação entre os conteúdos dos manuais didáticos e os resultados obtidos em nossa sala de aula.

## 5. Considerações Finais

Este ensaio teve por objetivo refletir sobre as possíveis contribuições que uma reflexão crítica sobre tecnologia e Letramento de Dados poderiam trazer para o Letramento Científico. A discussão se iniciou pelo papel da tecnologia na sociedade, passando então a discutir o extrativismo de dados e as possibilidades de resistência e possíveis transformações ao se agregar tal perspectiva à LA. Ao abordarmos o ensino de língua por meio do trabalho com dados, estamos mostrando que as escolhas e estruturas usadas pelos falantes são motivadas pelos seus diferentes contextos de situação (HALLIDAY, 1978), construindo ferramentas de compreensão do papel das diversas linguagens no processo comunicacional.

Como coloca Albagli (2015), nossa sociedade contemporânea passa por um processo de captura e privatização de nossos comportamentos pessoais e coletivos. Ao meu ver, ao trazermos uma perspectiva baseada no Letramento Científico e no Letramento de Dados, podemos possibilitar o surgimento de uma cultura de resistência, que passa imprescindivelmente por uma ocupação do espaço que a utilização de dados pode gerar. Outrossim, o trabalho com dados pode levar a uma reflexão crítica sobre as armadilhas abertas no espaço digital e dadificado ao qual estamos expostos, especialmente se tomarmos como base os eventos políticos que, em muitos casos, têm sua origem na falta de transparência das diversas plataformas digitais ou mesmo seu uso em favor de um processo de desinformação e má informação de seus usuários. Acredito que de forma a pensar sobre possíveis aplicações desta reflexão devemos trazer à baila algumas perspectivas ao contexto da LA, para que, oxalá, reflita-se sobre a replicabilidade de pesquisa. A ideia de sermos capazes de repetir análise e experimentos de linguagem é importante para esta perspectiva. Parte significativa do material científico que produzimos nas ciências da linguagem não possui esta característica, algo relevante para o ensino baseado em letramento científico e de dados.

Uma reavaliação sobre o papel do computador, nossa definição de tecnologia e dos métodos nos contextos de pesquisa e da escola merecem destaque. Parte significativa de nossa relação com a tecnologia em LA reflete aquilo que Bridle (2019) classifica como um fascínio quase religioso. Melhor dizendo, procura-se compreender pouco, ou quase nada, dos processos ideológicos, discursivos e sociais que estão por detrás das diferentes plataformas tecnológicas, guardando um deslumbramento sobre suas possibilidades técnicas, muitas vezes pouco realizáveis. Este fascínio seria o princípio básico para os processos de colonização epistemológica colocada por Hui



(2020) e Morozov (2011, 2018) e extrativismo de dados, como destacam Segura e Waisbord (2019) e José Van Dijk (DIGILABOUR, 2019).

Minha contribuição efetiva talvez seja estabelecer um cenário no qual o conhecimento é construído coletivamente por meio da análise, geração e manipulação de dados sobre a linguagem. Algo que tem impacto transformativo na forma como interagimos com nossa sociedade contemporânea.

### Agradecimentos

O autor agradece ao CNPq (311099/2021-1) pelo financiamento desta pesquisa e ao Prof. Dr. Wagner Rodrigues Silva (UFT) pela leitura de versões anteriores deste trabalho.

## 1. Referências

- ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Org.). . **Ciência aberta, questões abertas**. Brasília: IBICT/UNIRIO, 2015. p. 9–26.
- BANKS, William P. Beyond Modality. In: ALEXANDER, Jonathan; RHODES, Jacqueline (Org.). . **The Routledge Handbook of Digital Writing and Rhetoric**. 1. ed. [S.l.]: Routledge, 2018. p. 341–351. Disponível em: <<https://www.taylorfrancis.com/books/9781315518480/chapters/10.4324/9781315518497-32>>. Acesso em: 24 dez. 2019.
- BARRETT, Bridget; KREISS, Daniel. Platform Transience: Changes in Facebook's Policies, Procedures, and Affordances in Global Electoral Politics. **Internet Policy Review**, v. 8, n. 4, 31 dez. 2019. Disponível em: <<https://policyreview.info/node/1446>>. Acesso em: 16 nov. 2022.
- BHARGAVA, Rahul. Data Literacy. In: HOBBS, Renee; MIHAILIDIS, Paul (Org.). . **The International Encyclopedia of Media Literacy**. [S.l.]: Wiley, 2019. . . Acesso em: 6 maio 2020.
- BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. **Information Communication and Society**, v. 15, n. 5, p. 662–679, 2012.
- BRASIL. **App Auxílio Brasil**. Disponível em: <<https://www.gov.br/cidadania/pt-br/auxilio-brasil/app-auxilio-brasil>>. Acesso em: 6 jan. 2023.
- \_\_\_\_\_. **Decreto N° 7724 - Regulamenta a Lei no 12.527**. . Brasília-DF: Presidência da República, Secretaria-Geral, Subchefia para Assuntos Jurídicos, 16 maio 2012. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2012/decreto/d7724.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/decreto/d7724.htm)>. Acesso em: 16 nov. 2022.
- \_\_\_\_\_. **Decreto N° 8777 - Institui a Política de Dados Abertos do Poder Executivo Federal**. . Brasília-DF: Presidência da República, Secretaria-Geral, Subchefia para Assuntos Jurídicos, 5 nov. 2016a. Disponível em: <[http://www.planalto.gov.br/CCIVIL\\_03/\\_Ato2015-2018/2016/Decreto/D8777.htm](http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8777.htm)>. Acesso em: 16 nov. 2022.
- \_\_\_\_\_. **Lei Geral de Proteção de Dados Pessoais (LGPD) nº 13.853**,. . Brasília-DF: Presidência da República Secretaria-Geral Subchefia para Assuntos Jurídicos, 2019. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/lei/l13709.htm)>. Acesso em: 23 jan. 2023.
- \_\_\_\_\_. **Lei N° 12.527: Lei de Acesso à Informação, LAI**. . Brasília-DF: Câmara dos Deputados, 18 nov. 2011. Disponível em:

<<https://www2.camara.leg.br/legin/fed/lei/2011/lei-12527-18-novembro-2011-611802-norma-pl.html>>. Acesso em: 16 nov. 2022.

\_\_\_\_\_. **Meu INSS Gov BR – Aplicativo | Previdência Social**. . [S.l: s.n.]. Disponível em: <<https://www.previdenciasocial.net/meu-inss-gov-br-aplicativo-extrato-agendamento/>>. Acesso em: 6 jan. 2023b. , 2016

\_\_\_\_\_. **Resolução Nº 23.669**. . Brasília-DF: Tribunal Superior Eleitoral Secretaria de Gestão da Informação e do Conhecimento Coordenadoria de Jurisprudência e Legislação, 2021. Disponível em: <<https://www.tse.jus.br/legislacao/compilada/res/2021/resolucao-no-23-669-de-14-de-dezembro-de-2021>>. Acesso em: 15 fev. 2023.

BRIDLE, James. **A nova idade das trevas**. São Paulo: Todavia, 2019.

CALZADA PRADO, Javier; MARZAL, Miguel Ángel. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. **Libri**, v. 63, n. 2, jan. 2013. Disponível em: <<https://www.degruyter.com/document/doi/10.1515/libri-2013-0010/html>>. Acesso em: 25 maio 2022.

CARLSON, Jacob et al. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. **Portal: Libraries and the Academy**, v. 11, n. 2, p. 629–657, 2011.

CARMI, Elinor et al. Data Citizenship: Rethinking Data Literacy in the Age of Disinformation, Misinformation, and Malinformation. **Internet Policy Review**, v. 9, n. 2, 28 maio 2020. Disponível em: <<https://policyreview.info/node/1481>>. Acesso em: 30 ago. 2021.

CHAKRABORTY, Rajarshi et al. Online Shopping Intention in the Context of Data Breach in Online Retail Stores: An Examination of Older and Younger Adults. **Decision Support Systems**, v. 83, p. 47–56, mar. 2016.

CHEN, Daqing; SAIN, Sai Laing; GUO, Kun. Data Mining for the Online Retail Industry: A Case Study of RFM Model-Based Customer Segmentation Using Data Mining. **Journal of Database Marketing & Customer Strategy Management**, v. 19, n. 3, p. 197–208, set. 2012.

COULDREY, Nick; MEJIAS, Ulises A. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. **Television & New Media**, v. 20, n. 4, p. 336–349, maio 2019.

DIGILABOUR. **A Sociedade da Plataforma**. . [S.l: s.n.]. Disponível em: <<https://digilabour.com.br/a-sociedade-da-plataforma-entrevista-com-jose-van-dijck/>>. Acesso em: 9 ago. 2021. , mar. 2019

D'IGNAZI, Catherine; BHARGAVA, Rahul. Data visualization literacy: A feminist starting point. In: ENGEBRETSEN, Martin; KENNEDY, Helen (Org.). . **Data Visualization in Society**. [S.l.]: Amsterdam University Press, 2020. p. 207–222. Disponível em: <<http://www.jstor.org/stable/10.2307/j.ctvzgb8c7>>. Acesso em: 19 out. 2020.

ESTEVÃO, Janete Saldanha Bach; STRAUHS, Faimara do Rocio. Letramento informacional para reuso de dados nas ciências sociais: requisitos e competências. **Informação & Informação**, v. 25, n. 2, p. 1–25, 2 jul. 2020.

FECHER, Benedikt; FRIESIKE, Sascha. Open Science: One Term, Five Schools of Thought. In: BARTLING, Sönke; FRIESIKE, Sascha (Org.). . **Opening Science**. Cham: Springer International Publishing, 2014. p. 17–47. Disponível em: <[http://link.springer.com/10.1007/978-3-319-00026-8\\_2](http://link.springer.com/10.1007/978-3-319-00026-8_2)>. Acesso em: 6 jul. 2022.

FILGUEIRAS, Fernando; FLÁVIO, Cireno; PALOTTI, Pedro. Digital Transformation and Public Service Delivery in Brazil. **Latin American Policy**, v. 10, n. 2, p. 195–219, 2019.



- FLUSSER, Vilém. **Língua e Realidade**. São Paulo: Annablume, 2007a.
- \_\_\_\_\_. **O mundo codificado**. São Paulo: Cosac Naif, 2007b.
- FREIRE, Paulo. **Pedagogia da autonomia: saberes necessários à prática educativa**. Rio de Janeiro: Paz & Terra, 2011.
- \_\_\_\_\_. **Pedagogia do oprimido**. São Paulo: Paz e Terra, 2014.
- GOOGLE. **Cloud APIs**. Disponível em: <<https://cloud.google.com/apis>>. Acesso em: 9 jan. 2023.
- GOOGLE. **COVID-19 Community Mobility Report**. Disponível em: <<https://www.google.com/covid19/mobility?hl=en>>. Acesso em: 15 jun. 2022.
- GOVERNO DO ESTADO DE SÃO PAULO. **SP Contra o Novo Coronavírus**. Disponível em: <<https://www.saopaulo.sp.gov.br/coronavirus/>>. Acesso em: 6 jan. 2023.
- GREENBERG, Jane. Metadata and Digital Information. In: BATES, Marcia J.; MAACK, Mary Niles (Org.). **Encyclopedia of Library and Information Sciences, Third Edition**. [S.l.]: CRC Press, 2009. p. 3610–3623. Disponível em: <<https://doi.org/10.1201/9780203757635>>. Acesso em: 25 ago. 2020.
- GUTIÉRREZ, Miren. Algorithmic Gender Bias and Audiovisual Data: A Research Agenda. **International Journal of Communication**, v. 15, n. 0, p. 23, 6 jan. 2021.
- \_\_\_\_\_. How interactive maps mobilize people in geoactivism. In: ENGEBRETSEN, Martin; KENNEDY, Helen (Org.). **Data Visualization in Society**. [S.l.]: Amsterdam University Press, 2020. p. 441–456. Disponível em: <<http://www.jstor.org/stable/10.2307/j.ctvzgb8c7>>. Acesso em: 19 out. 2020.
- \_\_\_\_\_. Participation in a datafied environment: questions about data literacy. **Comunicação e sociedade [Online]**, v. 36, 2019.
- HALLIDAY, Michael Alexander Kirkwood. **Language as social semiotic: the social interpretation of language and meaning**. London: University Park Press, 1978.
- \_\_\_\_\_. The Notion of “Context” in Language Education. In: GHADESSY, Mohsen (Org.). **Text and Context in Functional Linguistics**. Amsterdam: Benjamins, 1999. p. 19–24.
- HUI, Yuk. **Tecnodiversidade**. São Paulo: Ubu Editora, 2020.
- HUYNH, Huy M. et al. Efficient Methods for Mining Weighted Clickstream Patterns. **Expert Systems with Applications**, v. 142, p. 112993, mar. 2020.
- JARDIM, José Maria. A lei de acesso à informação pública. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 5, n. 1, 2 jan. 2012. Disponível em: <<https://revistas.ancib.org/index.php/tpbci/article/view/266>>. Acesso em: 30 nov. 2022.
- JENKINS, Henry. **Convergence culture: where old and new media collide**. New York: New York University Press, 2006.
- JOHNSON, R. Burke; ONWUEGBUZIE, Anthony J.; TURNER, Lisa A. Toward a Definition of Mixed Methods Research. **Journal of Mixed Methods Research**, v. 1, n. 2, p. 112–133, abr. 2007.
- KITCHIN, Rob. Big Data – Hype or Revolution? In: SLOAN, Luke; QUAN-HAASE, Anabel. **The SAGE Handbook of Social Media Research Methods**. 1 Oliver’s Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd, 2016. p. 27–38. Disponível em: <<http://sk.sagepub.com/reference/the-sage-handbook-of-social-media-research-methods/i698.xml>>. Acesso em: 27 fev. 2019.
- \_\_\_\_\_. Big Data, new epistemologies and paradigm shifts. **Big Data & Society**, v. 1, n. 1–2, p. 1–12, 2014.
- KNELL, Mark. The Digital Revolution and Digitalized Network Society. **Review of Evolutionary Political Economy**, v. 2, n. 1, p. 9–25, 1 abr. 2021.

- KRESS, Gunther. Against Arbitrariness: The Social Production of the Sign as a Foundational Issue in Critical Discourse Analysis. **Discourse & Society**, v. 4, n. 2, p. 169–191, 1993.
- LEE, Jungah. **Cuidado, a sua TV Samsung pode estar espiando você**. Disponível em: <<https://economia.uol.com.br/noticias/bloomberg/2015/02/10/cuidado-a-sua-tv-samsung-pode-estar-espiando-voce.htm>>. Acesso em: 6 jan. 2023.
- LIMA-LOPES, Rodrigo Esteves de. Ciência Aberta e suas Contribuições para a Educação Aberta. **Revista Linguagem & Ensino**, v. 25, n. especial, p. 141–155, 22 dez. 2022.
- LIMA-LOPES, Rodrigo Esteves de; CÂMARA, Marco Túlio; OLIVEIRA, Maria Luiza Tápias. Reflexões sobre formação de professores, linguagem e tecnologias. **Comunicação & Educação**, v. 26, n. 1, p. 179–191, 15 jul. 2021.
- MANN, Monique; DALY, Angela. (Big) Data and the North- *in* -South: Australia's Informational Imperialism and Digital Colonialism. **Television & New Media**, v. 20, n. 4, p. 379–395, maio 2019.
- MANOVICH, Lev. Novas mídias como tecnologia e idéia: dez definições. In: LEÃO, Lucia (Org.). . **O chip e o caleidoscópio: reflexões sobre as novas mídias**. São Paulo: Senac, 2005. p. 25–50.
- MEIE, Kenneth J. Replication: A View From the Streets. **PS: Political Science and Politics**, v. 28, n. 3, p. 456–459, 1995.
- MEJIAS, Ulises A.; COULDREY, Nick. Datafication. **Internet Policy Review**, v. 8, n. 4, 29 nov. 2019. Disponível em: <<https://policyreview.info/concepts/datafication>>. Acesso em: 30 ago. 2021.
- MOROZOV, Evgeny. **Big Tech**. São Paulo: Ubu Editora, 2018.
- \_\_\_\_\_. **The net delusion**. 1st ed ed. New York: Public Affairs, 2011.
- MOTTA-ROTH, Désirée. Letramento científico: sentidos e valores. **Notas de Pesquisa**, v. 1, n. 0, p. 12–25, 2011.
- NEMER, David. Wired mobile phones: the case of community technology centers in favelas of Brazil. **Information Technology for Development**, v. 24, n. 3, p. 461–481, 3 jul. 2018.
- NOBLE, Safiya Umoja. **Algorithms of oppression: how search engines reinforce racism**. New York: NYU Press, 2018.
- OECD. **PISA 2018 Science Framework**. . Paris: OECD Publishing. Disponível em: <<https://www.oecd-ilibrary.org/content/component/f30da688-en>>. , 2019
- OLBRICH, Rainer; HOLZING, Christian. Modeling Consumer Purchasing Behavior in Social Shopping Communities with Clickstream Data. **International Journal of Electronic Commerce**, v. 16, n. 2, p. 15–40, dez. 2011.
- OLIVEIRA, Joacyr Tupinambás de. **Um corpus de aprendizes de tradução**. 2019. Doutorado em Estudos Lingüísticos e Literários em Inglês – Universidade de São Paulo, São Paulo, 2019. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8147/tde-31102019-183855/>>. Acesso em: 9 fev. 2023.
- POELL, Thomas; NIEBORG, David; VAN DIJCK, José. Platformisation. **Internet Policy Review**, v. 8, n. 4, 29 nov. 2019. Disponível em: <<https://policyreview.info/node/1425>>. Acesso em: 16 nov. 2022.
- PYTHON. **Python.Org**. Disponível em: <<https://www.python.org/>>. Acesso em: 9 fev. 2023.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2021. Disponível em: <<https://www.R-project.org/>>.



- RIEDER, Bernhard. **Engines of Order**. Amsterdam: Amsterdam University Press, 2020. Acesso em: 11 jul. 2020.
- SEGURA, María Soledad; WAISBORD, Silvio. Between Data Capitalism and Data Citizenship. **Television & New Media**, v. 20, n. 4, p. 412–419, maio 2019.
- SILVA, Wagner Rodrigues. Educação científica como abordagem pedagógica e investigativa de resistência. **Trabalhos em Linguística Aplicada**, v. 59, p. 2278–2308, 5 mar. 2021.
- SINGH, Gregory Matthew. **The death of Web 2.0**. London/New York: Routledge Taylor & Francis Group, 2019.
- STRAUBHAAR, Joseph D.; SPENCE, Jeremiah. Usage Dynamics of Telecenters in São Paulo. **INFORMÁTICA NA EDUCAÇÃO: teoria & prática**, v. 10, n. 1, p. 11–26, 2007.
- TRE-SP. **TRE-SP vai usar dados biométricos do Denatran para identificar eleitores em outubro. Justiça Eleitoral**. [S.l.: s.n.]. Disponível em: <<https://www.tre-sp.jus.br/comunicacao/noticias/2022/Setembro/tre-sp-vai-usar-dados-biometricos-do-denatran-para-identificar-eleitores-em-outubro>>. Acesso em: 15 fev. 2023., 9 jan. 2022
- TWITTER. **Twitter API for Academic Research | Products**. Disponível em: <<https://developer.twitter.com/en/products/twitter-api/academic-research>>. Acesso em: 9 jan. 2023.
- TYGEL, Alan Freihof; KIRSCH, Rosana. Contributions of Paulo Freire for a Critical Data Literacy: A Popular Education Approach. **The Journal of Community Informatics**, v. 12, n. 3, p. 108—121, 2016.
- WAJCMAN, Judy. Reflections on Gender and Technology Studies: In What State Is the Art? **Social Studies of Science**, v. 30, n. 3, p. 447–464, jun. 2000.
- WHEELER, David A. **Free-Libre / Open Source Software (FLOSS) is Commercial Software**. Disponível em: <<https://dwheeler.com/essays/commercial-floss.html>>. Acesso em: 15 fev. 2023.
- WOLFF, Annika et al. Creating an Understanding of Data Literacy for a Data-Driven Society. p. 9–26, 2016.
- WOLFF, Annika; MONTANER, Jose J Cavero; KORTUEM, Gerd. Urban Data in the Primary Classroom: Bringing Data Literacy to the UK Curriculum. **The Journal of Community Informatics**, v. 12, n. 3, p. 57–82, 2016.
- YEE-LOONG CHONG, Alain et al. Online Banking Adoption: An Empirical Analysis. **International Journal of Bank Marketing**, v. 28, n. 4, p. 267–287, 15 jun. 2010.
- ZUBOFF, Shoshana. **The Age of Surveillance Capitalism**. Main edição ed. [S.l.]: Profile Books, 2019.