

Contribuições dos estudos de discurso na análise da comunicação em plataformas digitais

Ruth Reis¹ e Daniela Zanetti²

Resumo

O propósito deste artigo é discutir a contribuição dos estudos de discurso no contexto da comunicação contemporânea, a construção de *corpora* e estratégias de análise dos processos discursivos em ambiente on-line. Marcado pelas mídias sociais, o ecossistema comunicacional da atualidade coloca em cena uma miríade de atores que se movimentam numa trama de múltiplas conexões e têm nas suas textualidades uma das formas de produção do social. Inicialmente, são apresentadas as principais contribuições teóricas dos estudos de discurso que são apropriadas pelo campo da Comunicação, incluindo autores como Foucault, Pêcheux, Fairclough, Dijk, Laclau, Mouffe, Paveau entre outros. Numa segunda parte, são discutidos alguns dilemas metodológicos e procedimentos possíveis de coleta e análise dos dados. Tendo a revisão bibliográfica como caminho metodológico, bem como a experiência no desenvolvimento de trabalhos de análise do discurso em ambientes digitais, apresentamos reflexões sobre os desafios de pesquisa e destacamos a importância da conexão entre o trabalho humano desempenhado pelo pesquisador e o trabalho maquínico desenvolvido por softwares na obtenção de resultados profícuos para análise do discurso em mídias sociais digitais.

Palavras-chave

Comunicação; Discurso; Plataformas digitais; Metodologia; Textualidades.

¹Professora do Programa de Pós-Graduação em Comunicação e Territorialidades da Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil.
E-mail: ruth.reis@ufes.br.

²Professora do Programa de Pós-Graduação em Comunicação e Territorialidades da Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil.
E-mail: daniela.zanetti@ufes.br.

Contributions of discourse studies to the analysis of communication on digital platforms

Ruth Reis¹ and Daniela Zanetti²

Abstract

The purpose of this article is to discuss the contribution of discourse studies to contemporary communication, the construction of corpora, and processes of discursive analysis strategies in online environments. Shaped by social media, today's communicational ecosystem features a myriad of actors put on the scene, navigating a web of multiple connections and having in their textualities one of the forms of social production. The article begins by presenting the main theoretical contributions of discourse studies that have been adopted in the field of communication, including authors such as Foucault, Pêcheux, Fairclough, Dijk, Laclau, Mouffe, and Paveau, among others. The second part discusses some methodological dilemmas and possible procedures for data collection and analysis. Through a bibliographic review of the methodological approach, as well as our experience in conducting discourse analysis in digital studies, we reflect on research challenges and highlight the importance of connecting human effort, performed by the researcher, with machinic work carried out by software, to achieve meaningful results for discourse analysis on digital social media platforms.

Keywords

Communication; Discourse; Digital platforms; Methodology; Textualities.

¹Professora do Programa de Pós-Graduação em Comunicação e Territorialidades da Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil.
E-mail: ruth.reis@ufes.br.

²Professora do Programa de Pós-Graduação em Comunicação e Territorialidades da Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil.
E-mail: daniela.zanetti@ufes.br.

Introdução

As transformações da comunicação desde o início da era digital, especialmente a partir dos primeiros anos deste século, trazem novas indagações a respeito das apropriações e dos usos e efeitos produzidos nos sistemas de mediação comunicacional disponíveis atualmente, configurados como uma rede distribuída que permite inúmeras conexões em diversos níveis e escalas, envolvendo humanos e não humanos (Latour, 2012). Uma parte dessas indagações diz respeito aos sentidos produzidos pelos atores que se instituem e se movimentam nessa trama, que tem nas textualidades o principal meio de produção do social em suas múltiplas dimensões. Entendidos em sentido amplo e múltiplo, os textos inscritos no atual sistema de comunicação inserem os integrantes dessa rede em processos sociais e discursivos que se forjam por meio das ferramentas tecnológicas disponíveis para as interações, hoje, fortemente administradas por dispositivos de inteligência artificial criados para otimizar os resultados materiais e simbólicos auferidos por quem as controla.

O avanço das grandes plataformas de mídias sociais neste século, e a adoção de estratégias de governança de dados baseadas em inteligência artificial resultaram num ambiente comunicacional sob domínio dos gigantes da comunicação digital, lideradas por empresas como Google, Microsoft, Meta, X, ByteDance e suas respectivas plataformas de mídias sociais, como YouTube, Facebook, Instagram e TikTok.

Essas grandes corporações conseguiram estabelecer um paradigma hegemônico na gestão de redes sociais digitais – e também na sociabilidade contemporânea –, financiadas por meio da publicidade comercial e do setor público, otimizada pelo controle que exercem sobre os usuários ao coletarem seus dados textuais e comportamentais, processo que passou a ser conhecido como plataformação da web (Helmond, 2015) e sociedade da plataforma (Dijck; Poell; Waal, 2018).

À medida que nos apossamos dos recursos oferecidos por esses dispositivos, inscrevemos neles nossa singularidade, que por sua vez é constituída por um complexo jogo de determinações sociais, culturais, psicológicas e econômicas. Essa inscrição se dá por meio da produção discursiva de múltiplos sujeitos, de forma voluntária, nas interfaces digitais, e, também, dos rastros digitais que deixamos de um modo involuntário quando os utilizamos.

Trata-se de algo que se manifesta, numa face, como textos (orais, escritos, audiovisuais, fotográficos ou mixagem de todos), que podem ser lidos e interpretados, e, noutra, como metadados – humanos e sociais – tais como sentimentos, emoções, preferências, imagem corporal, marcas ideológicas, localização etc., que se incrustam inadvertidamente nos territórios dessa máquina de comunicação, sendo passíveis de captura e governança, instituindo-se assim novas formas de poder, subjetivação e submissão.

Interessa-nos mais especificamente neste trabalho abordar a face das

textualidades e dos sentidos que podem ser produzidos nesses territórios digitais. Essas textualidades se colocam entre o visível e o legível, em especial nas plataformas de mídias sociais, e constituem o caminho para compreender os aspectos discursivos que se manifestam, pois são elas que, em boa parte, determinam as trocas comunicacionais realizadas nesse universo. Também consideramos que essa materialidade linguística e imagética é a expressão, por excelência, das políticas de comunicação dos usuários e de suas estratégias nesse território comunicacional.

Por meio de textualizações diversas – imagens em movimento ou não, textos grandes ou diminutos, memes, desenhos, *mashups*, entre outras formas discursivas – produzimos sentidos que são compartilhados, compondo uma territorialidade informacional digital. Para tal, vamos nos ater aos estudos de discurso, buscando compreender alguns aspectos que se colocam a partir dessa perspectiva teórico-metodológica, para em seguida delinear desafios metodológicos, especialmente no que se refere à constituição de *corpora* de pesquisa e processos de análise que possam nos conduzir à investigação propriamente discursiva da comunicação nas redes sociais digitais.

O que pretendemos, mais pontualmente, é desafiar esse campo dos estudos e a análise do discurso como fornecedores de chaves interpretativas e metodológicas para aprofundar a compreensão dos processos comunicacionais contemporâneos e seus impactos sobre o social, o político e o cultural, consideradas algumas de suas características como o descentramento, a fragmentação, a velocidade, a instantaneidade e sua grande escala.

Essas características trazem desafios metodológicos para o campo acadêmico que precisam ser enfrentados para se compreender as dinâmicas discursivas realizadas em rede nas mídias sociais e respectivos processos de participação na vida social, política e cultural. Não é pretensão deste trabalho responder a todo esse conjunto de temas, mas apenas dar relevo a alguns dos dilemas metodológicos com que temos nos deparado no presente no mundo da Comunicação e evocar conceitos que possam contribuir para criar um lugar mais seguro para nossos gestos de leitura (Orlandi, 2014, p. 16). Num primeiro momento faremos um breve apanhado dos estudos de discurso, sem o propósito de descrever todo o seu percurso. Em seguida nos deteremos sobre a questão da composição do *corpus* de pesquisa em redes sociais digitais, um dos primeiros desafios a considerar. Por fim, tratamos brevemente sobre recursos e dinâmicas de análise.

Estudos de discurso e seus fundamentos

Desenvolvidos a partir de um conjunto de autores provenientes da Linguística, Filosofia, Literatura, Psicanálise, Antropologia entre outros campos teóricos, os estudos de discurso sempre se propuseram a se colocar num intermeio dessas

ciências. Embora tenham se desdobrado em diversas vertentes, estas se unificam na concepção geral de que as atividades de linguagem têm imbricação profunda com o social, constituindo-se como parte integrante e irreduzível deste. Eles foram fortemente apropriados pelo campo da Comunicação nas investigações de produtos da indústria cultural à medida que se desenvolviam, mais expressivamente na segunda metade do século XX.

Essa contribuição teórica alcançou o século XXI em meio a diversas revisões, retomadas, silenciamentos e renovações. Os estudos do discurso ganham uma diversidade de abordagens que procuram dar conta de variáveis que incidem sobre a realização da atividade linguageira e dos sentidos que produzem. Essas variáveis podem se relacionar ao contexto em que o discurso se desenvolve, às influências do passado que se incrustam nas falas do presente, às condições psicológicas e sociais de cada indivíduo, aos processos materiais e simbólicos de mediação, aos lugares assumidos pelos falantes, enfim, a um conjunto de condições que compreendem conceitos como formações discursivas, interdiscurso, intradiscurso, condições de produção do discurso, forma e posições-sujeito, sistemas de autoridade, e outros, como propuseram Foucault, Pêcheux, Paul Henry, Denise Maldidier, Jean Jacques Courtine, entre outros autores.

De forma bastante sintética, vale recuperar algumas contribuições de autores fundadores que podem ser úteis aos estudos de discurso aplicados aos novos contextos comunicacionais. Na Análise de Discurso francesa, as contribuições de Foucault (2008) e Pêcheux (1975) são consideradas inaugurais, mesmo que ambos tenham trilhado percursos teóricos diferentes. Foucault, ao estudar as formas de emergência do saber das ciências sociais e humanas, ocorridas no século XIX, toma o discurso como conceito central para a constituição do conhecimento, e defende “não mais tratar os discursos como conjuntos de signos (elementos significantes que remetem a conteúdos ou a representações), mas como práticas que formam os objetos de que falam” (Foucault, 2008, p. 55).

Para Pêcheux (1975), o discurso é uma categoria teórica abrangente o suficiente por abarcar língua e fala em plena atividade, evocando suas dimensões históricas, sociais, psicológicas e linguísticas, e sendo um instrumento da prática política que tem como função transformar as relações sociais. Pêcheux entende que o discurso se estrutura em torno das relações de poder, da ideologia e das condições materiais de existência, imbricadas na singularidade da dimensão psicológica de cada indivíduo. Ele o vê sempre presente nas disputas que emolduram o social, como uma instância de materialização da ideologia.

Autores da Análise Crítica do Discurso (ACD), de tradição anglo-saxônica, ganham relevância nos anos 1990, a partir do trabalho do linguista britânico Norman Fairclough (2001). Ele adota uma concepção tridimensional do discurso, entendendo-o como composto pelas práticas sociais, práticas discursivas (envolvendo as condições

de produção, distribuição e consumo) e pelos textos (material semiótico). Para ele, a prática discursiva contribui para reproduzir a sociedade (identidades sociais, relatos sociais, sistemas de conhecimento e crença), e, também, para transformá-la.

Teun Van Dijk (2012), outro autor referência nos estudos de discurso críticos, defende a existência de modelos mentais que intercedem nas situações de comunicação, mediando o exterior e um interior. Ao destacar a importância do contexto nos processos interacionais, ele afirma que “são construtos dos participantes” (Dijk, 2012, p. 11), o que não significa que as estruturas sociais e políticas não tenham dimensões objetivas. Estas existem, mas só conseguem influenciar o discurso através das interpretações (inter)subjetivas dos participantes. Para Dijk, os contextos controlam “os processos da produção e a compreensão do discurso” (Dijk, 2012, p. 35).

Trazendo outras contribuições e instrumentos de análise, Laclau e Mouffe (Escola de Essex) defendem que o discurso não se resume apenas ao texto ou à fala, mas a uma totalidade que inclui o linguístico e o extralinguístico, uma vez que o que está fora da linguagem só pode ser apreendido dentro de uma “prática articulatória” (Laclau; Mouffe, 2015, p. 167) de natureza discursiva. Isso implica não considerar apenas as superfícies textuais *stricto sensu*, mas todo o conjunto de textos e ações que se dão no âmbito social. Conceitos centrais de Laclau e Mouffe (2015) são “antagonismo” e “deslocamento”, que designam o momento de disputa pela fixação dos sentidos.

Mais contemporaneamente, Paveau (2010; 2021) realiza incursões importantes na compreensão do discurso nas plataformas digitais, desenvolvendo conceitos como tecnodiscurso e tecnopalavras. A ênfase no prefixo *tecno* ressalta o propósito de deslocar a compreensão das formas discursivas no ambiente digital, mas principalmente de considerar que o tecnológico se imiscui no discurso porque também é produtor de sentidos.

A natureza multifacetada do discurso oferece condições para as múltiplas possibilidades de abordagem que buscam compreender desde os processos de produção aos seus efeitos, as estratégias adotadas pelos falantes até os modos de sujeição/subjetivação sociodiscursiva. Essa característica diversa e múltipla é sua riqueza, o que nos impõe uma atitude polideterminística quando buscamos compreender as inúmeras forças que concorrem para a produção discursiva e tudo o que dela advém. Também nos indica a possibilidade de encontrar no discurso, se não uma ontologia híbrida, como propõe Latour (2012), pelo menos um ponto de convergência de diversas potências que o constituem.

Os estudos de discurso oferecem um aporte teórico e metodológico importante para melhor compreender os processos comunicacionais que se dão no contemporâneo, nos quais se destaca um sistema midiático complexo, que amplia e diversifica as técnicas e as práticas de produção discursiva, oferecendo recursos expressivos e um sistema de interconexões historicamente inédito. Mas é necessário tomar as contribuições que os estudos de discurso nos legaram e desafiá-las a

convergir para as práticas discursivas contemporâneas, deslocando-as e atualizando-as teórica e metodologicamente.

O *corpus* de pesquisa e os novos dispositivos comunicacionais

Com a internet e a cultura das redes digitais, novas experiências discursivas foram adotadas como rotina na sociedade, configurando uma trama comunicacional historicamente inédita em quantidade e qualidade. É uma máquina movida por muitas mãos e vozes, que se alimenta das contribuições que cada um dos seus participantes oferece e que se transformou no centro da sociabilidade contemporânea. Os algorítmicos, dispositivos de inteligibilidade das máquinas, se tornaram centrais na organização e disponibilização do discurso em rede, acionado por uma miríade de singularidades produtoras. As textualidades que produzem podem ser consideradas o ponto para o qual convergem as dinâmicas instituidoras do social e de cada um dos sujeitos que atua na rede, um território não linear, em que o tempo e o espaço se ressignificam.

As práticas discursivas na internet assumem, assim, um papel central para a compreensão dos fenômenos sócio-históricos midiáticos, e tomam o lugar – substituindo ou complementando – de outras, provenientes do modo presencial e não midiático, e também as modifica. Diante dessas condições, Rogers (2013) chama a atenção para o desafio de se articular métodos de pesquisa tradicionais que se adaptem às formas digitais com aqueles que são nativamente digitais (por exemplo, os sistemas de recomendação das plataformas digitais).

Nessa linha, defendemos a importância de associar as diferentes abordagens da análise do discurso aos processos e métodos digitais e de explorar as perspectivas teóricas e metodológicas das vertentes dos estudos de discurso, considerando os elementos sócio-históricos que interpelam os indivíduos e os institui como sujeitos inscritos num determinado mundo, no qual desenvolvem seus dizeres e suas práticas e vivem a sua ordem e circuitos de poder.

O surgimento do digital transforma todo o ecossistema em que a vida se desenvolve, constituindo modos de existência – ou formas de subjetivação/sujeição – que diferem daqueles que vigoravam no momento inaugural dos estudos de discurso, e reconfiguram os mecanismos de poder. É nesse universo mediado pela inteligência artificial que se dá a produção discursiva, uma competência que deixa de ser apenas humana, pois agora abarca também a agência das máquinas por meios de algoritmos criados e treinados para atuar nessa sociedade com níveis variados de protagonismo.

Nos dispositivos de comunicação contemporâneos, os estudos de discurso requerem, em geral, a adoção de instrumentos informatizados para que se possa dar conta dos diversos problemas de pesquisa. Estudar as redes sociais nas plataformas de mídias digitais exige alguma capacidade de acessar e compreender bancos de dados

digitais, selecionar frações do grande universo de interações que nelas se realiza e promover leituras possíveis do conjunto de dados obtidos.

Hoje, no campo da Comunicação, há carência de formação e instrumentos adequados para essas ações, condição que precisa ser superada com a contribuição de outros campos de estudos, especialmente os das áreas das ciências da computação e de dados. Mesmo aqueles pesquisadores que dominam a linguagem das máquinas e conseguem agregar ferramentas e conhecimento para explorá-las, encontram limitações seja por determinações legais ou por regulagens internas das empresas que as controlam. As restrições de acesso às *Application Programm Interface* (APIs), em 2024, pelas plataformas mais estudadas no campo acadêmico (X e Facebook/Instagram) [1] criam novos obstáculos à pesquisa acadêmica e tornam ainda mais impermeável o conhecimento sobre o funcionamento dos algoritmos que governam as relações constituídas nesses dispositivos.

Pêcheux foi um dos primeiros teóricos a apostar na contribuição da informática e dos algoritmos como instrumentos para compreender as questões discursivas. Essa inclinação pela inteligência artificial o levou aos experimentos inaugurais do que chamou de análise automática de discurso (AAD69), no final dos anos de 1960, e ao desenvolvimento de softwares para criação e análise de *corpora* textuais e das operações discursivas com recursos informatizados. Esse projeto se mostrou sempre incompleto, devido às limitações tecnológicas da época e à complexidade do empreendimento teórico por ele proposto (Zanella, 2017).

A inclinação de Pêcheux (2014) para promover esse tipo de interdisciplinaridade se explica pela aversão ao distanciamento entre as “culturas literária e científica”, se referindo à oposição entre os trabalhos das Ciências Sociais e da Filosofia e os baseados na lógica matemática. Ao contrário, ele via um caminho para a convergência desses dois universos, questão que se torna ainda mais relevante hoje, quando se buscam alternativas teóricas e metodológicas.

Um dos desafios que se apresenta para qualquer pesquisador a partir da definição do seu problema e objetivos de pesquisa que envolvem dados digitais é a composição do *corpus*. Este corresponde a um determinado recorte, previamente definido pelo analista, que lhe servirá de matéria a ser observada e inquirida na fase de análise. O vasto poder de registro por parte dos dispositivos tecnológicos das plataformas digitais configurou o que se convencionou denominar *big data*, em alusão à quantidade, velocidade e variedade de dados coletados, processados e armazenados continuamente. Isso nos leva ao potencial de constituição de uma memória das práticas discursivas inédita na história da humanidade, pois permite que se analise os recortes de questões materializadas em arquivos de dados digitais mais diversos em complexidade e quantidade. Pêcheux (2014, p. 2) já apresentava interrogações sobre as dificuldades metodológicas de “ler o arquivo”, procurando sempre desmistificar a neutralidade dos processos de pesquisa.

O conceito de arquivo pode ser encontrado tanto nos escritos de Foucault (2008) quanto nos de Pêcheux (2014) com perspectivas relativamente similares: Foucault entendia arquivo como a fonte das regularidades que permite a existência de um conjunto de enunciados que encontraram seu lugar único no mundo. Estes são reunidos mediante políticas de discursividade atravessadas pelas permissões e vedações políticas, institucionais, culturais ou tecnológicas de cada momento histórico-social. “O arquivo é, de início, a lei do que pode ser dito, o sistema que rege o aparecimento dos enunciados como acontecimentos singulares [...]. É o sistema geral da formação e da transformação dos enunciados” (Foucault, 2008, p. 147).

Pêcheux toma o conceito de arquivo como esse grande universo delineado por Foucault e, também, como parte do processo de investigação, entendido como “campo de documentos pertinentes e disponíveis sobre uma questão”. Mas tem consciência, especialmente na segunda fase da AD, de que não se trata de algo já dado ou unitário, mas de um produto determinado por múltiplos processos de ordem social e linguística, sempre atravessado pela ideologia, conceito central na sua teoria. A esses *corpora* discursivos ele denominava “conjunto de superfícies”, gerados a partir do dispositivo informático (Pêcheux, 2015).

A constituição do *corpus* para implementação de pesquisa nas redes digitais precisa passar por um conjunto de decisões e procedimentos que se inicia quando o pesquisador estabelece a questão e delinea o problema que deseja investigar. O primeiro momento é acessar o grande arquivo resultante de todas as múltiplas trocas comunicacionais, que se coloca disponível no presente na forma de dados digitais, armazenados e controlados pelos inúmeros sistemas e plataformas, e dele segmentar o que interessa para responder as interrogações que dão origem à pesquisa. Como dito anteriormente, o auxílio de ferramentas computacionais se torna essencial nesse momento, uma vez que a capacidade humana se mostra limitada tanto para obter os dados que comporão o *corpus* quanto para analisá-los.

Devemos considerar quais plataformas pesquisar segundo os fins da pesquisa, tendo em conta que cada uma adota diferentes *affordances* que influenciam as formas de expressão e interação. Se tomarmos como exemplo o Instagram, uma das redes preferidas no Brasil, observamos que há três grandes fluxos numa mesma plataforma, o *feed*, os *stories* e os *reels*, que implicam estratégias discursivas diferentes. O *feed*, mais focado nas imagens e legendas, tem mais permanência na plataforma; os *stories* é o lugar para as mensagens efêmeras, porém oferece mais recursos de formação de redes e contexto por meio de *links* clicáveis, além de interações; os *reels*, dedicado aos conteúdos de vídeo, além de *streaming* ao vivo. Cada uma dessas trilhas permite criar produtos diferentes e reúne públicos diferentes.

Isso nos leva aos primeiros e importantes desafios para conciliar e, também, renovar as perspectivas colocadas pelos estudos de discurso, uma vez que para estes, a dimensão quantitativa tem menor peso do que a qualitativa. Mas é preciso enfrentar

a quantidade de dados que são gerados pela atividade comunicacional constituída pela agência de inúmeros actantes, e encontrar procedimentos para organizá-la e situá-la num nível de compreensão que permita abrir caminhos para a análise a ser realizada pelo pesquisador humano. Para isso, primeiro é necessário obter a amostra a ser investigada, o que se pode fazer pelo recurso a softwares de coleta de dados imediatamente visíveis (textos, imagens, registros de perfis cadastrados) e respectivos metadados nas plataformas de mídias digitais.

Se o objetivo de pesquisa é entender os processos discursivos travados sobre determinados temas pelos inúmeros perfis nas redes, a porta de entrada principal para a constituição do recorte que comporá o *corpus* de pesquisa são as inscrições propriamente linguísticas (as palavras) ou imagéticas (fotografias, desenhos, materiais audiovisuais, entre outros). Portanto, busca-se por palavras que evocam os temas de interesse durante o período definido. Para os estudos de discurso, as palavras não são consideradas plenas de sentido se não estiverem imersas nos seus contextos não apenas linguísticos como também sociais. Entretanto, devemos entender as palavras e outras materialidades textuais como pegadas para a compreensão dos valores investidos e, portanto, dos sentidos que os textos produzem. Se o objetivo é encontrar a atividade discursiva de actantes específicos, a busca de dados pode se restringir a apenas um perfil em redes sociais e circunscrever um tempo determinado. Esses dados iniciais obtidos são quantitativos e formam o caminho para uma primeira visada dos aspectos qualitativos.

Nesse momento a compreensão do arquivo (no sentido foucaultiano) que delinea a curiosidade da pesquisa é essencial, pois uma parte dos resultados obtidos na construção dos *corpora* decorre da leitura que se tem do mundo. Combinada com a clareza do problema de pesquisa, é essa compreensão que determina a escolha das palavras-chave que permitirão recuperar o conjunto de dados para posterior análise.

Ao trabalhar a questão *corpus* de discurso e *corpus* de textos, Charaudeau, tomando em conta a diversidade de definições de discurso e texto, propõe considerar que a relação de complementaridade entre análise de texto e análise de discurso não é simétrica, embora um e outro estejam entrelaçados. “O discurso não é o texto, mas ele é carregado por textos. O discurso é um percurso de significância que se acha inscrito num texto, e que depende de suas condições de produção e dos locutores que o produzem e o interpretam” (Charaudeau, 2011, p. 6). O autor ressalta ainda a importância de se ter clara a problemática de pesquisa quando se busca construir um *corpus*, e as interrogações que endereçam aos interesses de pesquisa.

A constituição de um *corpus* de análise se torna ainda mais complexa nos ambientes digitais, em especial nas mídias sociais, se levamos em conta características assinaladas por Dias (2016), baseada em Paveau, tais como: instabilidade (as inscrições em redes sociais são sujeitas a mudanças, com acréscimos e supressões); temporalidade (o arquivo digital é sempre trazido para o presente ou atualizado pelo acesso, não

importando quando foi criado); dimensão e heterogeneidade (grande quantidade de dados e profusão de proposições e pensamentos); autoria (diversidade de criadores) e leitura dispersiva (quebra do fio do discurso).

Acrescente-se ainda a diversidade de recursos enunciativos presentes na escrita digital, que se caracteriza por emular as estratégias utilizadas presencialmente, como desenhos, trechos de filmes, *emojis*, entre outros. Também podem se apresentar como expressões típicas do ambiente digital as mixagens ou manipulações de sons e imagens, exigindo um letramento específico tanto para o trabalho de escrita e leitura cotidiana, quanto para uma interpretação sistemática nas pesquisas científicas.

O sistema algorítmico presente hoje nas plataformas da internet, com todas as suas dinâmicas de abertura e fechamento, é um ator destacado não só na produção do grande arquivo preconizado por Foucault e Pêcheux, como na constituição dos *corpora* de pesquisa. Construir um arquivo como um *corpus* de pesquisa implica em submeter-se às vedações e permissões dos algoritmos, também estes constituídos com boas doses de cultura e não apenas da mais pura matemática.

Algoritmos são instruções para o desempenho de uma atividade e estas são definidas por seus criadores, hoje caracterizadas como grandes empresas do capitalismo global, como já mencionado antes. Por mais impermeáveis que ainda sejam, pesquisas científicas (Zuboff, 2021; Pariser, 2012; Doneda e Almeida, 2018) e os usos cotidianos dos sistemas algorítmicos demonstram que seu desempenho se dá a partir do conhecimento gerado pelo rastreamento de dados individuais, que revelam preferências, hábitos, comportamentos e outros traços usados para entregar conteúdos de forma hipersegmentada. Exemplos dessa dimensão cultural dos algoritmos são os achados demonstrando o racismo algorítmico que afeta a população negra e outros grupos sociais minorizados (Silva e Silva, 2024) e a presença de vieses político-partidários em situações de disputa eleitoral (Reis; Zanetti; Frizzera, 2020).

É importante saber também que os recortes que serão obtidos para análise são pequenas frações do vasto mundo inscrito no digital, sob forte controle das plataformas que se operam as redes sociais, o que não permite dimensionar o grau exato de representatividade das amostras obtidas. Além disso, essas amostras têm mudado suas características desde o evento da Cambridge Analytica (Kaiser, 2020), que desencadeou uma série de iniciativas para controlar apropriações e manipulação de dados pessoais, com novas autorregulações e regulações estatais. Essas questões, além de moldar a forma dos *corpora* de pesquisa e limitar o acesso acadêmico aos dados gerados no ambiente digital, promove novas questões éticas em todas as fases da pesquisa científica [2] (Franzke *et al.*, 2020), tema que não será tratado aqui dada a sua especificidade e aos objetivos deste artigo.

Na experiência contemporânea de pesquisa em mídias digitais, podem ser usadas estratégias manuais para extrair uma porção textual de menor tamanho, mas os maiores desafios residem na criação de grandes *corpora* (*big data*). Estes exigem

procedimentos computacionais, na forma de *scripts* desenvolvidos especificamente para uma pesquisa ou pela aquisição de serviços de terceiros. Para as buscas, mais frequentemente, são criados parâmetros por meio de palavras-chave. O desafio, no momento, é identificar palavras que mais se aproximem dos sentidos que se conectem com a questão investigada: as que apresentem maior investimento de textualidade, como propõe Orlandi (1995, p. 111), pois “quando uma palavra significa é porque ela tem textualidade, ou seja, porque sua interpretação deriva de um discurso que a sustenta, que a provê de realidade significativa”.

Percursos de análise

Os resultados obtidos em pesquisas que envolvem *big data*, por exemplo, correspondem a um conjunto de informações que ainda não têm o valor da significação. Esta só ganha corpo à medida que as informações forem selecionadas e organizadas para posterior análise do pesquisador, que traz a dimensão humana, importante e necessária, para a interpretação dos processos discursivos. O momento do encontro entre o pesquisador e seu *corpus* de análise é sempre um desafio árduo. Ambos são confrontados por mediações diversas, como o problema de pesquisa, os objetivos, as opções teóricas em que o pesquisador se situa, além de outros de natureza social ou subjetiva. É desse confronto que se dará a interpretação e a compreensão. Por isso, é importante que se assinale o lugar do pesquisador nesse processo, a quem caberá a dimensão interpretativa que realizará a partir dos *outputs* gerados pelos seus assistentes automatizados na busca de compreensão dos sentidos construídos pelas trocas discursivas.

Um primeiro caminho possível visando alcançar a complexidade da rede é entender, no *corpus* selecionado, as posições relativas e os indícios do contexto partilhado pelos actantes, sejam esses perfis humanos ou robóticos. Com isso é possível mapear uma parte das condições de produção do discurso, antes mesmo de explorar as textualidades. É importante que, ao constituir o arquivo destinado à exploração final do pesquisador sejam selecionados dados que não estão expressos nos textos ou imagens presentes na camada visível do discurso digital.

No sistema digital se situam também elementos que ficam total ou parcialmente escondidos na forma de metadados. Estes podem evidenciar tanto a dimensão social dos enunciadores e seus lugares de fala (por exemplo, usando algoritmos que permitem visualizar *clusters*, graus de entrada e saída que permitem perceber as redes de relações de influência), quanto sua relação com os próprios dispositivos de comunicação (mídias, linguagens ou instrumentos computacionais utilizados, entre outros). Para essas atividades, podem-se criar ferramentas próprias ou recorrer a softwares proprietários ou abertos, pagos (Excel/Microsoft, ou Numbers/Apple, por exemplo) ou gratuitos (Libre Office, Tableau Public, Gephi, entre outros [3]), por meio

das quais se terá uma visão, por exemplo, dos principais atores, posts mais ranqueados, datas e horários de circulação de mensagens, entre outras informações disponíveis.

Esses aspectos do processo discursivo digital, obtidos por meio dos metadados, remetem para um âmbito extralinguageiro e contribuem para se obter uma visão mais abrangente do que Pêcheux chamava de condições de produção do discurso, e que Paveau (2021) ressignifica para o conceito de ambiente, definindo-o como o “conjunto de dados humanos e não-humanos no âmbito dos quais os discursos são elaborados” (Paveau, 2021, p. 13-14).

Esses dados devem ser lidos como signos que apontam para a vida ordinária dos actantes, sua sociabilidade, conexões nas redes em que atuam, regularidades discursivas, entre outros elementos importantes para construir o dispositivo analítico. Esse percurso inicial permite ao pesquisador elaborar as primeiras visadas sobre o *corpus* que tem em mãos. É uma tarefa realizada numa articulação humano-máquina, continuamente revisada e reconfigurada, que vai permitir encontrar na dispersão os sentidos que os discursos constituem e por eles são constituídos.

Um segundo momento deve ser dedicado ao tratamento e análise dos dados linguísticos e imagéticos. No caso de grandes volumes de dados textuais sempre haverá necessidade do auxílio de softwares, que contribuirão para que se constituam as primeiras leituras do *corpus* selecionado. À medida que as pesquisas sobre linguagem natural informatizadas se incrementaram, surgiram também diversos instrumentos automatizados de análise, aplicáveis principalmente a *corpora* extensos, impossíveis de serem analisados sem eles. Para tal se recorre a aplicativos digitais, alguns de livre acesso (Voyant Tools, Iramuteq, Gephi, por exemplo), que permitem organizar e construir sínteses visuais na forma de grafos ou gráficos lastreados nos dados presentes no corpus que são tratados por meio de diferentes parâmetros. Isso favorecerá fortemente a interpretação dos dados e sua conversão em significação. Alguns exemplos de aplicações desses recursos podem ser encontrados em Reis, Zanetti e Frizzera (2020), Souza (2024), Jacobsen e Reis (2022), Medeiros (2022) e Haacke e Lima (2020).

Esses aplicativos podem abarcar uma diversidade de técnicas e métricas que dão conta da superfície textual de *corpus*, como ocorrências, frequência e coocorrência de termos, palavras em contexto, entre outros. Mais recentemente os recursos de *Large Language Model* (LLM), baseados em algoritmos que trabalham com dados não estruturados como os de linguagem natural também podem contribuir para essa etapa da pesquisa. As indagações do pesquisador também devem ser lançadas aos elementos imagéticos (fotografias, grafismos e ícones, mixagens de textos e imagens) que “falam” sobre emoções ou atitudes. Os procedimentos envolvem descrição e interpretação no percurso entre teoria e análise, uma relação de ressignificação de nossos instrumentos teóricos (Orlandi, 2013, p. 5), num contínuo confronto com a opacidade do texto e dos sentidos.

Do ponto de vista da análise, as textualidades são a superfície mais acessível, mas o caminho a percorrer leva a toda a densidade que elas evocam. Orlandi (1995, p. 114) apregoa: “O que nos interessa é o que o texto organiza em sua discursividade, em relação à ordem da língua e a das coisas: a sua materialidade”. O que se busca é o sentido e este não se dá nem na estrutura nem no acontecimento, mas no encontro, na relação de ambos (Pêcheux, 1990). Dessa forma, os primeiros resultados obtidos com o auxílio de instrumentos informatizados precisam de posterior aprofundamento por parte do pesquisador, momento em que se cotejam os elementos empíricos encontrados com o auxílio das categorias teóricas e analíticas selecionadas no processo de pesquisa.

Considerações Finais

No âmbito da Comunicação, uma questão que se coloca é a compreensão da produção e dos sentidos e conexões constituídas pelos processos discursivos nesse universo ruidoso e diversificado que o atual ecossistema midiático produz. Os estudos de discurso são referências teórico-metodológicas que propõem um trabalho de compreensão e interpretação do mundo a partir desse componente que se manifesta por meio da linguagem, cuja materialidade se expressa em diferentes textualizações, sejam elas da ordem do verbal, visual ou pela mixagem de ambas, e que organizam os sentidos. As marcas do discurso que apontam para suas condições de produção são sempre deixadas nas superfícies discursivas.

O papel do pesquisador também se insere nesse princípio, pois é sua capacidade de leitura que vai produzir os resultados de um investimento interpretativo e de compreensão dos princípios e fundamentos que estruturam e dão suporte para o que se fala e para os falantes. O pesquisador, assujeitado à ordem do discurso, realiza, com seu trabalho de análise e leitura, um novo discurso na tentativa de reconectar as manifestações discursivas submetidas ao seu olhar e aos modelos matemáticos disponíveis nas ferramentas automatizadas aos seus contextos e à sua exterioridade, buscando os efeitos que causa nas dinâmicas sociais, políticas ou culturais. A investigação das textualidades realizadas no complexo e vasto sistema de comunicação em que estamos mergulhados hoje é uma chave importante para a compreensão dos conceitos e processos que o movimentam e, portanto, das políticas de comunicação postas em ação pelos diferentes actantes que interagem nos territórios informacionais das redes digitais, um ecossistema comunicacional em permanente transformação.

Notas

[1] Em março de 2024, o X/Twitter fechou o acesso gratuito à sua API, de modo que a coleta de dados dessa rede se tornou possível mediante pagamento ou de forma gratuita através de raspagem (*crawler*). Em agosto de 2024, a Meta suspendeu o Crowdtangle, ferramenta que controlava a coleta de dados públicos do Facebook e do Instagram. O serviço foi substituído pela Biblioteca de Conteúdo da Meta, mediado pelo Consórcio Interuniversitário para Pesquisa Política e Social (ICPSR) na Universidade de Michigan, mediante inscrição e aprovação. Outras redes como Mastodon e Bluesky permitem acesso às APIs.

[2] Sobre esse tema, ver documentos a respeito de ética na pesquisa que têm sido produzidos pela *Association of Internet Researchers* (AoIR).

[3] Ressalve-se que o *status* dessas ferramentas podem se alterar algumas podem até desaparecer como temos visto ao longo do tempo.

Artigo submetido em 13/03/2024 e aceito em 13/12/2024.

Referências

CHARAUDEAU, P. Dize-me qual é teu corpus, eu te direi qual é a tua problemática. **Revista Diadorim**. Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro. v. 10, p. 23, 2011. DOI: <https://doi.org/10.35520/diadorim.2011.v10noa3932>.

DIAS, C. Análise do discurso digital: sobre o arquivo e a constituição do corpus. **Estudos Linguísticos**, v. 44, n. 3, p. 972–980, 2016. Disponível em: <https://abrir.me/tVHNI>. Acesso em: 20 dez. 2024.

DIJCK, J. V.; POELL, T.; WAAL, M. **The platform society: public values in a connective world**. EUA: Oxford University Press, 2018.

DIJK, T. A. V. **Discurso e contexto**. Uma abordagem sociocognitiva. São Paulo: Contexto, 2012.

DONEDA, D.; ALMEIDA, V. A. F. O que é a governança de algoritmos? In: BRUNO, F.; CARDOSO, B.; KANASHIRO, M.; GUILHON, L.; MELGAÇO, L. **Tecnopolíticas da vigilância: perspectivas da margem**. São Paulo: Boitempo, 2018. p. 141–148.

FAIRCLOUGH, N. **Discurso e mudança social**. Brasília: Editora da UNB, 2001.

FRANZKE, A. S.; BECHMANN, A.; ZIMMER, M.; ESS, M. and the Association of Internet Researchers. **Internet Research: Ethical Guidelines 3.0**. [S.l.]: Association of Internet Researchers (AoIR). 2020. Disponível em: <https://aoir.org/reports/ethics3.pdf>. Acesso em: 20 dez. 2024.

FOUCAULT, M. **A arqueologia do saber**. Rio de Janeiro: Forense Universitária, 2008.

HAACKE, V. A. R.; LIMA, F. L. M. de. “Você fez, agora aguenta”: análise das narrativas de violência obstétrica no Facebook. **Revista Eco-Pós**, v. 23, n. 3, p. 210–238, 2020. DOI: <https://doi.org/10.29146/eco-pos.v23i3.27616>.

HELMOND, A. The platformization of the web: making web data platform ready. **Social Media + Society**, v. 1, n. 2, 2015.

JACOBSEN, D. R.; REIS, R. O discurso hater contra Greta Thunberg no Twitter. **Revista Iniciacom**. Revista Brasileira de Iniciação Científica, v. 11, n. 03, 2022. Disponível em: <https://abrir.me/uoTim>. Acesso em 20 dez. 2024.

KAISER, B. **Manipulados**: como a Cambridge Analytica e o Facebook invadiram a privacidade de milhões e botaram a democracia em xeque. Rio de Janeiro: Harper Collins, 2020.

LACLAU, E.; MOUFFE, C. **Hegemonia e estratégia socialista**. Brasília: Intermeios, 2015.

LATOUR, B. **Reagregando o social**: uma introdução à teoria ator-rede. Salvador: Edufba, 2012.

MEDEIROS, F. A. B.; SANTOS, J. M. O.; MOTA, H. C. N.; ANDRADE, I. G. M. O IRAMUTEQ como ferramenta no processamento de dados em pesquisa qualitativa. **Revista Diálogos em Saúde Pública**, v. 1, n. 2, 2022. Disponível em: <https://revistadiálogos.saude.rn.gov.br/index.php/EPS/article/view/26/17>. Acesso em 20 dez. 2024.

ORLANDI, E. P. A materialidade do gesto de interpretação e o discurso eletrônico. *In*: DIAS, C. (org.). **Formas de mobilidade no espaço e-urbano**: sentido e materialidade digital. Série e-urbano, vol. 2. Campinas: Labeurb, 2013. Disponível em: <https://www.labeurb.unicamp.br/livroEurbano/volumeII>. Acesso em: 20 dez. 2024.

ORLANDI, E. P. (Ed.). **Gestos de leitura**: da história no discurso. 4. ed. Campinas. Ed. Unicamp, 2014.

ORLANDI, E. P. **Texto e discurso**. Organon, Porto Alegre, vol. 9, n. 23, 2012. DOI: <https://doi.org/10.22456/2238-8915.29365>.

PARISER, E. **O filtro invisível** – o que a internet está escondendo de você. Rio de Janeiro: Zahar, 2012.

PAVEAU, M. A. **Análise do discurso digital**: dicionário das formas e das práticas. 2. ed. Campinas: Pontes Editores, 2021.

PAVEAU, M. A. Une linguistique symétrique pour penser le discours. **La pensée du discours**. [S.l.], 23 abr. 2010. DOI: <https://doi.org/10.58079/ssi4>.

- PÊCHEUX, M. **O discurso**: estrutura ou acontecimento. Campinas: Pontes, 1990.
- PÊCHEUX, M. **Análise de discurso**. 4. ed., vol. 1. Campinas: Pontes Editores, 2015.
- PÊCHEUX, M. Ler o arquivo hoje. *In*: ORLANDI, E. P. (org.). **Gestos de leitura**: da história no discurso. Campinas: Editora da Unicamp, 2014, p. 55-66.
- ROGERS, R. **Digital methods**. Cambridge: MIT Press, 2013.
- REIS, R.; ZANETTI, D.; FRIZZERA, L. A conveniência dos algoritmos: o papel do YouTube nas eleições brasileiras de 2018. **Compolítica**, v. 10, n. 1, p. 35-58, 2020. DOI: <https://doi.org/10.21878/compolitica.2020.10.1.333>.
- SILVA, F. S. R.; SILVA, T. (org.). **Artificial intelligence and racial discrimination in Brazil**: key issues and recommendations. Belo Horizonte: Institute for Research on Internet and Society, 2024. Disponível em: <https://bit.ly/4dGXxVi>. Acesso em: 20 dez. 2024.
- SOUZA, C. S. Leitura direta e leitura informatizada com o Voyant. **EMAPS-Notas #05**. Rio de Janeiro, RJ - Brasil: SERG, Departamento de Informática, PUC-Rio, 2024. 10p. Disponível em: <https://shre.ink/gnJh>. Acesso em: 20 dez. 2024.
- ZANELLA, A. Um gesto de leitura em Pêcheux: sobre máquinas, informática e arquivo. **Entremeios**: revista de Estudos do Discurso, Pouso Alegre, v. 14, 2017, 135-147.
- ZUBOFF, S. **A era do capitalismo de vigilância**. Rio de Janeiro: Intrínseca, 2021.