

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
Programa de Pós-Graduação em Estudos de Linguagens

Luana Teixeira de Souza Cruz

PLATAFORMIZAÇÃO DO TEXTO:
reconfiguração de práticas de escrita e edição a partir de mediações algorítmicas do *Google*

Belo Horizonte
2023

Luana Teixeira de Souza Cruz

PLATAFORMIZAÇÃO DO TEXTO:

reconfiguração de práticas de escrita e edição a partir de mediações algorítmicas do *Google*

Tese apresentada ao Programa de Pós-Graduação em Estudos de Linguagens do Centro Federal de Educação Tecnológica de Minas Gerais, como parte dos requisitos para obtenção do título de Doutora em Estudos de Linguagens.

Orientadora: Profa. Dra. Ana Elisa Ferreira Ribeiro

Área de concentração: Edição, Linguagem e Tecnologia

Belo Horizonte
2023

Cruz, Luana Teixeira de Souza.
C957p Plataformização do texto : reconfiguração de práticas de escrita e edição a partir de mediações algorítmicas do Google / Luana Teixeira de Souza Cruz. – 2023.
280 f. : il.
Orientadora: Ana Elisa Ferreira Ribeiro.

Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais, Programa de Pós-Graduação em Estudos de Linguagens, Belo Horizonte, 2023.
Bibliografia.

1. Edição (Editoração). 2. Formação algorítmica. 3. Visibilidade. 4. Motores de busca. 5. Plataformização - Textos. I. Ribeiro, Ana Elisa Ferreira. II. Título.

CDD: 070.57973



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM ESTUDOS DE LINGUAGENS

LUANA TEIXEIRA DE SOUZA CRUZ

PLATAFORMIZAÇÃO DO TEXTO: reconfiguração de práticas de escrita e edição a partir de mediações algorítmicas do Google

Tese apresentada ao Programa de Pós-Graduação em Estudos de Linguagens do Centro Federal de Educação Tecnológica de Minas Gerais em 25 de setembro de 2023, como requisito parcial para obtenção do título de Doutora em Estudos de Linguagens, aprovada pela Banca Examinadora constituída pelos professores:

Prof.ª Dr.ª Ana Elisa Ribeiro

Centro Federal de Educação Tecnológica de Minas Gerais

Prof.ª Dr.ª Carla Viana Coscarelli

Universidade Federal de Minas Gerais

Prof.ª Dr.ª Verônica Soares da Costa

Pontifícia Universidade Católica de Minas Gerais

Dr.ª Pollyanna Mattos Vecchio

Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Rogério Barbosa da Silva

Centro Federal de Educação Tecnológica de Minas Gerais

Dedico esta tese aos meus filhos, Martin e Heitor, que nasceram em pleno processo de pesquisa. Vocês são meu combustível.

AGRADECIMENTOS

PARTO da premissa que “[...] pesquisar é mais perguntar do que responder” (HISSA, 2012, n.p.); portanto, agradeço a todos os pesquisadores e pesquisadoras – citados ou não em minha pesquisa – que precederam o meu trabalho e me permitiram, humildemente, fazer perguntas. Nossas re(descobertas) se alimentam das dúvidas de todos cientistas que compõem um campo.

Expresso imensa gratidão à minha orientadora, Dra. Ana Elisa Ribeiro, a quem carinhosamente chamo de “diva”, porque ela é inspiração de carreira e de vida; é, para mim, uma divindade. É dela que eu me lembro, é para ela que eu pergunto, é nela que eu me espelho. Ela costuma dizer: “Nunca vi ninguém piorar estudando, fazendo mestrado e doutorado”. Eu digo: “Nunca vi ninguém piorar estando junto com a Ana Elisa, porque ela sabe soprar as asas de seus orientandos, indicando, propondo, direcionando, questionando, fazendo parcerias”.

Sou grata por integrar uma família que sempre valorizou a educação e o conhecimento. Obrigada a meus pais, Neide e Teixeira, e irmãos, Léo e Bárbara, que compartilham minhas jornadas de estudo desde sempre. Ao meu marido, Marcos, pelas trocas gerais sobre assuntos de tecnologia, pela parceria na criação dos nossos filhos, em especial, durante o processo de escrita da tese. Aos meus filhos, Martin e Heitor, que me mostraram a grandiosidade da maternidade gemelar e isso me fez costurar esta pesquisa de doutorado com mais afeto.

Agradeço aos meus amigos e amigas, que me fazem enxergar meu potencial como pesquisadora, jornalista, professora e mãe, principalmente no período anuviado da pandemia da covid-19. Em especial, Verônica Soares, que parece morar nos meus pensamentos, tamanha a sintonia de ideias. A amizade só melhora porque ela sempre traz perspectivas da vida e dos desafios com mais lucidez, coerência e inteligência. Muita gratidão também a Eliziane Silva, porque partilhamos escutas muito carinhosas – acadêmicas e não acadêmicas. Ela anda de bicicleta no meu coração. Gratidão a Pollyanna Vecchio, minha contemporânea de mestrado e doutorado, com quem tive o prazer de escrever um artigo. Ela é autora de uma das pesquisas mais potentes que li nos últimos anos. Nunca cansarei de citar o trabalho dessa incrível pesquisadora.

Minha gratidão ao Programa de Pós-Graduação em Estudos de Linguagens (Posling), na figura de todos coordenadores, funcionários, colegas discentes e professores com quem convivi de 2012 até aqui. Afinal, são 11 anos no Programa, distribuídos nas

jornadas de disciplinas isoladas, mestrado e doutorado. Gratidão também ao CEFET-MG, que me acolhe desde o Ensino Técnico. Pertencço a essa instituição com muito orgulho.

Obrigada ao André Silva e ao Vitor Soares, que contribuíram no fluxo editorial da minha tese, revisando, refinando, ajustando, diagramando e redesenhando a minha escrita. Também fizeram parte desse fluxo pessoas para as quais pedi dicas e troquei figurinhas: Lorena Tárzia, Letícia Santana, Vanessa Fagundes, Patrick Simon, Caio César, Wagner Moreira, Paula Renata, Andrea Santos.

Sou grata à minha terapeuta, Cristiana Amorim, por me ensinar a reencontrar minhas essências e acreditar que tenho uma linda trajetória já construída e em construção. Nossos encontros foram fundamentais no último ano do doutorado.

Aos membros da banca minha gratidão pela leitura carinhosa da pesquisa em mais de uma etapa. Obrigada ao professor Dr. Carlos d'Andréa, que contribuiu com grande cuidado e delicadeza no parecer do projeto (aliás, do mestrado e doutorado); aos professores Dra. Carla Coscarelli, Dra. Úrsula Anacleto e Dr. Vicente Parreiras, que expandiram minha mente a partir dos comentários na qualificação.

Por fim, a expressão “só Deus sabe” se encaixa perfeitamente. Por isso, sou grata a Ele.

RESUMO

A produção, circulação e leitura de textos em ambiente digital estão atravessadas pelas plataformas *online*. Por isso, vivemos uma cultura digital pós-plataformas cada vez mais implicada pela comunicação mediada por algoritmos e pela tradução de tudo em dados numéricos. Esta pesquisa investiga a reconfiguração das práticas de escrita e edição a partir das lógicas algorítmicas do *Google*, que se firmou como uma plataforma infraestrutural capaz de ordenar, filtrar e selecionar a produção imaterial *online*. A incorporação de técnicas em resposta às demandas de ranqueamento do *Google* é comum e praticada de modo ambivalente por produtores e editores de texto. A ambivalência está no fato de que esses produtores se sentem, por um lado, na condição de ter que equilibrar seu trabalho entre a relevância para públicos e para algoritmos; por outro, muitos deles criam táticas de produção que parecem vencer a caixa-preta do *Google* em nome da performance de conteúdos. Esta pesquisa tem como suporte teórico a Linguística Textual, com referências que compreendem o texto não com base em sua essência, mas, sim, diante de suas condições de produção, circulação e leitura, ou seja, a partir da sua existência. Como inspirações metodológicas, temos os Estudos de Plataformas, Estudos de Algoritmos e Métodos Digitais, que têm matrizes conceituais originárias dos Estudos de Ciência e Tecnologia. Recorremos também aos Estudos de Busca na *Web*, que, por sua vez, alimentam-se de ideias do campo de Recuperação da Informação. No percurso de pesquisa, integramos duas etapas de métodos qualitativos: 1) análises documentais da plataforma de busca *Google*, para compreensão de funcionamento e governança; 2) experiências de escrita e edição de textos, em um espaço simulado no *WordPress.com*, para observar as práticas incorporadas por produtores e editores. O *corpus* da etapa 1 é composto por documentos que têm ênfase jurídica e tentam proteger o *Google* de atitudes danosas praticadas por seus usuários. Também há documentos que dizem respeito às expectativas da plataforma sobre o que é apropriado ou não durante o uso do motor de busca e, ainda, sobre o que é ideal para alcançar melhores performances de circulação de conteúdo. O *corpus* da etapa 2 é composto por cinco textos que representam diferentes modos de produção, objetivos, públicos e gêneros discursivos. Na análise, emergiram cinco categorias que nos permitem entrelaçar a elegibilidade algorítmica do *Google* com as práticas de escrita em ambiente digital. Concluímos que a plataformização do texto é o fenômeno composto pelo atravessamento de critérios algorítmicos e normatizações, definidas pelas plataformas de busca, na escrita e edição. Demonstramos que o *Google* mudou nosso jeito de escrever, porque consolidou como quase inevitável seu modelo de indexação e classificação, que

tensiona produtores de texto a redigir pensando em uma relevância mitológica. Esse tensionamento faz surgir o texto plataformizado, de caráter imanente e contingente, que se constrói de modo hipertextual, pela natureza do meio e dos processos de leitura, e em atenção ao desempenho para a visibilidade.

Palavras-Chave: Mediações algorítmicas. Edição. Visibilidade. Texto plataformizado. Plataformas de busca.

ABSTRACT

The production, circulation and reading of texts in a digital environment are crossed by online platforms. Therefore, we live in a post-platform digital culture, increasingly implicated in communication mediated by algorithms and the translation of everything into data. This research investigates the reconfiguration of writing and editing practices based on Google's algorithmic logic, which has established itself as an infrastructural platform capable of ordering, filtering and selecting immaterial production online. The incorporation of techniques in response to Google's ranking is common and practiced ambivalently by text producers and editors. The ambivalence lies in the fact that these producers feel, on the one hand, that they have to balance their work between relevance for audiences and for algorithms, on the other hand, many of them create production tactics that seem to beat Google's blackbox in the name of content performance. This research is theoretically supported by Text Linguistics, with references that understand the text not based on its essence, but rather on its production, circulation and reading conditions, that is, from its existence. As methodological inspirations, we have Platform Studies, Critical Algorithm Studies and Digital Methods, which have conceptual matrices originating from Science and Technology Studies. We are also inspired by Web Search Studies, which in turn feed on ideas from Information Retrieval. In the course of the research, we integrated two phases of qualitative methods: 1) document analysis of Google, to understand its operation and governance; 2) experiences of writing and editing texts, in a simulated space on *Wordpress.com*, to observe the practices incorporated by producers and editors. The corpus of phase 1 is composed of documents that have a legal emphasis and try to protect Google from harmful attitudes practiced by its users. There are also documents regarding the platform's expectations about what is appropriate or not when using the search engine, and also about what is ideal to achieve better content circulation performances. The corpus of phase 2 is composed of five texts that represent different modes of production, objectives, audiences and discursive genres. In the analysis, five categories emerged that allow us to connect Google's algorithmic eligibility with writing practices in a digital environment. We conclude that the platformization of the text is the phenomenon composed by the crossing of algorithmic criteria and norms, defined by the search engine, in writing and editing. We demonstrate that Google has changed our way of writing, because it has consolidated its indexing and classification model as almost inevitable, which forces text producers to write with a mythological relevance in mind. This tensioning gives rise to the platformed text, of an

immanent and contingent character, which is built in a hypertextual way due to the nature of the medium and the reading processes, but also in attention to the performance for visibility.

Keywords: Algorithmic mediations. Editing. Visibility. platformed text. Search engine.

LISTA DE SIGLAS E ABREVIACÕES

Aleam	Assembleia Legislativa do Estado do Amazonas
ALMG	Assembleia Legislativa de Minas Gerais
Anatel	Agência Nacional de Telecomunicações
API	Interface de Programação de Aplicativos
Apps	Aplicativos
Batx	<i>Baidu, Alibaba, Tencent, Xiaomi</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
Cade	Conselho Administrativo de Defesa Econômica
CEFET-MG	Centro Federal de Educação Tecnológica de Minas Gerais
CGI	Comitê Gestor da Internet no Brasil
Cide	Contribuição de Intervenção no Domínio Econômico
CLS	<i>Cumulative Layout Shift</i>
CMS	Sistemas de Gerenciamento de Conteúdo
CNBC	<i>Consumer News and Business Channel</i>
CSS	Cascading Style Sheet
CTB	Código Brasileiro de Trânsito
DMA	<i>Digital Market Act</i>
DSA	<i>Digital Services Act</i>
EEAT	Experiência, Especialidade, Autoridade e Confiabilidade
EUA	Estados Unidos da América
Fapesp	Fundação de Amparo à Pesquisa do Estado de São Paulo.
Fenaj	Federação Nacional dos Jornalistas
FID	<i>First Input Delay</i>
Funajor	Fundo Nacional de Apoio e Fomento ao Jornalismo
Gafam	<i>Google Alphabet Inc., Apple, Facebook (Meta), Amazon e Microsoft</i>
GPS	Sistema de posicionamento global
HTML	<i>HyperText Markup Language</i>
IA	Inteligência Artificial
INP	<i>Net PaiNt</i>
IP	Protocolo da Internet
IR	Information Retrieval
KDP	<i>Kindle Direct Publishing</i>

LaMDA	Modelo de linguagem neural conversacional
LCP	<i>Largest Contentful Paint</i>
LGPD	Lei Geral de Proteção de Dados
LLM	<i>Large Language Model</i>
MPF	Ministério Público Federal
Natu	<i>Netflix, Airbnb, Tesla, Uber</i>
NLP	Processamento de linguagem natural
NQ	<i>Natural Questions</i>
PF	Polícia Federal
PL	Projeto de Lei
Posling	Programa de Pós-Graduação em Estudos de Linguagens
PR (Digital)	Relações Públicas Digital
QA	Sistema de perguntas e respostas
SEO	<i>Search Engine Optimization</i>
SES-MG	Secretaria de Estado de Saúde de Minas Gerais
SGE	<i>Search Generative Experience</i>
SERP	<i>Search Engine Results Page</i>
STS	<i>Science and Technology Studies</i>
OCR	Reconhecimento óptico de caracteres automatizado
TAR	Teoria Ator-Rede
Unisinos	Universidade do Vale do Rio dos Sinos
URL	<i>Uniform Resource Locator</i>
YMYL	<i>Your Money Your Life</i>

LISTA DE ILUSTRAÇÕES

Figura 1 – Camadas do processo de edição de texto em ambiente digital no contexto da plataformização	30
Figura 2 – Movimentos atuais da cultura escrita.....	47
Figura 3 – Perfil do projeto <i>Insta Repeat</i> no <i>Instagram</i>	51
Figura 4 – Projeto <i>Viral Post Generator</i>	52
Figura 5 – Página inicial do <i>Google</i> em 1 de maio de 2023.....	69
Figura 6 – Desenho metodológico.....	101
Figura 7 – Produtos do <i>Google</i>	112
Figura 8 – Confiança em notícias em plataformas x confiança em notícias gerais.....	116
Figura 9 – O que está por trás da barra do <i>Google</i> ?	122
Figura 10 – Receita de publicidade das principais empresas de venda de anúncios digitais em todo o mundo, em 2022. (em bilhões de dólares americanos)	124
Figura 11 – Como funciona a busca no <i>Google</i>	135
Figura 12 – Composição da SERP	140
Figura 13 – Conteúdo principal e suplementar.....	146
Figura 14 – Sinais vitais considerados pelo <i>Google</i> para metrificar boa experiência de uso	149
Figura 15 – Escaneabilidade.....	151
Figura 16 – Readability	153
Figura 17 – Hierarquização <i>heading 1</i> , <i>heading 2</i> , <i>heading 3</i>	154
Figura 18 – Termos mais buscados no <i>Google</i> em 2022 – Brasil	175
Figura 19 – <i>Rich snippet</i>	178
Figura 20 – <i>Graph search</i>	179
Figura 21 – <i>Knowledge Graph</i>	179
Figura 22 – <i>Google Business Profile</i>	180
Figura 23 – <i>Setup</i> https://producaoegestaodeconteudo.com/ no <i>WordPress.com</i>	197
Figura 24 – <i>Yoast SEO</i>	199
Figura 25 – <i>Rank Math</i>	200
Figura 26 – Pontuação no <i>Rank Math</i>	201
Figura 27 – Código-Fonte indicando uso do <i>WordPress.com</i>	202
Figura 28 – Código-Fonte indicando uso do <i>Yoast SEO</i>	203
Figura 29 – Código fonte indicando uso do <i>Rank Math</i>	204
Figura 30 – Análise de frase-chave no texto 1 pelo <i>Yoast SEO</i>	208

Figura 31 – <i>Snippet</i> da SERP	211
Figura 32 – Análise do título, metadescrição e URL do <i>Yoast SEO</i> / texto 1	214
Figura 33 – Análise do título, metadescrição e URL do <i>Rank Math</i> / texto 1.....	215
Figura 34 – <i>Snippet</i> da SERP/ texto 4	216
Figura 35 – Análise do título, metadescrição e URL do <i>Yoast SEO</i> / texto 4.....	216
Figura 36 – <i>Snippet</i> da SERP/ texto 5	217
Figura 37 – Análise do título, metadescrição e URL do <i>Yoast SEO</i> / texto 5.....	217
Figura 38 – Exemplo 1 de texto-âncora.....	222
Figura 39 - Exemplo 2 de texto-âncora	222
Figura 40 – Análise de legibilidade do <i>Yoast SEO</i> para o texto 1.....	224
Figura 41 – Análise de legibilidade do <i>Yoast SEO</i> para o texto 5.....	225
Figura 42 – Demarcação de necessidade de edição feita pelo <i>Yoast SEO</i>	227
Figura 43 – Intertítulos texto 2	236
Figura 44 – Código-Fonte atributo <i>alt-tag</i> / texto 5	238
Figura 45 – Imagem do texto 5.....	239
Figura 46 – SEO marca de bebida vegetal <i>Nude</i>	243
Quadro 1 – Elementos da cadeia de valor da informação e do conteúdo	72
Quadro 2 – Articulação de elementos da cadeia de valor da informação e do conteúdo	73
Quadro 3 – Posição das <i>Big Five</i> no mercado mundial.....	87
Quadro 4 – Concorrência entre empresas de tecnologia	87
Quadro 5 – Abordagens metodológicas para Estudos de Algoritmos.....	93
Quadro 6 – Abordagens metodológicas para Estudos de Busca na <i>Web</i>	98
Quadro 1 – <i>Corpus</i> etapa 1.....	106
Quadro 8 – Tipos de intenção de pesquisa	172
Quadro 9 – Movimentos do leitor vistos nas ferramentas de <i>analytics</i>	184
Quadro 10 – <i>Corpus</i> etapa 2	202
Quadro 11 – Categorias de análise	205
Quadro 12 – Entrelaçamento entre plataforma de busca e escrita na <i>Web</i> (categoria 1 – palavra-chave)	210
Quadro 13 – Titulação: relevância para leitores <i>versus</i> relevância para algoritmos	213
Quadro 14 – Entrelaçamento entre plataforma de busca e escrita na <i>Web</i> (categoria 2 – título, metadescrição e URL)	218

Quadro 15 – Entrelaçamento entre plataforma de busca e escrita na <i>Web</i> (categoria 3 – hipertextualidade).....	223
Quadro 16 – Inserção de palavras de transição	229
Quadro 17 – Inserção de palavras de transição	232
Quadro 18 – Distribuição de cabeçalhos texto 1	235
Quadro 19 – Distribuição de cabeçalhos texto 2	236
Quadro 20 – Entrelaçamento entre plataforma de busca e escrita na <i>Web</i> (categoria 4 – legibilidade e escaneabilidade).....	237
Quadro 21 – Entrelaçamento entre plataforma de busca e escrita na <i>Web</i> (categoria/ imagem)	240
Quadro 22 – Critérios de plataformas de busca x lógicas de produção textual.....	249
Tabela 1 – Tamanho do texto	226

SUMÁRIO

1 INTRODUÇÃO	19
1.1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA.....	24
PARTE I: SENTIDOS	37
2 PRODUÇÃO, CIRCULAÇÃO E LEITURA DE TEXTO EM AMBIENTE DIGITAL	38
2.1 TEXTO PLATAFORMIZADO	43
2.2 CULTURA DIGITAL PÓS-PLATAFORMAS	61
2.3 CADEIA DE VALOR DO TEXTO PLATAFORMIZADO	71
2.3.1 Seleção	73
2.3.2 Acesso, facilidade e preços	76
2.3.3 Agregação, intertextualidade e flexibilidade.....	77
2.3.4 Navegação, escala e possibilidade de pesquisa.....	78
2.3.5 Desenvolvimento, capacidade de atualização, portabilidade e multimídia.....	80
2.3.6 Autoridade.....	80
2.3.7 Desdobramentos de uma nova cadeia de valor	81
3 METODOLOGIAS	85
3.1 INSPIRAÇÕES METODOLÓGICAS	85
3.1.1 Estudos de Plataformas e Estudos de Algoritmos	86
3.1.2 Métodos digitais	95
3.1.3 Estudos de Busca na <i>Web</i>	96
3.2 MÉTODOS DESTA PESQUISA.....	99
PARTE II: MATERIALIDADES E EXPERIÊNCIAS	100
4 A PLATAFORMA DE BUSCA <i>GOOGLE</i>	105
4.1 ANÁLISE DOCUMENTAL	106
4.1.1 Seleção do <i>corpus</i>	106
4.2 <i>DON'T BE EVIL</i> : A ONIPOTÊNCIA DO <i>GOOGLE</i>	107
4.3 LÓGICAS DOS MOTORES DE BUSCA E GOVERNANÇA DO <i>GOOGLE</i>	117
4.3.1 Políticas.....	118
4.3.2 Funcionamento	128

4.3.2.1 Rastreamento	136
4.3.2.2 Indexação.....	138
4.3.2.3 Ranqueamento	139
4.4 COMO O <i>GOOGLE</i> CONSTRÓI “MÉTRICAS” DE RELEVÂNCIA	145
4.4.1 Páginas e conteúdos que reflitam o objetivo principal.....	148
4.4.2 Boa experiência de uso	149
4.4.3 Utilidade e originalidade	156
4.4.4 Atualizações e frequência de publicações.....	157
4.4.5 <i>Links</i> rastreáveis, títulos descritivos e úteis	158
4.4.6 Transparência e autoridade.....	159
4.4.7 Qualidade e quantidade de conteúdo.....	161
4.4.8 Fatores EEAT	162
4.4.9 Mito da relevância	166
4.5 MEDIAÇÃO E CURADORIA	167
4.6 MOVIMENTOS DE BUSCA	170
4.6.1 Intenções de busca	171
4.6.2 Leitura-Escritura.....	174
4.6.3 Leitura-Navegação	176
4.6.4 Interação pós-SERP	182
4.7 DESCOMPACTANDO O CONJUNTO SOCIOTÉCNICO DE ALGORITMOS.....	187
4.7.1 BERT	189
5 PLATAFORMIZAÇÃO DO TEXTO: PRÁTICAS DE ESCRITA E EDIÇÃO INCORPORADAS	193
5.1 ESPAÇO SIMULADO DE ESCRITA E EDIÇÃO	196
5.1.1 <i>Yoast SEO</i>	198
5.1.2 <i>Rank Math</i>	199
5.1.3 Seleção do <i>corpus</i>	201
5.1.4 Categorias de análise	204
5.2 RECONFIGURAÇÃO DE PRÁTICAS: ANÁLISES E DISCUSSÕES.....	206
5.2.1 Palavra-Chave.....	206
5.2.2 Título, meta descrição e URL	210
5.2.3 Hipertextualidade (<i>links</i>)	218
5.2.4 Legibilidade e escaneabilidade	223

5.2.5 Tamanho do texto	226
5.2.5.1 Tamanho do parágrafo.....	227
5.2.5.2 Frases consecutivas.....	228
5.2.5.3 Palavras de transição	228
5.2.5.4 Tamanho das sentenças	230
5.2.5.5 Voz passiva.....	231
5.2.5.6 Hierarquia do texto – distribuição de intertítulos ou cabeçalhos (<i>headings</i>)	233
5.2.6 Imagem	237
5.3 IMPLICAÇÕES DA RECONFIGURAÇÃO DAS PRÁTICAS DE ESCRITA E EDIÇÃO	241
5.3.1 Incorporação generalizada de técnicas de SEO	242
5.3.2 Homogeneização e surgimento de gêneros textuais	244
5.4 PRÁTICAS SUBVERSIVAS: INDÍCIOS E CAMINHOS PARA RESISTÊNCIAS À CURADORIA ALGORÍTMICA	250
6 CONSIDERAÇÕES FINAIS	260
6.1 TEXTO PLATAFORMIZADO	263
6.2 CULTURA DIGITAL PÓS-PLATAFORMAS	264
6.3 MITO DA RELEVÂNCIA.....	265
6.4 PLATAFORMIZAÇÃO DO TEXTO	266
REFERÊNCIAS	268

Minha amiga Ana Elisa,

Certa vez, durante uma discussão sobre a influência dos algoritmos na circulação de informação, em uma turma de pós-graduação em Comunicação, um aluno pediu a palavra e se posicionou assim: “Não vejo problema nenhum que as plataformas online me recomendem conteúdo. Fico super à vontade de que elas escolham e me digam o que consumir”. Compreendi que ele tem a noção da influência algorítmica, mas não entende as implicações sociais, econômicas e culturais disso em nível macro. Argumentei que a discussão não pode se limitar ao gosto individual ou à perspectiva privada, mas, sim, a uma escala ampla de circulação de informações públicas que têm implicações sociais.

Dei como exemplo o momento em que plataformas mediam a circulação massiva de desinformação durante a pandemia da covid-19 e isso ajuda na construção de narrativas antivacina que, por sua vez, atrapalham e interferem em políticas de saúde pública nacionais e internacionais. Se uma tecnologia influencia diretamente na construção de uma política pública de saúde por causa da mediação da informação, chegamos a um resultado coletivo que precisa ser debatido. E esse debate deve estar acima de qualquer visão neoliberal que permita empresas privadas escolherem filmes para um indivíduo se entreter no fim de semana ou músicas para compor a playlist da academia.

O aluno se deixou enredar pelo determinismo tecnológico de que plataformas online estão aí para servir a personalização e ao gosto do consumidor, centrando seus negócios em relevância para esses públicos, quando, na verdade, o interesse desses grandes negócios é outro, enraizado nos métodos de propaganda direcionada. A conversa me fez lembrar a ideia de barganha faustiana do educador e crítico social Neil Postman (2013, p. 2): “[...] a tecnologia dá e a tecnologia tira, e nem sempre na mesma medida. Uma nova tecnologia às vezes cria mais do que destrói. Às vezes, destrói mais do que cria. Mas nunca é unilateral”.

Obrigada pela escuta, sempre.

Abraços,

Luana

1 INTRODUÇÃO

A ordenação algorítmica da informação, a tipologia mais densa de todos os modos de ordenação, tornou-se uma prática onipresente em nossa experiência *online*. Por isso, é necessário e imediato entender como a capacidade de organizar, filtrar, classificar ou recomendar textos, imagens, opiniões, indivíduos, populações e tudo o mais – de maneira altamente dinâmica e orientada a objetivos privados – se relaciona com as formas contemporâneas de capitalismo.

As dinâmicas de “produção, circulação e leitura” de texto em ambiente digital, mediadas por algoritmos de plataformas *online*, subverteram as lógicas tradicionais, praticadas antes por grandes conglomerados de mídias e por produtores profissionais. A viabilidade e a existência lucrativa dos negócios ancorados na distribuição de informação estão alteradas pela ascensão de grandes provedores, mais especificamente o grupo que compõe as *Big Five* ou *Gafam* (*Google Alphabet Inc., Apple, Facebook, Amazon e Microsoft*) ou, ainda, *Big Techs*¹, que ancoraram as lógicas de circulação de texto no engajamento das pessoas e passaram a protagonizar o regime capitalista contemporâneo. Aliás, Jin (2013, p. 167 *apud* POELL; NIEBORG; VAN DIJCK, 2020, p. 5), considera a existência de um “imperialismo de plataforma”, argumentando que o rápido crescimento dessas empresas demonstra que o imperialismo americano continuou por meio da exploração de plataformas. Apesar de já buscarmos uma construção subversiva e resistente, em que algumas iniciativas se alinham ao “socialismo de plataforma” (MULDOON, 2022), tentando reinventar a internet e reimaginar usos para as plataformas *online*, ainda há um longo caminho.

Essas empresas privadas são capazes de ditar o funcionamento de vários mercados, como do jornalismo, da música, do livro (D’ANDRÉA, 2020), influenciando, especialmente, a indústria cultural. Destacamos o atravessamento das lógicas privadas no trabalho cultural (imaterial) tanto dos produtores, quanto dos consumidores dos conteúdos. Por meio de suas plataformas *online* (*Twitter, Instagram, Facebook, Google Search, e-commerce Amazon, YouTube* etc.), os provedores de informação atuam fortemente para reorganizar as relações interpessoais, o consumo de bens culturais, as discussões políticas e as práticas urbanas. Como donos de sistemas computacionais que suportam o trabalho criativo, esses provedores definem não somente quais textos devem circular, mas redesenham práticas

¹ Esses termos são encontrados de formas diversas em pesquisas da área. As combinações tornam-se infinitas, caso venhamos a destacar a hegemonia de empresas em setores da indústria (cultural e outras). Mais recentemente, criaram o Natu (*Netflix, Airbnb, Tesla e Uber*) e também surgiu o Batx (*Baidu, Alibaba, Tencent e Xiaomi*) para se referir aos gigantes chineses.

e percepções coletivas. Mesmo que futuramente venham a ser outras empresas, outras plataformas e outras lógicas programáticas (como Inteligência Artificial (IA)) a operar a ordenação das informações, o trabalho cultural continuará sob as implicações dessas tecnologias.

É compreensível que ferramentas e modos de fazer texto sejam reconfigurados para atender as necessidades produtivas, principalmente para produtores de texto que se adaptam contingencialmente às lógicas capitalistas de audiência. Aliás, compreendemos o texto em perspectiva semiótica expandida, como definiremos no Capítulo 2, e com caráter multimodal (KRESS; VAN LEEUWEN, 1998); portanto, estamos nos referindo a representações de sentido verbais, verbovisuais, infográficas, fotográficas e outras modulações possíveis. É preciso olhar criticamente para a quantidade de camadas que são, atualmente, criadas na circulação textual: há algoritmos definidores de parâmetros de relevância que motivam o uso de ferramentas (*plugins* e outros) que, por sua vez, ensinam ao humano produtor/editor como alcançar a parametrização e melhorar a performance de circulação.

Vale demarcar que entendemos algoritmos como um complexo conjunto sociotécnico que faz parte de uma família de sistemas autoritários para produção de conhecimento ou tomada de decisão (GILLESPIE, 2016). Algoritmos são regras criadas para que sistemas executem ações de busca, seleção, organização, classificação, recomendação, entre outras. Nesses sistemas, os leitores/usuários/consumidores envolvidos são legíveis como dados e colocados em relações matemáticas entre si e a informação que consomem. As relações que se estabelecem parecem ser mensuráveis, mas as métricas postuladas nesse conjunto sociotécnico, quase sempre, não conseguem dar conta da complexa atividade social, e os valores nela contidos são traduzidos em uma interação funcional de variáveis, indicadores e resultados. Usaremos diversas vezes a expressão “mediações algorítmicas”, compreendida aqui como dispositivo que permite interações entre espaços de produção e leitura com aspectos técnicos extremamente mutantes (WINQUES; LONGHI, 2022). É por causa dessa intensa mediação que os funcionamentos dos algoritmos precisam ser decifrados.

Destaca-se nesta tese um fluxo de produção e edição sem precedentes na história da escrita. Tal modelo implica a vinculação dos produtores a uma agenda de informações e de visões de mundo que é definida pela lógica das plataformas *online*, em especial, os motores de busca, como *Google* e seus concorrentes. As interferências dessa ferramenta no trabalho de edição entrelaçam modos humanos de pensar e organizar textos com hierarquias e lógicas de *softwares*. Os produtores tendem a naturalizar (ou não) a existência de ferramentas em suas

rotinas, além de acomodar em seus processos a crescente influência de tecnologias. São, na verdade, os provedores de informação (*Big Five*) que fazem parecer tão natural a plataformização da produção textual e a parametrização dos conteúdos aos “olhos” dos algoritmos. Talvez, por isso, soe tão distópico pensar em um cenário de circulação desses conteúdos em ambiente digital sem plataformas.

Dissertamos sobre distopia, mas tentando evitar uma leitura fatalista do presente e do futuro, porque isso poderia nos colocar na contramão de uma visada que pretenda superar determinismos tecnológicos (D’ANDRÉA, 2020); portanto, vamos considerar as diversidades de usos feitas por quem cria texto em ambiente digital. A própria tecnologia, em sua ambivalência, permite usos diversos e ressignificações, conforme argumenta Cupani (2011). O autor sustenta que há jogos de determinadas classes e grupos para manutenção ou derrubada de projetos políticos e sociais. Assim, quando pensarmos em circulação de conteúdos mediados por plataformas *online*, vamos esbarrar em iniciativas de ciberativismo que subvertem a posse de criação de ideias e conhecimento das grandes corporações, mas também seremos confrontados com a ação de empresas que usam estrategicamente as mediações algorítmicas, como, por exemplo, as buscas do *Google*, para lavar a reputação e apagar o passado criminoso de alguns clientes (já condenados por vários crimes). Em 17 de fevereiro de 2023, uma reportagem da *Folha de S.Paulo* trouxe à tona um desses esquemas de criação de *fake news*, manipulação de políticas de moderação de conteúdo e aplicações escusas de técnicas de otimização feitas por uma multinacional para limpar a memória digital de informações sobre traficantes, corruptos e torturadores.

É por isso que, na tentativa de dizer algo relevante sobre como os algoritmos transformam o discurso público, devemos resistir firmemente à tentação de tratar a “[...] tecnologia como condutora nas explicações” (CUPANI, 2011, p. 98). Há ambivalências e, mesmo que produtores de texto ignorem a caixa-preta que manejam, eles não necessariamente se contentam com a programação prévia das plataformas *online* e fazem usos nada ingênuos das materialidades. Sendo assim, a questão que guia esta pesquisa é: em que medida produtores de texto, visando alcançar visibilidade no ambiente digital, incorporam – ambivalentemente – práticas e técnicas que respondem às lógicas de mediação algorítmica criadas por plataformas de busca? Nosso gesto metodológico foi de simular processo de escrita e edição em um Sistema de Gerenciamento de Conteúdo (CMS) para olharmos, especialmente, para a influência que o *Google* tem na reconfiguração das práticas e na conformação das cadeias editoriais. É nesse sentido que nos encontramos com outros

trabalhos da linha de pesquisa Edição, Linguagem e Tecnologia do Posling que estudam as relações entre linguagens, processos de criação, edição e circulação.

O objetivo principal desta pesquisa foi verificar a reconfiguração de práticas de escrita e edição, analisando as implicações da inserção da lógica das plataformas de busca na produção e circulação de texto em ambientes digitais, em especial, a incorporação de parametrizações estimuladas pelas demandas algorítmicas do *Google*. Para isso, nos dedicamos, especificamente, a 1) caracterizar o *Google* como plataforma de busca, compreendendo com isso o processo de plataformização do texto, e 2) discutir as consequências da plataformização do texto e da reconfiguração de práticas de escrita e edição, tanto nas rotinas de produção, quanto nos tipos de texto que circulam socialmente.

Esses passos foram cumpridos com duas etapas metodológicas de abordagem qualitativa. Na etapa 1, olhamos para os modos como a plataforma de busca *Google* estabelece regras e negocia condutas com os usuários, ou seja, a forma como estabelece governança. Para isso, investigamos e analisamos patentes, guias de produção e desenvolvimento, relatórios de transparência, entre outros documentos, fazendo um trabalho de leitura e anotação à luz dos fundamentos da nossa pesquisa. Na etapa 2, como já citado, criamos um espaço simulado para verificar a incorporação de práticas de escrita e edição, submetendo cinco textos já publicados na *Web* a dois *plugins* do *WordPress.com*, o *Yoast SEO* e o *Rank Math*. Nossa intenção foi testar e parametrizar essas produções textuais, editando-as para que pudéssemos – simuladamente – fazer esses conteúdos serem mais elegíveis aos algoritmos da plataforma de busca *Google*. O método consistiu em inserir o texto escolhido no rascunho de um *post* de *WordPress.com*, dentro do *site* <https://producaoegestaodeconteudo.com/>, criado exclusivamente para atividades de ensino, aprendizagem e pesquisa. Ao submeter os textos às análises oferecidas pelos *plugins*, conseguimos detalhar as fases de edição desses textos para verificar o quanto é possível alterar uma produção para que ela seja mais ranqueável, já que os *plugins* têm a função de automatizar esse serviço e propiciar uma escrita mais encontrável.

Durante essas duas etapas debatemos, então, a reconfiguração de práticas de escrita e edição na medida em que as plataformas de busca definem o desempenho da circulação de uma produção textual a partir de certos parâmetros, como a quantidade de caracteres ou modelos de títulos que performam melhor para a entrega a públicos leitores; a descrição detalhada de imagens e fotos; a hierarquização do texto redigida por técnicas de otimização; entre outros. Assim, investigamos a homogeneização e o sugrimento de gêneros textuais a partir das demandas algorítmicas, especialmente, das plataformas de busca.

Consideramos o termo “desempenho” à luz dos estudos sobre performance em plataformas *online* desenvolvidos por pesquisadores da área de comunicação, em especial, com recortes para o jornalismo digital. Muitos deles estão reunidos no livro *Performance em ciberjornalismo: tecnologia, inovação e eficiência*, de 2017, que trata de vários mecanismos de determinação de desempenho para textos jornalísticos em ambiente digital. Em *The culture of connectivity: a critical history of social media*, José Van Dijck também lida com a noção de audiência digital, analisando a medição de desempenho sob a lógica capitalista de quanto mais eficiente uma ação se revela, mais é reproduzida.

Refletimos ainda a pressão por frequência de publicações, qualidade e reputação baseadas em parâmetros de Experiência, Especialidade, Autoridade e Confiabilidade (a sigla EEAT, em inglês)², além de outros mecanismos técnicos que atravessam a escrita para que textos sejam encontráveis aos “olhos” do *Google*, principalmente após a mudança algorítmica dessa plataforma em 2019. Olhamos criticamente para as reconfigurações da escrita de modo contingencial aos algoritmos que parecem ser necessárias para os produtores e editores com pretensões de alcançar visibilidade. Sendo assim, a coerência com a linha editorial e o senso criativo precisam também responder as expectativas de performance em plataformas *online*. As dimensões editoriais e um profundo conhecimento sobre os públicos não são mais suficientes; afinal, o caminho para alcançar os leitores está mediado por algoritmos.

Tudo isso parece impactante? A nós parece muito intrigante, a ponto de despertar a necessidade de destrinchar o quanto as plataformas de busca atravessam a “produção, circulação e leitura” textual em ambiente digital. Essa dinâmica parece definir gerações de textos que funcionam como molas sociais e espelhos culturais do tempo em que foram produzidos e publicados. “Os textos são mais bem entendidos não como artefatos fixos e determinados num meio específico, mas como potenciais. Todas as versões implicam uma forma ideal que nunca é totalmente realizada, apenas parcialmente percebida e expressa por qualquer uma delas.” (MCKENZIE, 2018, p. 73).

Na lógica das plataformas de busca, o que se produz, em qualquer fluxo editorial – dos mais organizados e padronizados aos mais indefinidos e bagunçados –, é uma eterna versão beta. Esse texto em ambiente digital precisa se conformar com demandas de audiência extremamente mutantes; portanto, ele também se faz muito mutante, conforme discutiremos no Capítulo 2 desta pesquisa.

² Parâmetros inseridos pelo *Google*, em 2019, para elegibilidade de conteúdos para indexação e ranqueamento.

1.1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA

Apesar de o termo “plataforma” ter se tornado uma *buzzword*³, assim como, em outras épocas, expressões como interatividade, hipertexto, cibercultura, convergência e redes, é preciso localizar essa definição. Partimos da noção de plataformas *online* como infraestruturas digitais robustas e (re)programáveis que moldam atividades sociais em arquiteturas computacionais, provendo *softwares* e *hardwares* para usuários finais. Esses usuários, por sua vez, interagem com apropriações criativas e performativas, geram coletivismos (VECCHIO, 2022) e cooperativismos (SCHOLZ, 2016)⁴ e recriam, cotidianamente, as dinâmicas *online*. As plataformas são organizadas por meio de coletas sistemáticas, processamentos algorítmicos, monetização e circulação de dados, conforme resumem os autores do livro *The platform society: public values in a connective world*: “Uma plataforma é alimentada com dados, automatizada e organizada por meio de algoritmos e interfaces, formalizada por meio de relações de propriedade orientadas pelos seus modelos de negócios e regidas por acordos de usuários.” (VAN DIJCK; POELL; WAAL, 2018, p. 9, tradução nossa)⁵.

Essa definição perpassa uma visão de estudos de *software*, porque aponta para a natureza programável das plataformas, e inclui uma perspectiva das Ciências Sociais Aplicadas em diálogo com os Estudos de Linguagem quando aponta para as influências das plataformas *online* em moldar práticas e percepções dos públicos leitores e de suas conexões. É uma conceituação presente nas discussões de pesquisadores brasileiros, como Carlos d’Andréa (2020), Rafael Grohmann (2020), Fernanda Bruno (2020) e Amanda Jurno (2020), e estrangeiros, como Poell, Nieborg, Van Dijck (2020) e Anne Helmond (2015). Ademais, a definição se adequa a esta pesquisa, que tentou compreender fluxos textuais mediados por tecnologias, com reflexos em outros fluxos – econômico, social, administrativo e do trabalho – do mercado multilateral de produção de texto.

³ A expressão está na introdução da entrevista feita pelo Digilabour (Unisinos) com a autora José Van Dijck, em 2019, logo após o lançamento do livro *The platform society: public values in a connective world*. Disponível em: <https://digilabour.com.br/2019/03/06/a-sociedade-da-plataforma-entrevista-com-jose-van-dijck/>. Acesso em: 18 mar. 2022.

⁴ As iniciativas podem ser compreendidas visitando o Observatório do Cooperativismo de Plataforma. Disponível em: <https://cooperativismodeplataforma.com.br/en/home-english/>. Acesso em: 02 jan. 2023.

⁵ Tradução nossa para: “[...] a platform is fueled by data, automated and organized through algorithms and interfaces, formalized through ownership relations driven by business models, and a governed through user agreements”.

Por isso, contextualiza-se também a noção de “plataformização”, a “[...] inextricável relação entre plataformas online e estruturas societárias” (VAN DIJCK, 2019)⁶. A plataformização é a penetração e o imbricamento das lógicas de plataformas na produção de vários setores, criando uma superdependência que vai além das dimensões computacionais. A partir da leitura de Poell, Nieborg, Van Dijck (2020) e Nieborg e Poell (2018), é possível observar que muitos setores passaram a existir dentro das plataformas, sendo forçados a ajustar suas estruturas legais às suas lógicas de funcionamento. No caso do mercado de produção textual (jornalismo, conteúdos para marcas, divulgação científica, mercado dos livros etc.), a crescente dependência das plataformas *online* nos faz enxergar o cenário de ausências dessas arquiteturas computacionais como um estado irreal.

Interessa ainda contextualizar a ideia de “plataformização da *Web*”, que engloba as consequências da programabilidade das plataformas, materializada por códigos e algoritmos. Ser programável é uma característica inerente às plataformas e que transformou completamente a ideia que tínhamos de *Web*, principalmente aquela do início dos anos 2000, cunhada por Tim O’Reilly (2006) como *Web 2.0*. Naquele período, vigoraram discursos otimistas e dotados de imaginários neoliberais sobre o futuro convergente, democrático e participativo que tornaria a *Web* um oásis para construção da inteligência coletiva. Esse discurso se frustrou justamente quando começamos a compreender as lógicas das plataformas e os interesses privados que vêm a reboque. Juanals (2007) constatou que graças à capacidade de armazenamento de imensas bases de dados digitalizados em interconexão, a *Web* parecia se apresentar e como um meio capaz de realizar esse sonho fantástico e enciclopédico (no sentido de que “reúne um conjunto de conhecimentos”) de uma democracia cognitiva, por meio do acesso universal ao conhecimento. Mas a própria autora alertou que era preciso cuidar dos “modos de acesso” e, como veremos, eles foram muito alterados.

A programabilidade das plataformas *online* se baseia em tornar os *sites* na *Web* acessíveis a interfaces técnicas — ou Interface de Programação de Aplicativos (a sigla API, em inglês)⁷. Como modelos de infraestrutura, as plataformas ofertam um sistema tecnológico no qual outros podem construir (HELMOND, 2015). Essa possibilidade de construção de interfaces tem como consequência o crescente intercâmbio de dados mediados pelas plataformas e a capilaridade dos modelos de negócios (D’ANDRÉA, 2020). Esses fluxos de

⁶ Disponível em: <https://digilabour.com.br/2019/03/06/a-sociedade-da-plataforma-entrevista-com-jose-van-dijck/>. Acesso em: 18 mar. 2022.

⁷ As APIs surgiram em 2000, lançadas pelas empresas de comércio eletrônico Salesforce e eBay, e se proliferaram rapidamente. O Flickr começou a usar em 2004, o Google Maps, em 2005, e Twitter e Facebook começaram em 2006. As APIs não são usadas exclusivamente em plataformas comerciais, na verdade, são um protocolo cada vez mais comum em iniciativas governamentais de dados abertos. (PLATIN et al, 2016).

dados, conforme destaca Helmond (2015, p. 50), “[...] não apenas configuram canais para que os dados fluam entre plataformas de mídias sociais e terceiros, mas também funcionam como canais de dados que preparam os dados externos para a configuração de plataformas”.

Sendo assim, perdemos a noção prática do que seria a *Open Web* sem essas trocas de dados entre plataformas *online*. Elas são, atualmente, o modelo de funcionamento computacional que temos em mente, sendo difícil imaginar, por exemplo, a publicação de textos sem Sistemas de Gerenciamento de Conteúdo (CMS) e a circulação deles sem plataformas de busca e de mídias sociais. “Para funcionar (em termos técnicos) e sobreviver (em termos econômicos), blogs, sites pessoais, portais etc. são levados a se adequar aos protocolos de acesso e intercâmbio de dados.” (D’ANDRÉA, 2020, p. 20).

O que chamamos aqui de *Open Web* se refere ao ambiente fora do espaço de redes controladas por *gatekeepers* específicos. O conceito de *gatekeeper* é emprestado da Teoria da Comunicação, mas tratado de forma mais atualizada e contextualizada. A teoria do *gatekeeper* (TRAQUINA, 2012) pode ser compreendida como uma série de escolhas em que o fluxo de informações tem de passar por *gates*, isto é, “portões”. Eles se configuram como áreas de decisão em relação às quais o *gatekeeper* (jornalistas e outros) tem que definir a relevância de uma informação ou não. A expressão *gatekeeper* é também usada pelos setores de economia, políticas concorrenciais e regulação de mercado, quando é necessário definir agentes e imputar obrigações a eles. Para se ter uma ideia, diante da discussão mundial de regulação das plataformas *online*, o *Digital Market Act* (DMA)⁸, documento aprovado pela União Europeia, em março de 2022, para fins de regulação, precisou colocar critérios numéricos para demarcar *gatekeeper*, evitando que as empresas de tecnologia se digam desobrigadas a seguir regras ao se (auto)retirarem da posição de *gatekeeper*.

Sem o domínio de grandes *gatekeepers*, a *Open Web* seria feita por e para todos usuários, a partir de uma visão da *Web* como um recurso público global e acessível. A ideia de *open* se relaciona a conceitos técnicos, como código-fonte aberto, que permitem a qualquer pessoa criar conteúdos visíveis, localizáveis e vinculáveis que são codificados usando padrões públicos (PLANTIN et al., 2016). Mas *open* também engloba conceitos democráticos, como liberdade de expressão e inclusão digital, que podem ser comprometidos quando plataformas *online* assumem o papel infraestrutural da *Open Web*.

⁸ Segundo o DMA, é considerado um *gatekeeper* a empresa que teve faturamento igual ou superior a 6,5 bilhões de euros nos últimos 3 anos e mais de 45 milhões de usuários finais e mais de 10 mil usuários comerciais.

A Web é efetivamente um bem comum global, em grande parte além do controle de qualquer corporação ou governo em particular. Finalmente, a Web está incorporada na vida e na prática cotidiana. Com entretenimento, finanças, saúde, educação e muito mais agora “na Web”, o acesso à Web tornou-se tão crucial para a vida social moderna quanto eletricidade, telefones e esgotos. Claramente, então, a Web é uma infraestrutura. (PLANTIN et al., 2016, p. 13, tradução nossa)⁹.

Importante frisar que esse cenário de “substituição” à *Open Web* pode ser temporário porque os debates sobre *Web 3.0* e metaverso, que engatinham em meio a controvérsias e debates (de otimistas *versus* incrédulos), apontam para a descentralização de fluxos de informação, enfraquecimento de intermediadores na monetização e transparência no controle de dados em ambiente digital (FOLHA DE S.PAULO, 2022). Alguns consideram que a *Web 3.0*, chamada também de *Web Semântica*, já vem se desenvolvendo há pelo menos 10 anos a partir da construção de *softwares* que “[...] conseguem compreender os dados em seus diferentes contextos sem sobreposição de interpretações” (CARVALHO, 2017, p. 82). Segundo essa vertente, tornaram-se mais reais os trabalhos de produção textual automatizada e organização do “caos informativo” em ontologias para que conectemos um conteúdo de maneira muito mais ampla ao “todo” produzido sobre alguma temática. A quem já adiante a crise das plataformas *online* em curto e médio prazo, com a ascensão de modelos mais descentralizados como o Mastodon¹⁰. Para Ronaldo Lemos (2022)¹¹, advogado, diretor do Instituto de Tecnologia e Sociedade do Rio de Janeiro, a ideia de descentralização está na raiz da internet, como abordaremos no Capítulo 4 desta pesquisa, ao destrinchar a construção política e técnica dos motores de busca. É por isso que o desejo pela descentralização pode ser a força motriz para a inovação e o surgimento de novos modelos para praticamente qualquer tipo de aplicação. “A crise nas plataformas nos permite sonhar. E o futuro é inventado primeiro através do sonho.” (LEMOS, 2020).

Enquanto esse debate é incipiente e o poderio das *Big Five* é real, há ainda outra camada que se une ao fato de não enxergarmos mais a *Web* sem as infraestruturas de construção de aplicativos trazida pelas plataformas *online*. Em diálogo com a ideia de plataformização da *Web*, está o conceito de plataformização da produção cultural (NIEBORG; POELL, 2018), que ajuda a entender como as produções tradicionais se adaptam e como

⁹ Tradução nossa para: “The Web is effectively a global commons, largely beyond the control of any particular corporation or government. Finally, the Web is embedded in everyday life and practice. With entertainment, finance, health, education, and so much else now ‘on the Web’, Web access has become as crucial to modern social life as electricity, telephones, and sewers. Clearly, then, the Web is an infrastructure.”

¹⁰ É uma plataforma social, inspirada no modelo microblog do *Twitter*, e que faz parte do Fediverso, uma rede descentralizada e interligada de servidores operados de maneira independente.

¹¹ Disponível em: <https://www1.folha.uol.com.br/colunas/ronaldolemos/2022/12/crise-das-plataformas-vai-criar-opportunidades-em-2023.shtml>. Acesso em: 27 dez. 2022.

novos modelos surgem, atendendo radicalmente as lógicas de plataformas *online* ou usando taticamente essas lógicas.

As plataformas permitem formas de produção direcionadas pela demanda, mas definem princípios de governança (regras econômicas e padrões tecnológicos) que orientam os produtores de conteúdo e desenvolvedores de *software*. Dessa forma, um conteúdo passa a funcionar mais como uma “mercadoria cultural contingente” (*contingent cultural commodity*) (NIEBORG; POELL, 2018, p. 11) e isso se relaciona com a performance desse conteúdo. É que a coleta de dados dos usuários (*feedbacks* e rastros de uso) permite saber que tipo de conteúdo “funciona” ou “não funciona” para as audiências, o que impulsiona remodulações contingentes e revisões nas formas de produção em plataformas *online* (CRUZ; VECCHIO, 2022). A esse processo de rastreamento e organização de dados de bilhões de usuários, damos o nome de “datificação” (VAN DIJCK, 2017). O que alimenta esse processo é certa repetição de fórmula de sucesso para desempenho: “[...] na contingência da performatividade – os perfis que ganham proeminência são aqueles mais recorrentes na interação entre o usuário e o algoritmo” (CASTRO, 2019, p. 21). Ponderamos que tal contingenciamento da produção cultural não é uma novidade da era das plataformas *online*, mas acontece em maior escala por causa das condições ferramentais propiciadas pela datificação. Escritores, jornalistas e editores, por exemplo, nunca tiveram *feedbacks* tão rápidos e visualmente amigáveis sobre a performance de suas produções perante os públicos.

Em concordância com esse debate sobre datificação, Gillespie (2018) analisa a influência exercida pelos algoritmos de plataformas *online* por meio de personalização de resultados de pesquisa ou recomendação de conteúdo. A partir do recolhimento de rastros, *cookies* e outras informações que doamos ao assinar as políticas de uso, as plataformas *online* fazem “vigilância permanente” das atividades dos leitores, dinâmica que Gillespie (2018) nomeia criação de “públicos calculados”. Em seminário promovido pelo comitê editorial especial da *Revista LIINC*, o pesquisador Sérgio Amadeu Silveira (2022) argumentou sobre o cenário panóptico em vários segmentos da sociedade. Citando Zuboff (2019), o pesquisador descreveu o “capitalismo de vigilância”, que busca derivar o excedente monetário da datificação da experiência humana, como uma forma de mercado inimaginável fora do meio digital. Para Silveira (2022), as plataformas controlam a visualização e modulação de informações a partir de um modelo de gestão algorítmica para exercer controle da atenção daqueles que tiveram atenção capturada. Segundo o pesquisador, as plataformas não param de coletar dados para predizer o futuro por meio de registros do passado, mantendo-se em terreno

neutro, quando na verdade a neutralidade é impossível, porque essas tecnologias sequer são auditadas.

Um dos efeitos da datificação é a criação de um ciclo de contingenciamento: produtores de conteúdo rastreiam preferências de audiências (públicos leitores/consumidores), remodelam processos de produção com ajuda das possibilidades de reconfiguração de *design* que as plataformas oferecem. Um exemplo dessas possibilidades de remodelações contingentes é o surgimento de milhares de *plugins*¹² para *Wordpress.com* que injetam funcionalidades nesse CMS, sacramentando a participação ostensiva de empresas e desenvolvedores terceiros e atendendo a demandas de produtores de texto que precisam se adequar aos *trends* de audiência definidos pelas plataformas de busca. É praticamente um ciclo de “comodificação” de conteúdo em ambiente digital. E estamos nos referindo somente à busca orgânica, aquela em que o motor de busca entregará resultados de ordenação de informação não patrocinada. Sobre os *links* e anúncios pagos, certamente cumprem um papel peculiar na circulação de conteúdo, mas não os estudamos especificamente.

Retomando, Castro (2019) corrobora a discussão sobre o ciclo de contingenciamento quando avalia que as informações geradas no interior de uma plataforma *online*, a partir da interação com usuários, prestam-se a retroalimentá-la. Isso se dá nomeadamente no espaço restrito de uma plataforma, cada qual funcionando de modo relativamente estanque:

A interpelação algorítmica é recursiva não apenas por respaldar-se em dados do usuário, mas também por responder a uma demanda dele. Num mecanismo de busca como o Google, digitamos ou falamos algo para pôr em marcha uma pesquisa; numa rede social como o Facebook, acedemos a nosso *feed* de notícias com a expectativa do que vai cativar nossa atenção; numa loja on-line como a Amazon, buscamos algum produto ou consultamos as sugestões. Todas essas situações equivalem a uma espécie de demanda de nossa parte, porquanto nos disponibilizamos mais ou menos ativamente a receber uma interpelação externa. Há uma adaptação recíproca nos casos de demanda à plataforma pelo usuário e interpelação deste pela plataforma. (CASTRO, 2019, p. 6).

Partimos da hipótese de que produtores de texto trabalham para cumprir o ciclo de contingenciamento visando atender as demandas algorítmicas completamente influenciados pelos *vieses* das plataformas de busca, sendo esse trabalho ambivalente – ora com ações de obediência, ora com usos estratégicos. Assim, produtores se munem de ferramentas, redefinem suas práticas produtivas, como, por exemplo, o uso de *plugins* de edição de texto e

¹² *Plugins* são programas de computador que funcionam como módulo de extensão para adicionar funções em outros programas. Em 8 de abril de 2022, o próprio *Wordpress* contabilizava 59.78 *plugins*.

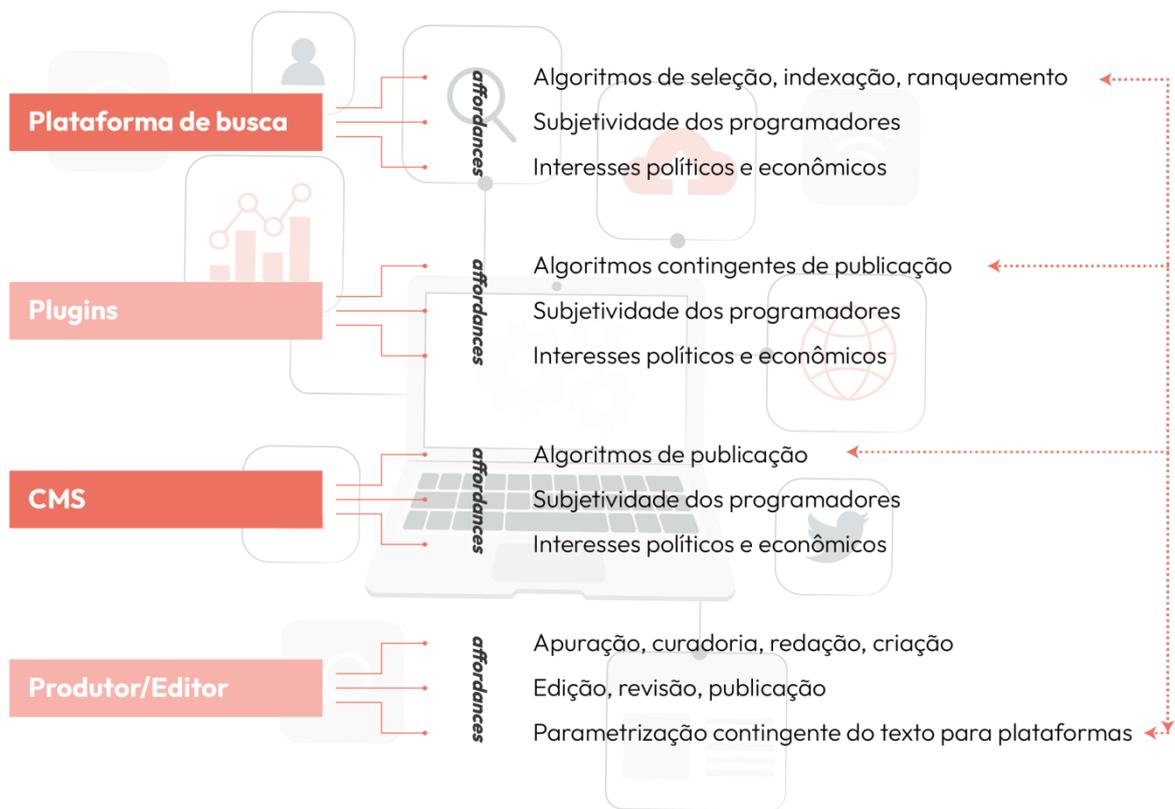
as práticas de engenharia reversa¹³. Não sabemos em que medida – frequência e abrangência – essas práticas ocorrem, mas parece ser comum. Por isso, propusemos-nos a verificar como produtores e editores de conteúdo em ambiente digital são confrontados com escolhas que vão além da linguagem.

Alguns jornalistas, por exemplo, vão defender que essa reconfiguração é parte de um processo de compreensão do meio digital, tal como os profissionais fizeram em outros momentos, ao adaptar suas produções textuais à TV, ao rádio, ao impresso. A estrutura textual feita para cada meio é diferente e contingente às características inerentes aos *media*. No caso do ambiente digital plataformizado, os profissionais do texto, muitas vezes, tentam deduzir os critérios dos algoritmos e isso ocorre porque as ferramentas computacionais passaram a influenciar os meios de expressão e circulação de informação. Essa dedução está refletida em algumas práticas incorporadas que nos fazem pensar sobre o quanto a humanidade sujeitou seus textos, discursos e os conhecimentos às lógicas procedimentais sustentadas pela computação.

Defendemos também a ideia de grandes camadas do processo de produção e edição de texto em CMS, que podem ajudar na descrição de fluxos plataformizados, estes que dependem de plataformas *online* para existir. Amanda Jurno (2021) trabalhou com a ideia de camadas de mediação no acesso ao *Facebook* para explicar a plataformização do jornalismo. A pesquisadora se preocupou com aspectos que influenciam o acesso às informações nas plataformas de mídia digital e, em nossa pesquisa, o foco é nos elementos que (re)moldam a produção textual.

Figura 1 – Camadas do processo de edição de texto em ambiente digital no contexto da plataformização

¹³ No decorrer da tese, o uso da expressão “engenharia reversa” ficará mais recorrente e contextualizado, mas, em suma, engenharia reversa é um conjunto de técnicas incorporadas por criadores de texto para que suas produções circulem e tenham visibilidade em ambiente digital, principalmente, cumprindo requisitos de indexação e ranqueamento do *Google*.



Fonte: Elaborado pela autora, 2023

Tomamos como hipótese que a camada “Produtor/Editor” reúne as práticas de escrita, com suas subjetividades e modulações, e as outras camadas concentram as relações e experiências dos usuários com os artefatos tecnológicos projetados/ofertados pelos sistemas, as *affordances* de CMS, *plugins* e plataformas de busca. Já a camada “CMS” engloba toda a lógica programada de espaços para publicação de textos projetados para atender, em configuração mínima, a leigos que não sabem criar uma página na *Web*.

Consideramos ainda a camada “Plugins”, por conta da interdependência entre essas ferramentas e os CMS. Mas, por que os *plugins* funcionam como camada para as cadeias editoriais do texto digital? Porque muitos deles são usados como instrumentos (re)definidores de práticas produtivas com a intenção de auxiliar produtores e editores de texto na criação e publicação de conteúdos mais elegíveis aos motores de busca, em especial do *Google*. Muitos *plugins* se colocam como “bússolas editoriais” e poderão sugerir alterações de tamanho do texto, escolhas semânticas (como mudar a frase-chave), uso de conectivos (conjunções e pronomes), entre outras práticas, como mostraremos no Capítulo 5.

Enfim, sobre a camada “Plataforma de busca”, acreditamos ser a mais determinante de contingenciamentos. Martha Gabriel (2010) discute como o crescimento do conteúdo digital criou grande *continuum* textual e a complexidade de leitura/consumo

propiciaram “[...] uma era da busca” (GABRIEL, 2010, p. 83). Chartier (2002, p. 20) também previu essa dinâmica: “[...] o mundo da comunicação eletrônica é um mundo de superabundância textual cuja oferta ultrapassa a capacidade de apropriações dos leitores”. Fatores como colaboratividade, computação ubíqua (alavancada pelas plataformas móveis), tentativa de veículos de comunicação em se colocar como produtores do “tempo real” colaboram para a explosão da quantidade de conteúdo *online*. Assim, fomos conduzidos à “solução” construída pelas plataformas de busca, em especial, o *Google*, que moldaram seus negócios e modelos tecnológicos para ocupar espaços de oráculos modernos. Atualmente, o *Google* exerce um poder que vai muito além da função de porta de entrada da internet ou espaço de consulta, tornando-se umas das plataformas *online* que melhor estabelece suas lógicas de poder na detenção das informações mundiais, afinal “[...] o que não é encontrado, não existe” (GABRIEL, 2010, p. 217).

Por todo o exposto anteriormente, nossa tese pode ser relevante. Ela se justifica pela importância dos Estudos de Plataforma (*Platform Studies*) e Estudos de Algoritmos (*Critical Algorithm Studies*), pois temos visto um crescente número de pesquisas sobre o “[...] complexo e assimétrico jogo de poder” (D’ANDRÉA, 2020, p. 14) que se estabelece entre os usuários e plataformas *online*. *Twitter*, *Facebook* e *YouTube*, *Uber*, em especial, são muito estudados, o que resulta em bibliografia significativa em inglês e português, diferentemente dos motores de busca, cujos estudos de perspectiva crítica são poucos em português.

García-Orosa, Canavilhas e Vázquez-Herrero (2022) fizeram uma revisão sistematizada da literatura sobre “algoritmos e comunicação” e, das 243 publicações analisadas, 18% tinham como tema de estudo as plataformas, ficando atrás apenas dos enfoques em “esfera pública, democracia e comunicação política” e “jornalismo”. Dar continuidades a essas abordagens possibilita um conhecimento amplificado que resulta da experimentação do mundo que, por sua vez, permitiria sua transformação.

Destacamos alguns fatos que ajudam a refletir a importância de debates sobre platformização e os desdobramentos nos espaços sociais e simbólicos. Aliás, esses fatos nos mostraram que a repercussão da pesquisa científica sobre o tema pode ser um subsídio para mudanças da percepção social sobre essas plataformas *online*. O primeiro deles é a relação entre uso de plataformas e inclinação política. Uma reportagem¹⁴ da *Folha de S.Paulo* sobre a pesquisa da antropóloga Rosana Pinheiro Machado relatou o quanto o trabalho em aplicativos (*Uber*, *Ifood*, *WhatsApp*) está empurrando pessoas para a direita na política, pois, nesses

¹⁴ Disponível em: <https://www1.folha.uol.com.br/mercado/2022/03/trabalho-por-app-pode-estar-empurrando-pessoas-para-a-direita-diz-antropologa.shtml>. Acesso em: 21 mar. 2022.

espaços de circulação de informações, os trabalhadores de países emergentes encontram ambientes férteis para discursos de extrema direita. Já o Núcleo de Jornalismo detalhou a máquina¹⁵ empenhada pelo *site Brasil Paralelo* para dominar as buscas no *Google* e tornar visíveis conteúdos alinhados ao bolsonarismo no Brasil, em especial, com a compra de palavras-chave sobre assuntos politicamente sensíveis para a extrema direita. Em complementaridade, e como um clímax da narrativa política brasileira dos últimos anos, o *TikTok*¹⁶ relatou ter excluído mais de 10 mil vídeos golpistas no período dos ataques à Esplanada dos Ministérios, em janeiro de 2023.

O segundo fato é uma percepção, mesmo que inicial, sobre a precarização do trabalho no cenário das plataformas *online*, aliás, fenômeno nomeado “emprecariado” (empreendedorismo + precariado) por Silvio Lorusso (2023). Em março de 2022, foram divulgados os resultados da pesquisa *Fairwork Brasil*¹⁷; o levantamento, que é uma colaboração de pesquisadores da Universidade do Vale do Rio dos Sinos (Unisinos), Universidade de Oxford e Instituto WZB Berlin Social Science Center, mostrou que *Uber*, *99*, *Rappi* e *Ifood* têm notas pífias na avaliação sobre condições do trabalho. Alguns veículos de imprensa brasileiros repercutiram os baixos pagamentos e queixas de tratamento desumano a que os trabalhadores são submetidos. A luta por direitos dos trabalhadores segue em instâncias acadêmicas e mobilizações dos próprios profissionais, como podemos acompanhar nas iniciativas relatadas pelo *Observatório do Cooperativismo de Plataforma*¹⁸.

O terceiro fato é o investimento em guerras de narrativas em plataformas *online*. A reportagem da *Agência Pública*¹⁹, publicada em abril de 2020, desvela a máquina oculta de propaganda do *iFood* para desacreditar e enfraquecer o movimento de motociclistas em busca de melhores condições de trabalho. A estratégia brutal da empresa, contratando agências especializadas em *marketing* digital, demonstra uma capacidade sem fim de investimento na manutenção de poder. Além disso, evidencia como as empresas detentoras das plataformas têm urgência em se manter vivas no sistema político e econômico e usam a guerra de narrativas em plataformas de mídias sociais para contingenciar discursos que as favorecem. Por fim, escancara a “transgressão lucrativa” (CHARAUDEAU, 2010, p. 263)

¹⁵ Disponível em: <https://nucleo.jor.br/especiais/2023-01-31-a-maquina-do-brasil-paralelo/>. Acesso em: 17 fev. 2023.

¹⁶ Disponível em: <https://www1.folha.uol.com.br/poder/2023/02/tiktok-relata-105-mil-videos-golpistas-derrubados-em-81-demais-big-techs-silenciam.shtml>. Acesso em: 21 mar. 2023.

¹⁷ Disponível em: <https://fair.work/wp-content/uploads/sites/131/2022/03/Fairwork-Report-Brazil-2021-PT-1.pdf>. Acesso em: 21 mar. 2022.

¹⁸ Disponível em: <http://cooperativismodeplataforma.com.br/>. Acesso em: 17 fev. 2023.

¹⁹ Disponível em: <https://apublica.org/2022/04/a-maquina-oculta-de-propaganda-do-ifood/>. Acesso em: 08 abr. 2022.

praticada pelas agências de *marketing*. Estas sobrevivem alimentando lógicas das plataformas, ao recriar jogos de conteúdos feitos para performar na guerra de narrativas. Charaudeau (2010), ao discutir a deontologia e questões de responsabilidade no discurso, argumenta a inexistência de um sistema de autorregulamentação midiática, em especial do *marketing* e da publicidade, para penalizar “comportamento desviante” (p. 262). Pelo contrário, quanto maior a desobediência à regra (isto é, transgressão), melhor assegura a captação do público. Sendo assim, tudo acontece “[...] como se a ausência da transgressão não pudesse garantir impacto junto ao público” (CHARAUDEAU, 2010, p.263). Nesse caso do *iFood*, a agência de *marketing* e publicidade parece servir à plataforma *online*, sendo parte do jogo desviante de conduta ética adotado pela plataforma.

Esses três fatos – a relação entre uso de plataformas e inclinação política, a percepção sobre a precarização do trabalho no cenário das plataformas *online* e o investimento em guerras de narrativas em plataformas *online* – nos ajudam, com casos brasileiros, a extrapolar o que foi mostrado no documentário *Dilema das redes*²⁰. Esses fatos não constroem narrativas destrutivas às plataformas, trazendo pessoas de dentro das grandes empresas para desvelar o “mal”, soando à traição. Esses casos revelam as engrenagens do “imperialismo de plataforma”, que é oculto ao grande público, e comprovam que não há limites (financeiros, éticos, narrativos) para o estabelecimento e manutenção da ordem de poder. Isso posto, nossa tese pode ser relevante e aplicável por continuar discutindo as lógicas de plataformas *online* cumprindo a missão de empoderar²¹ produtores de texto e leitores sobre a influência desses artefatos tecnológicos nos modos de vida e relações sociais. Pode contribuir também para a busca (talvez utópica) de subverter o atravessamento dessas plataformas *online* em nossas práticas profissionais e na vida. Consideramos que a tese é concebida a partir de uma visão tripa de fazer o leitor entender o jogo – destrinchando o cenário de plataformas online; jogar o jogo – detalhando funcionamentos do Google e parametrização textual para ranqueamento; e subverter o jogo – convidando à criação de táticas e estratégias para o futuro.

Registramos como pertinente certo ineditismo desta pesquisa na tentativa de explorar a caixa-preta, mesmo que parcialmente, da plataforma *Google*, em uma interrelação

²⁰ O filme, dirigido por Jeff Orlowski, revela como as plataformas *online* têm sido usadas para coletar dados, prever o comportamento humano e impulsionar mudanças de preferência de consumo àquelas que dão mais lucro. Também elucida os bastidores das grandes empresas de tecnologia. Um dos trechos mais populares do filme é o momento em que Tristan Harris, ex-designer do *Google*, citando uma máxima das empresas de internet, afirma: “Se você não está pagando pelo produto, então você é o produto”.

²¹ Aliás, o trabalho de empoderamento de trabalhadores tem sido feito de forma incrível por pesquisadores do *Projeto Fairwork* (<https://www.youtube.com/channel/UC-DyfwUBCeryF2j1d8CIM-A>), do *Observatório do Cooperativismo de Plataforma* (<https://cooperativismodeplataforma.com.br/>), entre outras iniciativas.

complexa com a produção textual, com o olhar dos Estudos de Linguagens. Caixa-Preta é uma metáfora muito acionada para especificar a dificuldade ou a impossibilidade de se conhecer detalhadamente como opera um artefato tecnológico tão contingente quanto uma plataforma *online*. Pesquisadores dos Estudos de Plataformas se apropriaram do termo também para dizer sobre a estabilização de uma rede sociotécnica. (D'ANDRÉA, 2020). É provável que, mesmo com as pesquisas acadêmicas, a caixa-preta permaneça e isso não é um problema, pois não abrir-la em sua totalidade poderá nos dar apontamentos sobre a opacidade do buscador *Google*. Mais do que fornecer explicações claras e matizadas do desenvolvimento tecnológico, temos a chance de pensar nas práticas de produção de vários profissionais do mercado, como jornalistas, escritores, publicitários, professores, profissionais de *marketing*, divulgadores científicos, abordando as consequências na rotina desses trabalhadores. Falar sobre essa rotina mostra nossa preocupação com a condição desses trabalhadores, afastando as críticas de Winner (1993), por exemplo, a pesquisadores que revelam e rastreiam atores sociais, processos das caixas-pretas tecnológicas, mas não questionam os compromissos e projetos básicos da sociedade tecnológica moderna, principalmente, no que diz respeito à classe trabalhadora.

Nossa tese se justifica como retomada do assunto desenvolvido no mestrado, a partir de uma visão revisional e mais madura. Entre 2012 e 2014, quando pesquisamos o processo de produção de repórteres de um jornal mineiro que incorporava práticas de análises de resultado em audiência e gerenciamento de *tags* ao trabalho jornalístico (CRUZ, 2014), denominamos algumas dessas habilidades de técnicas de *folksonomia* reversa e hierarquia reversa. Naqueles casos analisados, já havia uma consciência mínima de critérios algorítmicos que davam visibilidade ou não a conteúdos no ambiente digital, mas os efeitos político-sociais do atravessamento de *softwares* não foram problematizados.

Os trabalhos nesse sentido amadureceram, apesar de os pesquisadores (tanto quanto os produtores de texto) terem pouco acesso ao universo “encaixapretado” dos algoritmos. Os cientistas têm se dedicado mais frequentemente a revelar as consequências sociais de ter os programas como mediadores do conhecimento coletivo. Para Gillespie (2018, p. 106), “[...] a maioria dos usuários não se debruça sobre os critérios algorítmicos e tende a tratá-los como ferramentas não problemáticas a serviço de uma atividade maior”. Apenas entrelaçar as lógicas computacionais às práticas de produção textual demonstra falta de compreensão do quanto o mundo é alterado por essas entregas das plataformas *online*, que são instrumentos de (grandes) negócios. Quando o produtor luta para vencer o incompreensível

algoritmo, sem questioná-lo politicamente, poderá alcançar a visibilidade *online*, mas nunca elevará seu trabalho ao contrafluxo de poder.

A minha proximidade com o objeto de pesquisa – sendo jornalista, produtora e editora de conteúdo digital – também me motivou a estudá-lo e, a partir do processo de pesquisa, fazer um exercício de leitura do mundo. Pertencendo ao universo de quem produz texto em ambiente digital, pretendo interpretar esse mundo estando disposta a enfrentar os confortos e desconfortos que isso venha a trazer a mim como “sujeito de pesquisa”. “O sujeito que pergunta para construir sua pesquisa interroga a si próprio.” (HISSA, 2012, n.p.).

Por fim, registramos que um dos pontos de partida desta pesquisa é a possibilidade de mapeamento das práticas a partir de uma perspectiva teórica crítica. Esse movimento pode ser importante para as áreas de Linguística Textual e Comunicação. Para corroborar, recorro a duas pertinentes produções sobre estudos de texto e escrita. Para Bakhtin (1992, p. 267), “[...] os textos são correias de transmissão entre a história da sociedade e a história da linguagem”. Em cada época da evolução da linguagem, um tom diferente é dado às produções. Ribeiro (2015, p. 115) reforça esta perspectiva:

Não estamos alheios à história da escrita, como se ela fosse um modo de fazer que corresse ao largo dos cidadãos. A adesão a novas máquinas, novos modos de produzir texto, novos gêneros textuais são “criações” sociais, menos ou mais inusitadas, inovadoras, que correm conosco na história da leitura e dos modos de escrever. As técnicas e tecnologias da escrita de que dispomos hoje são mais uma fase dessa história, que não despreza nenhuma outra anterior.

Portanto, temos um cenário no qual convivem processos de variadas formas do texto, misturadas aos nossos modos de vida, às nossas vivências, ao nosso modo de operar em sociedade. É uma grande oportunidade de nos debruçarmos sobre esse momento social, penetrado pelo “imperialismo de plataforma”, que certamente deixará legados à história da escrita. É também um grande desafio estudar cultura e sociedade e fundamentar as descobertas com a internet.

Além desta Introdução, esta tese está organizada em duas partes, distribuídas da seguinte maneira: na parte I – *Sentidos* –, apresentamos ideias fundadoras e inspiradoras para as percepções e interpretações que fizemos no decorrer da pesquisa. Esta parte é composta pelo Capítulo 2 – *Produção, leitura e circulação de texto em ambiente digital* –, em que argumentamos nosso entendimento de texto e textualidade, em uma conceituação moderna e atualizada, pensada a partir do atravessamento de plataformas online na escrita em ambiente digital. O capítulo é também uma releitura crítica de definições propostas por linguistas,

chegando ao estado da arte sobre o tema e colocando em questão especificidades sobre o texto plataformizado. No Capítulo 3 – *Metodologias* –, inspiramo-nos em metodologias que estudaram criticamente plataformas online e algoritmos, em especial motores de busca, para repensar o quanto a materialidade dessas tecnologias são parte da reconfiguração de práticas de produção e edição de texto. Com essas inspirações em mente, organizamos nosso método de pesquisa.

A parte II desta tese – *Materialidades e experiências* – é um esforço para demonstrar que as plataformas que medeiam a produção textual não são objetos neutros e constituem a rede digital com uma materialidade radical, sendo necessário desdobrar as interações em camadas, como as descritas na Figura 1 desta Introdução. O Capítulo 4 – *A plataforma de busca Google* – é uma reunião de leituras documentais e um panorama histórico sobre a *Google* como plataforma de busca, como empresa e como mediadora da circulação textual na atualidade. Reunimos e analisamos documentos disponibilizados pelo *Google* para compreendê-lo como um sistema de mediação que gera demandas contingentes de produção textual, em especial, a partir do *Core Update Bidirectional Encoder Representations from Transformers* (BERT), de 2019. No Capítulo 5 – *Práticas de escrita e edição incorporadas ao texto plataformizado* –, demonstramos detalhadamente a parametrização contingenciada da produção textual, analisando as implicações. Nele também nos dedicamos a reunir iniciativas subversivas aos algoritmos e modos de funcionamento das plataformas de busca comerciais.

Nas Considerações Finais, está a apresentação de resultados obtidos, além da retomada da pergunta que guiou esta pesquisa. Lá também compactamos algumas das principais contribuições da pesquisa, como a concepção de texto plataformizado, a discussão sobre o mito da relevância e o retrato de práticas textuais incorporadas por produtores, e apontamos rastros para estudos futuros.

O leitor perceberá que uma carta introduz cada capítulo, revelando detalhes da experiência como pesquisadora, jornalista, professora e cidadã no que diz respeito ao atravessamento das plataformas de busca em minha vida. A intenção é me fazer presente e grata aos que se dispuserem a apreciar o processo desta investigação.

PARTE I: SENTIDOS

Qualquer tecnologia de que dependamos constitui um desafio e uma ameaça. Quanto maior é a dependência em relação a uma tecnologia, maior é a necessidade de analisar e expor seu funcionamento interno, independentemente de direitos ou questões legais (The Critical Engineering Manifesto, 2011-2021)

Prezado, Sírio Possenti.

O dia em que eu dei uma carona para você, em Belo Horizonte, foi bem curioso.

Observador, você entrou no carro e encontrou um monte de revistas Minas Faz Ciência empilhadas, e eu, claro, não perdi a oportunidade de oferecer uma para que você lesse. Assim, você conheceria parte do meu trabalho e da ciência produzida em terras mineiras.

Te levei até o CEFET-MG, onde daria uma palestra, e durante boa parte do caminho você ficou folheando a revista. Perguntei se havia achado bom o material e você deu um “sim” bem modesto, ao seu modo. Um linguista sabe quais são as condições para um bom texto, aliás, você falaria sobre isso logo mais tarde para estudantes, professores e pesquisadores. Você desceu do carro e deixou para trás a pilha de revistas.

Da palestra, destaco os nove elementos elencados por você para que um texto seja bom: 1) adequação: texto adequado ao meio e ao destinatário; 2) estilização: texto melhorado de uma primeira visão e que tenha estética textual e visual dando luz aos modos de escrita; 3) senso de direção: texto que “escolhe” o leitor; 4) subversão de gênero: texto que não se prenda a limites de um gênero, que extrapole condições previamente impostas às pressuposições, de forma a surpreender o leitor; 5) código linguageiro especial: texto em que se reconhece o autor porque ele pratica uma variante linguística no interior da língua; 6) manejos intencionais: texto que faz escolhas competentes de palavras para substituir outras que poderiam se repetir, afinal, não existe sinônimo, existe escolha; 7) que sabe dar voz: falamos e escrevemos a partir dos outros; 8) que preze pela tessitura: não é a ideia que faz o texto, mas a tessitura em si; e 9) que faça uso competente de verbos dicendi.

Achei tudo aquilo tão legal que, além de anotar ponto a ponto, incrementei umas aulas minhas com o material. Atrevi-me a fazer uma releitura daquela fala para disciplinas sobre o texto produzido em ambiente digital, acrescentado um ponto de reflexão que nos traz exatamente para as discussões aqui da minha pesquisa: relevância para visibilidade – atender a um sistema de expectativas do leitor e ao sistema de demandas contingenciais algorítmicas que farão um texto ser encontrado e lido.

Você nem sonha que eu ando por aí associando os nove elementos elencados na palestra a esse lance da relevância, fazendo virar dez elementos. Ao final da palestra, fomos embora de carro. Na porta do hotel onde estava hospedado, antes de descer, você pediu para levar uma revista Minas Faz Ciência, pois faria uma leitura mais cuidadosa. “Opa, é claro”, eu disse.

O dia em que eu dei uma carona para você, em Belo Horizonte, foi bem curioso.

Lembra deste dia?

Abraços,

Luana.

2 PRODUÇÃO, CIRCULAÇÃO E LEITURA DE TEXTO EM AMBIENTE DIGITAL

Textos não devem ser compreendidos com base em sua imanência (a partir da essência), mas, sim, diante de suas condições de “produção, circulação e leitura” (ou seja, a partir da sua existência). Textos são fenômenos emergentes que se efetivam a partir das múltiplas relações estabelecidas, por exemplo, nos fluxos editoriais em que estão inseridos e com as materialidades em que são produzidos e nas quais circulam. Por isso, os esforços de muitos pesquisadores em consolidar o entendimento de texto são, na verdade, trabalhos para expandir o conceito, assim como fizeram Costa Val (2000), trazendo princípios de Insenberg (1968) e as definições de van Dijk (1973); Leal (2018), apresentando texto e textualidade como processos; e o próprio Sírío Possenti (2020), ao reconhecer que os nove elementos elencados são moldados a partir da situação de comunicação na qual se insere um texto. Todos parecem dizer que a identidade de um texto é sempre contingente, uma vez vinculada ao trabalho da comunicação, como se observa neste trecho:

Todo texto é um plasma provisório de um processo em curso, no qual elementos heterogêneos se apresentam articulados, em mais de um nível. Isso é dizer que todo e cada texto é necessariamente “híbrido”, “intertextual”, “dialógico”, “impuro” e “heterogêneo”, ainda que nem sempre polifônico, a depender do nível que se observa e da definição de polifonia mobilizada. As metáforas da rede e similares não expõem apenas nós, linhas visíveis e espaços aparentemente (só aparentemente) vazios, mas apelam ao caráter multidimensional, multiforme, impreciso, inconcluso e diverso da textualidade e dos processos comunicacionais (LEAL, 2018, p. 26).

Costa Val (2000) retoma os princípios constitutivos da textualidade, tal como elaborados na Linguística Textual: coerência, coesão, intencionalidade, aceitabilidade, informatividade, situacionalidade e intertextualidade. Esse movimento é quase repetido por Possenti (2020), mas ambos os autores relativizam a necessidade de ir além da identificação de elementos e características formais que parecem compor um “todo” homogêneo, inclusive para ensinar as pessoas a produzir textos. Assim:

Quando se pensa no texto não como produto em si, mas como resultado de uma atividade linguístico-cognitiva socialmente situada, o ensino da escrita começa por explicitar aos alunos a necessidade de pautar o trabalho de redação por perguntas voltadas para a dimensão interacional: por que e para que estou escrevendo? Quem é meu leitor? Em que suporte meu texto vai circular, em que condições vai ser lido? Quando essas questões orientam a produção textual, fica claro para o aluno-autor que o processo envolve escolhas, decisões, detecção e resolução de problemas, e que é preciso empenhar-se no gerenciamento da propriedade, eficiência e eficácia do texto que está sendo construído, definindo estratégias de dizer adequadas a seus objetivos e à situação dos leitores previstos. (COSTA VAL, 2000, p. 50).

É útil também entender texto como tudo que produz sentido, uma combinação de signos interpretada de forma dinâmica e em meios e contextos sociais de leitura diversos. Todavia, produzir sentido é mais do que ser “veículo de significado” (LEAL 2018), mas, sim, compartilhar sensações, sentimentos, afetos: “[...] nenhum texto é um veículo, mas um modo de agir, uma forma de experiência, de múltiplos sentidos e, por que não, múltiplas referências” (LEAL, 2018, p. 33). Une-se a isso a utilidade de considerarmos como texto todo modo que inclua dados verbais, visuais, orais e numéricos, em forma de mapas, fotos, impressos e música, de arquivos de registro de som, de filmes, vídeos e qualquer informação conservada em computador ou nuvem (MCKENZIE, 2018). Por isso, consideramos que todo texto tem um caráter multimodal (KRESS; VAN LEEUWEN, 1998), assim, podemos nos distanciar de uma visão verbocêntrica e midiocêntrica (LEAL, 2018). Por todos esses aspectos, defendemos o uso do termo texto em visão expandida, principalmente ao ser referir às produções em ambiente digital, que transformam as materialidades do texto:

Há inclusive aqueles que recusam o uso do termo “texto” no âmbito das pesquisas sobre esses temas, em nome da tradição verbocêntrica e semântica à qual ele é frequentemente associado. Se se tem em vista a expansão da noção de texto, porém, a dissociação das textualidades e os afetos, materialidades e presenças, ou seja, e as demais relações de sentido, não se justifica. Reconhecer a riqueza das textualidades é também respeitar e ter em conta as complexidades dos processos comunicacionais, enfrentando, então, os desafios que aí se apresentam. (LEAL, 2018, p. 27).

Ribeiro (2016) já se dedicou muito a estudar o texto em seus multimodos, reconhecendo que, por muito tempo, o conceito de texto se focalizou apenas nas palavras, mesmo os textos (e seus conceitos) se alterando com o passar do tempo.

Os critérios para definição de texto mudam, assim como suas materialidades são alteradas, com esta ou aquela tecnologia; suportes provocam ou propiciam novas práticas e, assim, a teoria só pode escolher abrir ou fechar seus recortes, ao tratar do que seja um texto. Este está sempre em movimento, quer a teoria queira ou não. (RIBEIRO, 2016, p. 35).

Portanto, torna-se fundamental (re)construir uma noção de texto e de escrita em ambiente digital para, enfim, refletir sobre a reconfiguração de práticas de produção. O texto atravessado por sua forma, por sua relação com outras semioses e, especialmente, pelos dispositivos das plataformas *online*. É um desafio pensarmos em como experimentávamos produzir texto antes de existirem CMS e plataformas de redes sociais, de busca, de vídeo, de áudio, de mensagem instantânea.

2.1 TEXTO PLATAFORMIZADO

Vamos pensar a plataformização como um dos desdobramentos do entrelaçamento entre texto e tecnologia, e demarcar o quanto isso influencia o campo das Linguagens. Partimos da ideia de que somos uma sociedade muito textualizada, muito antes de sermos uma sociedade plataformizada. Na obra *Livro: edição e tecnologias no século XXI* e no verbete “Livro”, inscrito em *Tarefas da edição*, Ana Elisa Ribeiro dialoga com Roger Chartier revisando os “[...] movimentos atuais da cultura escrita” (RIBEIRO, 2018b, p. 28) modificados pelo mundo eletrônico. Pensando especialmente no livro, a autora debate reflexões relevantes do historiador, entre elas as rupturas da ordem do discurso, ou seja, a fissura na relação que se estabelece entre objetos/materialidades, categorias de textos e formas de leitura.

Herdamos da cultura impressa a relação com o objeto material e a obra intelectual, quase sempre identificados por autor, paginação, capa, lombada (como vemos em livros, jornais, revistas). A isso Chartier (2002, p. 22) dá o nome de “estética identificada”, e Ribeiro (2020, p.88) chama de “arquitetura”, referindo-se à organização de um texto – digital ou impresso – que demarque certa “genealogia”. “É essa ordem do discurso que se transforma profundamente com a textualidade eletrônica” (Chartier 2002, p. 22), pois a tela (de computador, celular e outros) vira suporte único para qualquer forma ou gênero textual, que se misturam em um *continuum* abundante e difícil de diferenciar os “diversos discursos a partir de sua própria materialidade”.

Aliás, Ribeiro (2020, p. 85) vai além e chama a dupla *hardware/software*, tais como *e-reader/aplicativo*, de “objetos não-dedicados”. Segundo ela, “não-dedicados” porque permitem a leitura de livros na mesma medida que podem ser um console de jogo ou uma televisão. Ribeiro (2018) vai lembrar que a tela pode ser suporte para “produção, circulação e leitura”, sendo que na cultura impressa de cada uma dessas “funções” estaria distribuída em suportes diferentes, a exemplo da prensa, cujo encargo era apenas de imprimir. Chamamos atenção para o fato de que as plataformas *online* são dispositivos dentro do suporte tela que reúnem a possibilidade de concretização do ciclo “produção, circulação e leitura”, portanto, são parte integrante das rupturas da ordem do discurso, conforme detalharemos a seguir.

Plataformas *online* podem ser “espaços de escrita” (BOLTER, 1991) pensados como campo físico e visual em que “acontece” o texto, afinal todas as formas de escrita são espaciais e exigem um lugar em que se inscreva/escreva. E podem ser também “equipamentos de leitura” (FURTADO, 2003), compreendendo leitura como um[...] entrelaçamento entre

habilidades e estratégias de ler e navegar (COSCARELLI, 2016). Essa noção de leitura é útil, especialmente, ao nos referirmos a plataformas de busca que exigem do leitor habilidades como selecionar palavras-chave, avaliar se a informação é pertinente ou não e reconhecer elementos que sinalizam a presença de um *link* que o levará ao destino. Castro (2019) afirma que a metáfora por excelência que elucida essas operações interativas (leitor-plataformas) é a da navegação, que engloba as ideias de movimento, de destino, de retorno à origem, de risco e de pilotagem. Leitura e navegação são partes da dinâmica das plataformas *online*, porque o tempo gasto pelo leitor no ambiente plataformizado permite coleta dos dados de uso, comportamentos de consumo e outros elementos que contribuem para sustentar a datificação. “Os novos media digitais não são agentes externos que vêm causar uma rotura numa cultura inocente. Emergem do interior de contextos culturais e remoldam outros media que estão inseridos no mesmo ou em similares contextos.” (BOLTER; GRUSIN, 1999, p. 19 *apud* FURTADO, 2003, p. 31).

A “não-inocência” das plataformas está justamente nos modos como algoritmos interpelam os leitores durante as interações, segundo argumenta Castro (2019). Quando um leitor preenche a lacuna do buscador *Google* com alguma demanda ou responde aos questionamentos de *status* “No que está pensando”, do *Facebook*, e “O que está acontecendo”, do *Twitter*, ele está se colocando em um processo de leitura dirigido por um agente impessoal. Assim, esse leitor certamente vai se deparar com conteúdos selecionados e hierarquizados, sendo esse agente impessoal que subjaz à filtragem.

Nas plataformas algorítmicas, pode-se afirmar que somos compelidos a tomar como nossa uma simulação de nossa própria estrutura mental, mas feita por outrem e otimizada conforme seus interesses. O conteúdo da abordagem via algoritmos não é alheio ao receptor, mas escora-se em traços oriundos dele, reverberando de algum modo suas reações e seu *input*. A interpelação parte de alguém e regressa a ele modificada. (CASTRO, 2019, p. 5).

Muitos pesquisadores se dedicaram a explicar a complexidade das operações interativas que resultam em processos de leitura em ambiente digital, principalmente, pensando na presença intermediadora e complexa dos algoritmos de motores de busca. Manovich (2001) postula que essas operações consistem em uma externalização de nossas operações mentais. Se outrora uma imagem despertava em nós uma determinada associação mental, agora, com o uso das plataformas de busca como equipamentos de leitura, essa associação está pré-programada no sistema, na forma de um *link* que nos é disponibilizado para clicar. Somos induzidos a tomar a estrutura mental de outra pessoa (responsável pela

programação) como a nossa. Vaidhyathan (2011) cita a pesquisa de psicólogos da Universidade da Califórnia²² cuja conclusão aponta que a técnica de pesquisa do *Google* reproduz o modo como o cérebro humano armazena informações; portanto, é compreensível que acreditemos na classificação dessa plataforma de busca como mera extensão dos nossos juízos de valor coletivos, o que não passa de uma crença equivocada. Essa é uma questão também discutida por estudiosos da Ciência da Informação: quando alguém desconhece qualquer dado sobre a informação que busca, a entrega da plataforma de busca já será a primeira moldadora não só da experiência de leitura, mas de absorção de informação e, talvez, construção de conhecimento.

Na apresentação de lançamento da ferramenta *Google Instant Streaming*²³, em 2010, Sergey Brin, um dos fundadores do *Google*, declarou: “Nós queremos tornar o Google a terceira metade do seu cérebro”. Conforme mostramos em pesquisa anterior (CRUZ, 2014), os buscadores trabalham na direção de imitar o pensamento humano, o que Santaella (2005) chama de “modelo computacional da mente”. O cérebro humano e o computador funcionam como sistemas processadores de informação, com alguns aspectos que os diferenciam. Para o computador ter uma boa capacidade semântica, ainda é necessário, além da criação de processadores simbólicos, o desenvolvimento de uma linguagem natural – que envolva a habilidade de manipular símbolos – para compor a automatização competente. É nessa direção que as grandes plataformas de busca caminham e já tivemos sinalizações disso com lançamentos recentes²⁴. Santaella (2005) propõe que as operações realizadas em ambiente digital externalizam as operações da mente e as interatividades nas plataformas externalizam a essência mais profunda do dialogismo de Bakhtin e Pierce²⁵, quando estes colocam em primeiro plano a natureza coletiva dos sentidos da linguagem e o caráter eminentemente social do signo, respectivamente. Ainda nesta linha, Rieder (2020, p. 12), citando Burrell (2016), afirma que a capacidade dos computadores de executar bilhões de pontos de dados por meio de bilhões de iterações de pequenos passos de cálculo significa que eles “pensam” de maneiras que não são apenas opacas, mas potencialmente estranhas e difíceis para se encaixar

²² GRIFFITHS, Thomas L.; STEYVERS, Mark; FURL, Alana. Google and the mind: Predicting fluency with PageRank. *Psychological science*, v. 18, n. 12, p. 1069-1076, 2007.

²³ O *Google Instant* é um recurso do *Google* que visa deixar o resultado mais rápido, trabalhando em conjunto com o *Google Suggest*, os resultados são exibidos antes que o leitor termine de digitar na barra de busca.

²⁴ Em 2023, o consórcio *Open AI*, o *Google* e a *Microsoft* lançaram modelos de língua (Large Language Model – LLM) em funcionalidades separadas, como *ChatGPT*, ou acopladas aos buscadores *Google* e *Bing*.

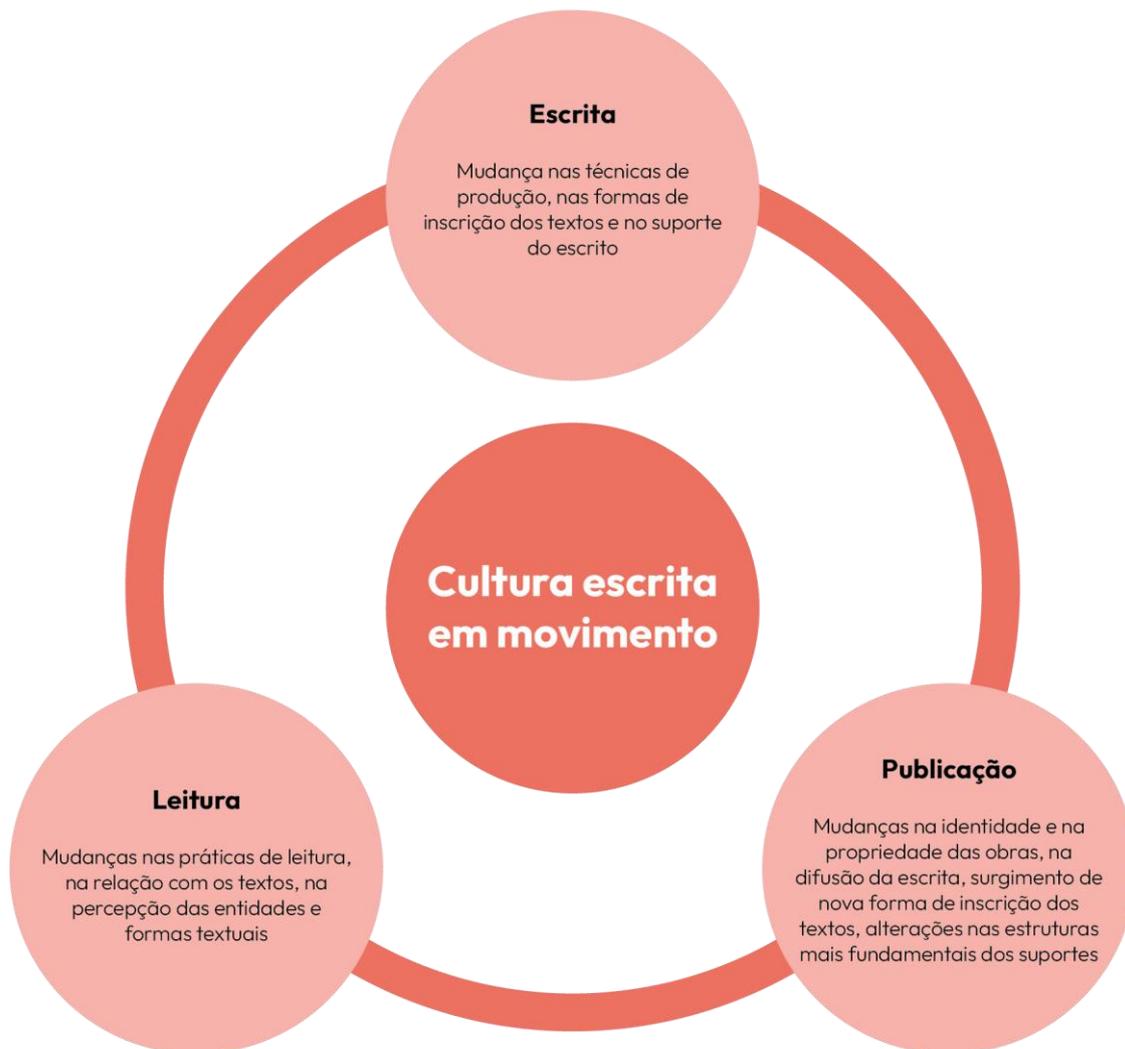
²⁵ Santaella trabalha a comparação de ideias dos dois autores partindo da afirmação de S. C. Pierce de que todo pensamento é dialógico na forma e de Mikhail Bakhtin de que não é o nosso ego que dá significado à linguagem, mas é a linguagem que dá significado ao ser humano e esse significado só pode emergir nas interações de vozes, nas trocas e interseções entre o falante e seu ouvinte.

em categorias já estabelecidas. Para o autor, técnicas como aprendizado de máquina, algoritmos de rede ou sistemas de gerenciamento de banco de dados relacional não são apenas meios poderosos para produzir e aplicar conhecimento, estabelecer preferências de valor ou controlar a prática; eles participam da própria definição do que é conhecimento.

No entanto, não é tão simples o reconhecimento das plataformas *online* como equipamentos de leitura. Chamamos *Spotify* de plataforma de *streaming* de áudio, mas também de *player* de áudio, o lugar para ouvir. Pensamos o *YouTube* como plataforma de vídeo, mas também como *player* de vídeo, o lugar para assistir. No entanto, não consideramos *Instagram*, *Facebook* e *LinkedIn*, muito menos o buscador do *Google*, como *players* de textos, ou seja, como lugares para ler. Temos o hábito de pensá-los mais como plataforma de seleção, produção e circulação de texto, atrelada a outras funções, como a de rede social. Talvez, as noções mais próximas de *players* de textos que tenhamos sejam os *e-readers* ou “*streamings* de leitura” (VECCHIO, 2022), como *Kindle* e *Kobo*, mas é uma ideia muito associada à leitura do objeto livro (em uma referência à materialidade do livro impresso).

Um dificultador em pensar na ideia de *players* de textos, ou lugares de ler, é que nem sempre conseguimos ampliar a noção de leitura. Essa ampliação poderia se dar em duas medidas: a) compreender toda a interação que fazemos em ambiente digital como um processo de leitura que engloba experimentarmos “leitura-escritura” e “leitura-navegação” (JUANALS, 2007); b) construir uma dimensão mais realista sobre o texto em ambiente digital, reforçando a ideia do ciclo “produção, circulação e leitura” como um sistema complexo. Aliás, esse tripé interage produzindo um comportamento organizado, sendo que alguns elementos textuais, como *tags* e *hashtags*, funcionam como vórtices atratores no sistema. Um sistema complexo como esse que aqui argumentamos é dotado de dinamismo, auto-organização, ressonância e não linearidade (CRUZ, 2014). Pensar esse conjunto como sistema é exatamente o que fez Ribeiro (2018), ao propor o diagrama sobre “movimentos atuais da cultura escrita”, baseado em Roger Chartier. A Figura 2 traz a ideia de dinamismo e não linearidade, detalhando as mudanças na cultura escrita a partir dos pilares “escrita, publicação e leitura”, que se assemelha ao sistema complexo argumentado em nossa pesquisa: “produção, circulação e leitura”.

Figura 2 – Movimentos atuais da cultura escrita



Fonte: RIBEIRO, 2018.

Seguimos discutindo sobre esse sistema complexo agora pensando na circulação de textos como um termo para substituir a noção de transmissão e distribuição, que vigorou em outros modelos de imperativo midiático. Conforme destaca Grohmann (2020), para pensar em circulação no contexto das plataformas, é necessário tratar de alguns eixos-chave – como codificação e decodificação, emissão e recepção – de forma menos funcionalista. Circulação é uma palavra que funciona como guarda-chuva epistemológico de onde se olha a comunicação e a linguagem e, dessa forma, afastamo-nos de uma visão linear/transmissionista e passamos a entendê-las em facetas circulares – com deslocamentos, arestas, rupturas, continuidades e conflitos. No caso de plataformas *online*, esse olhar menos funcionalista “[...] envolve gramaticalidades e lógicas algorítmicas, com normas e valores das plataformas inscritos nos rastros da circulação.” (GROHMANN, 2020, p. 3). Um exemplo interessante é pensar o

quanto a datificação influencia a circulação, porque a redesenha de forma volátil e contingente, conforme abordamos na Introdução desta pesquisa. Mas, afinal, o que significa um texto circular em plataformas *online* ou por meio delas? Tem a ver com vencer disputas de sentido tanto para alcançar públicos visados, quanto para “agradar” e ser legitimado por algoritmos, como detalharemos a seguir.

Muito se diz sobre ter presença digital, o quanto ela é importante e quando feita de forma consciente e estratégica pode se tornar um trunfo de autoridade nesse ambiente. Mas essa presença se firma, na verdade, quando uma marca, uma empresa ou uma personalidade circula seus textos, seus sentidos, suas imagens, seus produtos. A partir de leitura de teóricos dos estudos culturais, Grohmann (2020) afirma que a circulação dos produtos culturais reflete sobre as condições especificamente capitalistas de produção e consumo; portanto, circular sempre estará associado aos valores produzidos.

Nada (nem ninguém) circula indefinidamente, pois há materialidades, limites e contradições na vida social e comunicativa. Os próprios meios e suas infraestruturas possibilitam determinadas maneiras de circulação em detrimento de outras e podem sugerir – eles mesmos – formas de circular a comunicação. (GROHMANN, 2020, p. 3).

Assim, o autor propõe três dimensões para apreender a circulação. A primeira é a semiodiscursiva, que constata que os discursos se movem temporal e espacialmente a partir de suportes materiais. Os significados podem ser estabelecidos e contestados, sedimentados e ressignificados, com tentativas de controle dos modos de significar o mundo; no caso das plataformas *online*, esses feixes de sentido passam por controle algorítmico. Os discursos não acontecem no vazio, eles – como lugar de materialização e sedimentação dos sentidos – dão-se a ver na circulação. Segundo Grohmann (2020), as disputas pelo controle dos feixes de sentido acabam por se tornar “lutas por circulação” (GROHMANN, 2020, p. 7).

São signos que circulam com determinados acentos e valores e não outros [...] Deslocamentos, embates, rupturas, tensões, conflitos, reatualizações, reconhecimento de legitimidades e autoridade: algo marcado sempre por mudanças e permanências. Um lugar de onde se possa compreender as relações de poder marcadas nos processos comunicacionais: a circulação como lugar de embates. (GROHMANN, 2020 p. 5).

Algoritmos mediando a circulação de texto refletem modos de pensar de quem os programa, por isso falamos hoje, por exemplo, em preconceito, sexismo e racismo

algorítmico. A reportagem da *Revista Fapesp*, “Algoritmos parciais”²⁶, traz uma sólida discussão sobre como algoritmos podem replicar comportamentos indesejáveis porque absorvem, reproduzem e robustecem a discriminação e a intolerância vistas na sociedade nas mais variadas formas. No dossiê *Interseccionalidade e plataformas digitais*, da *Revista LIINC*, pesquisadores contribuem com o debate acerca do gênero e da raça, com elementos sobre sexualidade, corporeidades, território, classe, origem, entre outros, entrecruzando esses temas com conteúdos que circulam em plataformas *online*, em especial, plataformas de redes sociais. As pesquisas de d’Andréa (2020) também confirmam que os mecanismos de segmentação das plataformas *online* dão margem a direcionamentos de anúncios com viés claramente racista ou xenofóbico:

Em 2017, reportagem do site ProPublica revelou que o Facebook oferecia, como possibilidade de segmentação de público, categorias antissemitas como “Jew Hater” e “How to burn jews” (“odiador de judeus” e “como queimar judeus”, em uma tradução livre para o português). (ANGWIN; VARNER; TOBIN, 2017) Já o BuzzFeed descobriu que a Google permitia a associação de anúncios a expressões racistas como “black people ruin neighborhoods.” (em uma tradução livre “negros arruinam vizinhanças”). (KANTROWITZ, 2017) Após as denúncias, as plataformas afirmaram que ajustes foram feitos em seus sistemas. (D’ANDRÉA, 2020, p. 38).

Especialmente a pesquisa do *Google*, como há algoritmos alimentados por aprendizado de máquina e dados gerados por usuários humanos (defeituosos e tendenciosos), há grande vulnerabilidade em exibir os vieses. Por exemplo, em 2016, o *Google* teve que alterar suas sugestões de pesquisa (que aparecem quando um usuário começa a digitar uma consulta) para remover os resultados de preenchimento automático que apareciam quando alguém digitava “judeus são” ou “mulheres são”. Antes dessas mudanças, o algoritmo do *Google*, que usava aprendizado de máquina para completar automaticamente com frases, sugeriria terminar estas frases com a palavra “maus” ou “más”, respectivamente. (METAXA; TORRES-ECHEVERRY, 2017)

Grohmann (2020) considera classe, raça, gênero, geração, território e, enfim, as materialidades midiáticas das plataformas *online* como marcadores que atravessam a circulação de sentidos. Frisamos: a circulação de sentidos não pode se efetivar sem essas materialidades das plataformas.

A circulação de sentidos é atravessada, pois, pelos desenhos dos meios, sem esquecer que as próprias tecnologias, como mostra Vieira Pinto (2005), são frutos do trabalho humano. Ou seja, a governança e as materialidades dos dispositivos

²⁶ Disponível em: <https://revistapesquisa.fapesp.br/algoritmos-parciais-2/>. Acesso em: 14 set. 2022.

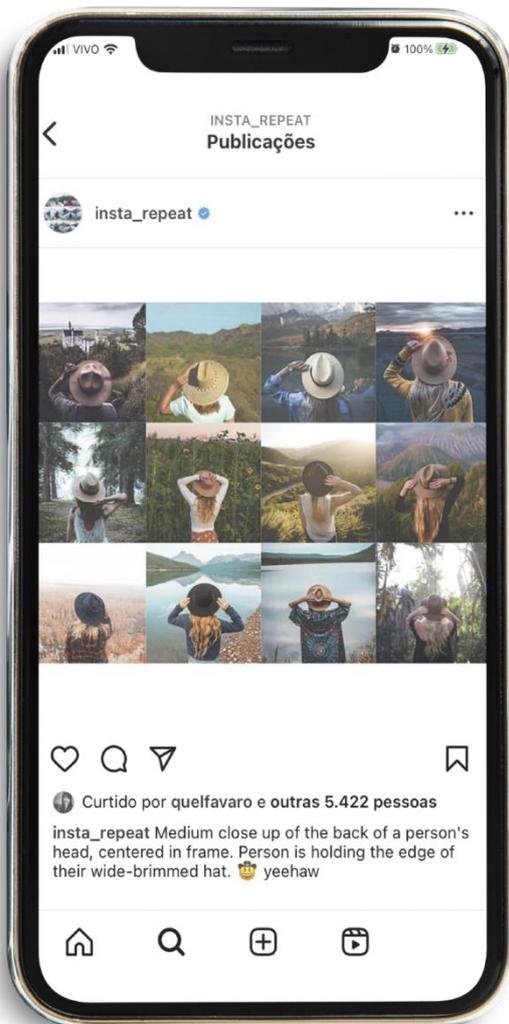
mediáticos podem impor barreiras à circulação. No caso de plataformas digitais, há circuitos e circulações de ordenações algorítmicas, que possibilitam determinados regimes de visibilidades em detrimento de outros. (GROHMANN, 2020, p. 8).

A segunda abordagem para apreender a circulação é culturalista, que reflete sobre “culturas de circulação”. Quando percebemos nas plataformas *online* algumas maneiras de produzir e consumir ideias, mercadorias e espaços, inclusive com determinados rituais, não somente midiáticos ou de consumo, é porque existe um esforço para fazer circular determinadas formas culturais. Um exemplo dessa perspectiva é a influência dos modos de produção e compartilhamentos de vídeos curtos promovida pelo *TikTok* e que passou a influenciar muito setores (KARHAWI, 2021). Fala-se em “tiktokização” da educação, da comunicação, dos conteúdos, das profissões, da política e até em “tiktokização” da vida. Issaaf Karhawi (2021) argumenta até a existência de um pânico moral, ou seja, um medo de que o *TikTok* ameace algo já estabelecido e já aceito culturalmente. A autora nos alerta, porém, para o quanto é importante compreender quem está bancando esse pânico moral: no caso do *TikTok*, esse discurso foi franqueado por concorrentes, em especial, o *Instagram*. Aliás, criadores de conteúdos e leitores do *Instagram* se uniram em um movimento²⁷, em 2021 e 2022, reclamando contra a “tiktokização” do *Instagram*, acusando a plataforma de valorizar modos de produção semelhantes à “rede vizinha” e, portanto, obrigando produtores a um contingenciamento excessivo para entregas de conteúdos aos públicos a partir da legitimação algorítmica.

Mas, pensando na abordagem de influência cultural, o próprio *Instagram* é responsável pelo movimento de criação de “ambientes instagramáveis” que incentiva a existência de espaços físicos (fora do digital) para fotografar e compartilhar digitalmente. É também uma plataforma que influenciou modos estéticos e plásticos de publicizarmos a vida pelo *feed*, pois funciona numa lógica de “*templatability*” (LEAVER; HIGHFIELD; ABIDIN, 2020). Se um formato e uma linguagem textual dão certo com algum produtor de conteúdo, isso se alastra como um *template* de sucesso para todos os outros.

²⁷ Essas mobilizações ocorreram nos momentos em que o *Instagram* anunciou mudanças nos algoritmos de seleção de conteúdo. Em julho de 2022, por exemplo, usuários fizeram campanha por “Instagram original”, e os responsáveis pela plataforma se manifestaram e acabaram recuando em algumas mudanças algorítmicas. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2022/07/15/nem-views-nem-curtidas-pessoas-relatam-insatisfacao-com-o-instagram.htm> e <https://canaltech.com.br/redes-sociais/usuarios-fazem-campanha-por-instagram-original-e-rede-se-manifesta-221598/>. Acesso em: 16 ago. 2022.

Figura 3 – Perfil do projeto *Insta Repeat* no *Instagram*

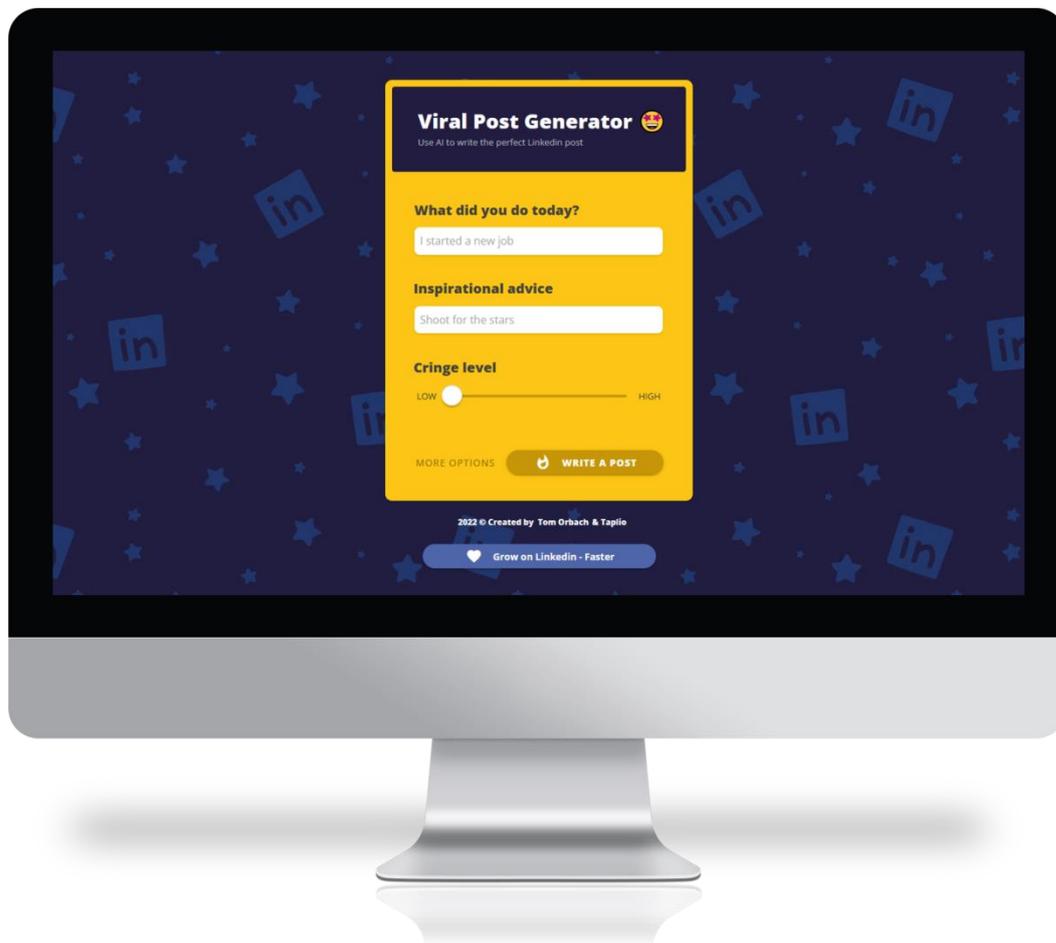


Fonte: *Instagram* (https://www.instagram.com/insta_repeat/), 2022.

O projeto *Insta Repeat*²⁸, por exemplo, critica essa dinâmica de imitação, replicação ou *déjàvu vibes*, ao expor imagens iguais ou muito parecidas, mas coletadas em perfis diferentes. Essa lógica de *template* também ocorre com o texto verbal, em *sites* e *blogs*, como mostraremos no Capítulo 5 desta tese. Outro exemplo é a sátira promovida pelo projeto *Viral Post Generator*²⁹, aos *posts* inspiracionais, feitos no *LinkedIn* de forma padronizada, que “passam do ponto” e acabam virando piada. Os responsáveis pelo projeto criaram uma ferramenta de IA que, a partir de algumas entradas, elabora um post de *LinkedIn*, em padrões altamente motivacionais, pronto para ser postado.

²⁸ Disponível em https://www.instagram.com/insta_repeat/. Acesso em 19 ago 2022.

²⁹ Disponível em <https://viralpostgenerator.com/>. Acesso em 31 ago 2022.

Figura 4 – Projeto *Viral Post Generator*

Fonte: Disponível em: <https://viralpostgenerator.com/>. Acesso em: 20 mar. 2022.

Voltemos às dimensões para apreender a circulação, propostas por Grohmann (2020). A terceira abordagem é nomeada “circulação comunicacional do capital”. É a partir do transporte e da comunicação que as mercadorias chegam aos sujeitos nos diferentes lugares e se colocam como essenciais para a concretização da produção material e da efetivação do mais-valor. No caso das plataformas *online*, elas não servem somente para se comunicar, mas são o meio de produção. Para Grohmann (2020), as tecnologias são, ao mesmo tempo, de comunicação, de produção e de circulação. O seu potencial comunicativo é fartamente explorado na pesquisa em comunicação, mas há uma relativa ausência de sua imbricação com as dimensões da produção e da circulação.

As tecnologias de circulação – assim como os próprios meios de comunicação – auxiliam a “encurtar o tempo de circulação do capital na produção e no mercado” (HARVEY, 2018, p. 99), ou seja, tem como função auxiliar a aceleração da própria circulação do capital, auxiliando, inclusive, na disseminação de dados. (GROHMANN, 2020, p. 9).

Vamos tomar emprestada a pergunta de Grohmann (2020, p. 9): “[...] como se produz valor a partir da circulação dos produtos culturais/midiáticos enquanto mercadorias por meio de plataformas digitais?” A partir desse questionamento, o autor argumenta a relação entre circulação de mercadorias culturais e circulação de capital que, atualmente, abastecem a lógica do “capitalismo de plataforma” (SRNICEK, 2016). A financeirização, a mediatização e a datificação formam o tripé dessa lógica de circulação em que é mais valorizado circular do que efetivamente os sentidos carregados por essa produção. É uma ideia de volume e aceleração da política de circulação em detrimento dos próprios sentidos colocados em disputa. Em suma, quando colocamos o foco em conteúdos feitos com a intenção de performar, parece ser mais importante o modo como os leitores chegam a esses textos do que a informação inscrita neles. Aqui constatamos uma grave desvalorização de processos de apuração, curadoria, criação, autoria e repertório do autor, capacidade de articulação textual em detrimento, por exemplo, de técnicas que fazem um texto se propagar.

Com tudo que demonstramos sobre circulação, cabe retomar o fracasso de discursos otimistas e dotados de imaginários neoliberais que, no início da *Web*, acreditaram em uma circulação livre das informações. Ochigame (2021a) afirma que isso sempre foi uma fantasia, uma vez que vários tipos de máquinas, estruturados por condições político-econômicas, sempre filtram quais as informações circulam e para quem. Para o autor, concepções tecno-utópicas de liberdade de informação nunca acertam.

Após elucidarmos o sistema complexo “produção, circulação e leitura” em plataformas *online*, percebendo as mutações das práticas de leitura, as novas modalidades de publicação e a redefinição da identidade das textualidades, vamos propor uma abordagem para o que nomeamos de “texto plataformizado”. Ribeiro (2018) demonstra as dificuldades em definir texto e apresentar seus componentes, principalmente quando entram em cena as tecnologias digitais. Inspirando-nos, no nome do livro da autora – *Escrever, hoje: palavras, imagem e tecnologias digitais na educação*, poderíamos nos perguntar: o que é escrever para *Web* hoje? Não há uma resposta simples e fechada, nem pretendemos apresentar qualquer normatividade (o que dever ser para alcançar sucesso), mas propomos a discussão a partir de dois pilares: 1) planejamento e execução de uma produção mais hipertextual possível e 2) preparação da performance do texto.

Percebemos que elementos “extratexto” – como a mediação dos motores de busca e algoritmos – estão incorporados como conhecimento e como impulsionadores de ações na rotina da produção e edição de texto em ambiente digital. O texto que se produz não é apenas

fragmentado, não linear, hiperconectado, ele é feito, quase sempre, para atender a requisitos de desempenho mensuráveis. A todos os conceitos de hipertexto, construídos por décadas na literatura – Landow (2006), Ribeiro (2006), Marcuschi (1999), Coscarelli (2009) e outros –, acrescenta-se agora a perspectiva da performance, que, como veremos, é uma característica imposta pela dinâmica de circulação em plataformas *online*.

O hipertexto é uma grande metáfora para a rede semântica que se desencadeia na *Web*, à luz do paradigma técnico, em que um conjunto de nós são ligados por conexões clicáveis, e do paradigma funcional, em que o hipertexto serve como estrutura para a organização de conhecimentos ou dados, a aquisição de informações e a comunicação. O papel do hipertexto não é só reunir textos, mas as redes de associações capazes de vincular pessoas e ideias na elaboração coletiva de um “hiperssignificado”.

Isso nos faz pensar que a leitura hipertextual transcende as telas de computador, considerando que os hipertextos já existiam anteriormente aos meios digitais: “Acredito que isso se justifique pelo fato de que todo texto seja um hipertexto e que toda leitura seja um processo hipertextual.” (COSCARELLI, 2009, p. 151).

A mente pode ser vista como um grande hipertexto que é muitas vezes espelhado na linguagem e, logo, nos textos que sempre produzimos mesmo muito antes do advento da escrita. Isso não significa que a informática não tenha modificado em nada o universo dos textos da escrita e da leitura. [...] Com o texto digital, usamos menos o lápis e a borracha. Escrever é um ato diferente, mesmo da datilografia, pois podemos cortar, colar, editar, inserir tabelas e imagens com muito mais facilidade que antes e, além disso, hoje podemos salvar, inserir sons e animações, além de usar links que podem nos levar diretamente a outros textos, a filmes, músicas ou imagens [...] Essa mudança, no entanto, não deve ser vista como uma substituição das habilidades que o leitor precisa ter para lidar com o texto impresso, por outras que serão exclusividade do meio digital, mas uma ampliação daquelas. (COSCARELLI, 2009, p. 152).

Por tudo isso, Coscarelli (2009) propõe que hipertextos em ambientes digitais não sejam vistos como uma ruptura em relação ao impresso, mas uma continuidade. Para a autora, talvez não seja o caso de pensarmos como a textualidade do hipertexto digital seja diferente da textualidade do texto impresso, mas em como outras linguagens se incorporam a ele (o vídeo, a animação e o som, por exemplo). Nessa direção, está nossa proposta de planejar uma produção o mais hipertextualmente possível como solução para a escrita na *Web* hoje. Uma oferta rica em rotas de navegação tende a oferecer mais caminhos e possibilidades de multisemioses ao leitor, além de atender a lógica de conectividade das plataformas *online*, considerando que elas são planejadas para manter leitores em imersão contínua.

Por fim, vale destacar que hipertexto exige busca, navegação, movimento de coautoria, habilidades que, segundo Coscarelli (2009), não costumamos desenvolver em nossos leitores em formação. “Não ensinamos os estudantes a navegar nos textos impressos, buscando informações no jornal, nas revistas, nas enciclopédias nem nas bibliotecas. Normalmente entregamos a eles o texto que precisam ler.” (COSCARELLI, 2009, p. 152). Nesse contexto, as plataformas de busca se firmam mesmo como oráculos modernos, como argumentaremos no Capítulo 4, pois selecionam com critérios próprios a informação ranqueada, diante da inabilidade das pessoas em encontrar e da impossibilidade perante o *tsunami* de documentos na internet. Essa retórica de ordenação da informação baseada no caos, complexidade e desorientação que se impôs no fim do século XX e início do XXI facilita o argumento de necessidade infraestrutural de plataformas de busca.

Ademais às questões da hipertextualidade, percebemos a crescente consciência de produtores de texto em ambiente digital sobre mecanismos de desempenho na circulação. O processo de edição dá o tom da performance, por meio dele um texto será reconfigurado para atender a demandas específicas de cada plataforma de publicação. Por isso, pesquisar e caracterizar a produção do texto plataformizado não diz respeito a identificar apenas a composição, mas se refere à reflexão do quanto eles respondem a lógicas mercadológica e sociais atualmente impostas por provedores da informação que são responsáveis por materializar, arquivar e circular textos.

Defendemos que a ideia de performance está conectada à noção de performance de produto que vem da inovação disruptiva, responsável por gerar “coisas” sempre melhores ou superiores que as antecedentes. O termo está na mesma órbita da ideia de progresso, em perspectiva positivista, na qual o que é novo é melhor do que o criado anteriormente. Outra palavra que vem a reboque é otimização, aliás, muito ligada ao entendimento de desempenho e qualidade para inovação, que bebe nas fontes dos negócios neoliberais. Problematisamos esses termos à luz de Benjamin (2006) e Cupani (2011), que vão criticar a retórica do progresso construída pela burguesia e relativizar a inserção da performance em tudo, afinal, nem sempre o que é progresso tecnológico se reverte em progresso para a humanidade. Atualmente, a facúndia do empreendedorismo e da inovação fagocita todos os setores, resumindo os trabalhos de produção cultural à performance por retorno financeiro. Ou seja, é uma conversa muito casada ao funcionamento das *Big Five*, pois elas praticam uma “[...] governança algorítmica que viceja na sociedade neoliberal” (CASTRO, 2019, p. 2).

A performance de textos na circulação de textos se processa, necessariamente, pelo domínio da tecnologia e pela eficiência, seja do ponto de vista técnico, seja do ponto de

vista das ações profissionais. Vejamos dois exemplos de como isso se configura em setores diferentes: no jornalismo e no mercado editorial.

A produção jornalística passou a se pautar não apenas para o leitor final e os interesses públicos, mas pensando nas máquinas que irão catalogar e organizar esse conteúdo. Os jornalistas sabem que se ater a orientar os mecanismos de busca e indexação pode fazer uma grande diferença no impacto e alcance de uma reportagem. “Esse entendimento transcende a atenção em se posicionar perante o algoritmo do Google, na compreensão da relevância em recuperar os dados, compartilhá-los e criar uma nova maneira de pensar e produzir conteúdo a partir desse ambiente.” (CARVALHO, 2017, p. 83). Aliás, Janaína Kalsing (2021) provou a existência de um “jornalista metrificado”, profissional que monitora de forma instantânea, incessante e rotineira o comportamento da audiência e, a partir disso, mobiliza esforços de reportagem e edição para alcançar o maior número possível de leitores. Essas métricas performáticas ressignificam critérios, pautas e edição textual.

Jornalistas e outros produtores de texto incorporam em seus processos criativos uma ideia além da escrita: o que meu leitor buscaria no *Google* para encontrar o conteúdo sobre o qual estou escrevendo? E, assim, trabalham equilibrando processo criativo, demandas de públicos e elegibilidade aos algoritmos. Os profissionais monitoram métricas em tempo real, por meio de ferramentas de *web analytics*, em que é possível medir performance de conteúdos, e compreender as contingências para a produção. Um exemplo desses contingenciamentos está nos estudos de Anderson (2014), que mostram o jornalismo se alimentando da cultura do clique; da cultura do Vale do Silício orientada por dados, métrica e obcecada por inovação; da cultura de medição, quantificação e viralidade. Christofolletti e Vieira (2015) se apropriam desses conceitos de Anderson (2014), mostrando *sites* brasileiros que se alimentam ou que resistem a esse foco nos dados brutos de audiência e visibilidade em ambiente digital.

No artigo *Eu, uma escrava de robôs*³⁰, a jornalista e diretora executiva da *Agência Pública*, Natália Viana, dialoga com essas questões do jornalismo atravessado por técnicas de *Search Engine Optimization* (SEO – otimização de *sites*) e pelas métricas. Em meio aos debates sobre o futuro das inteligências artificiais, a jornalista constata uma realidade do presente: “Queridos leitores, nós já trabalhamos para robôs. O tempo todo”. E a partir daí, ela analisa o quanto as plataformas *online*, em especial as de busca, funcionam como antigamente as capas dos jornais: “[...] grande mural que chama a atenção do público sobre as histórias que

³⁰ Disponível em: https://mailchi.mp/apublica.org/newsletter-xeque-na-democracia-030?e=9647d7e595&utm_source=substack&utm_medium=email. Acesso em: 20 mar. 2023

são importantes num determinado dia”. A jornalista descreve, com aplicação à realidade da *Agência Pública*, a dinâmica de mediação dos algoritmos que fazem a curadoria de conteúdo na *Web* seguir a tríade “pessoa-robô-capital”, ficando os jornalistas atravessados por essa dinâmica:

No meio disso ficamos nós, jornalistas, que perdemos a tranquilidade de focar apenas na qualidade do produto que a gente entrega. Temos que passar cada vez mais tempo estudando e observando como funciona esse ou aquele algoritmo. Criamos equipes inteiras que trabalham para robôs, que passam seus dias tentando entender por que uma frase funcionou e outra não funcionou para agradar aos robôs que fazem a mediação entre o nosso conteúdo e o público. Eu como diretora executiva, fui virando um pouco escrava dos robôs, dedicando parte das minhas horas a tentar entendê-los, acompanhar cada nova mudança não anunciada, pensar junto com a equipe soluções para a tarefa de levar o nosso jornalismo para mais longe. (VIANA, 2023).

Apesar de tudo que argumentamos aqui sobre performance e jornalismo, demonstrando o esforço contingenciado dos jornalistas para atender as demandas das plataformas de busca, o *Google* se posiciona discursivamente como grande apoiador do jornalismo por “[...] ajudar os leitores a acessar informações relevantes”³¹. A empresa reforçou esse discurso a partir de 2022, quando os debates sobre regulação das plataformas engrossaram a ideia das *Big Five* como *gatekeepers*. Além disso, explicitou os programas de apoio financeiro a veículos brasileiros, como *Google Destaques*³² e *Programa de Inovação de Storytelling*³³.

Hoje, a Busca do Google, por exemplo, envia brasileiros para sites de notícias mais de dois bilhões de vezes por mês – de graça. Isso dá às empresas jornalísticas a oportunidade de ganhar dinheiro e aumentar seus negócios e audiência, mostrando anúncios e ofertas que convertem as pessoas em novos assinantes. [...] Nossos esforços em apoiar as empresas brasileiras e os veículos jornalísticos são parte do nosso compromisso já de longa data em promover um ecossistema de notícias saudável e diversificado no Brasil. (GOOGLE, 2022e).

Segundo Kalsing (2021), com o atravessamento das plataformas de busca e a cultura das métricas, há uma transferência de responsabilidades, com o jornalista se corresponsabilizando pelo sucesso ou fracasso comercial do veículo de imprensa. Os profissionais, à procura de seguir e superar metas, colocam-se como empreendedores de si.

³¹ Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/apoiando-o-jornalismo/>. Acesso em: 06 jun. 2023.

³² Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/destaques-nosso-investimento-de-us-1/>. Acesso em: 05 jun. 2023.

³³ Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/web-stories-novas-formas-de-fortalecer-os-produtores-de-conteudo-jornalístico/>. Acesso em: 05 jun. 2023.

Movidos pela pressão por rendimento, eles não almejam apenas que seus textos sejam lidos, comentados e compartilhados, preocupam-se também com a rentabilidade do negócio para o qual trabalham, sob pena de comprometerem seus próprios empregos. Várias são as consequências dessa lógica de “mostrar-se relevante”, entre elas, a competitividade acirrada dentro das redações de jornal.

Além do jornalismo, vejamos o exemplo do mercado editorial, em especial, a produção de livros em plataformas digitais de autopublicação, estudada por Vecchio (2022). A pesquisadora analisou o imaginário do empreendedorismo que emerge no discurso de autores-empresendedores e que subsidia ideologicamente o capitalismo de plataforma. Esses escritores se submetem às lógicas da plataforma de autopublicação, produzindo livros com potencial performativo (mercadoria cultural contingente), ou investem na divulgação de suas obras para alcançar métricas de leitura, ou defendem o discurso da supremacia da tecnologia e do empreendedorismo, ou não se percebem como trabalhadores precarizados. Esses indivíduos, assim como os jornalistas metrificados, são supostamente capazes de alcançar sucesso por esforços próprios, fortalecendo o discurso da meritocracia baseada em estatísticas, *rankings*, *feedbacks* das ferramentas de *analytics*.

Kalsing (2021) e Vecchio (2022) demonstraram nas pesquisas que racionalidade neoliberal empresenedora, entranhada na lógica produtiva, torna esses jornalistas e autores profissionais automatizados e competitivos, dotados, em certa medida, de uma consciência da relação da produção com a performance algorítmica. Dardot e Laval (2016) vão dar o nome de “sujeito neoliberal” a esses empresenedores de si, que vivem buscando maximizar seus resultados com foco prioritário no “dispositivo de desempenho/gozo”. Há uma ambivalência desses profissionais, em diálogo com a própria ambivalência da tecnologia, que já argumentamos anteriormente. Os produtores de texto em ambiente digital também se beneficiam de jogar o jogo das plataformas, tornando a ausência delas um cenário distópico por vontade de todos os lados. Ou melhor, não há o lado dos provedores de conteúdo e o lado dos produtores de conteúdo de forma maniqueísta. Estão todos do mesmo lado, principalmente aqueles produtores que dominam o jogo e o usam taticamente. Muitos deles fazem o jogo sabendo que isso é obediência, mas sobrevivência. A submissão à lógica algorítmica pode ser vista como um modo de sobreviver, estrategicamente, porque centenas e milhares de negócios estão pendurados e dependentes das plataformas *online*.

Percebemos que o processo de apuração, concepção, escrita, criação e edição de textos se deslocou de uma economia de produção, baseada no autor, para uma economia fundada no leitor-consumidor que gera demandas contingentes e condicionadas por

algoritmos. O foco no leitor não é uma exclusividade da era das plataformas *online*, mas o atravessamento da lógica algorítmica muda o ritmo desse contingenciamento. Não se pode esquecer que esses algoritmos são criados para fins lucrativos. Ochigame (2021a) revela que a indústria da publicidade governa a maior parte da pesquisa e desenvolvimento no campo da recuperação da informação e ainda explica:

Cientistas da computação e engenheiros costumam medir a “relevância” de resultados potenciais e testar o “desempenho” dos algoritmos candidatos de acordo com a avaliação de *benchmarks* e a validação de dados ditados pelas prioridades da indústria. Os sistemas predominantes são projetados para maximizar as receitas de anúncios e métricas de “engajamento”, como “taxas de cliques”. Consequentemente, esses sistemas tendem a promover conteúdo que já é popular ou semelhante ao que usuários já viram ou curtiram antes. Se as previsões de popularidade e similaridade são baseados em correlação simples e análise de regressão ou em análise complexa de modelos de aprendizado de máquina, os resultados tendem a ser previsíveis e semelhantes. (OCHIGAME, 2021a, p. 108, tradução nossa)³⁴.

A performance em plataformas *online* se alimenta da repetição. Castro (2019) nos lembra que resultados performativos não são atos singulares, mas reiterações de normas. É por isso que a ideia de robôs interferindo na circulação e promoção de conteúdo e nas fábricas de cliques que impulsionam conteúdos, funciona bem nas plataformas *online*, pois são combustíveis para a repetição.

A performatividade, de qualquer forma, tende a confundir-se com a rotina de funcionamento das plataformas algorítmicas. Não se trata de um ato único, mas do saldo acumulado dos atos de interpelação ao qual ele é submetido. O aspecto iterativo da interpelação é tonificado por sua recursividade: quanto mais eficiente ela se revela, mais é reproduzida. E, independentemente dos ajustes cada nova interpelação, há elementos que se repetem. Daí provém o efeito performativo da interpelação, que consolida as operações de perfilamento: quanto mais se reproduz, maior é sua eficácia na construção de perfis. As plataformas algorítmicas não são, pois, meros veículos para a expressão de uma identidade preexistente do usuário; essa identidade é construída na própria plataforma, via interpelação, perfilamento e performatividade. (CASTRO, 2019, p. 17).

Enfim, performatividade no contexto da plataformação remete a performance na acepção de desempenho, que atravessa todo o ciclo “produção, circulação e leitura”. Portanto, preparar um texto para performar significa dar conta de responder, ao mesmo tempo, a uma

³⁴ Tradução nossa para: “Computer scientists and engineers often measure the “relevance” of potential results and test the ‘performance’ of candidate algorithms according to evaluation benchmarks and validation data sets dictated by industry priorities. The predominant systems are designed to maximize ad revenues and ‘engagement’ metrics such as ‘click-through-rates’. Consequently, these systems tend to promote content that is already popular or similar to what users have seen or liked before. Whether the predictions of popularity and similarity are based on simple correlation and regression analysis or on complex machine learning models, the results tend to be predictable and like-minded.”

demanda por parte do algoritmo e a uma demanda por parte de leitores interpelados pelo algoritmo. O ciclo é completamente impelido por resultados: quanto maior o nível de atividade, mais dados elas acumulam e maior seu potencial de lucratividade. Os algoritmos de plataformas *online* são projetados, inclusive, para um certo padrão de eficiência que tem relação com esse ciclo repetitivo, pois “agem” a partir de uma rede de ações sobre ações, conforme esclarece Gillespie (2016):

Para projetar algoritmos com eficiência que atinjam um objetivo alvo (em vez de alcançar uma resposta conhecida), os algoritmos são “treinados” em um corpus de dados conhecidos. Esses dados foram certificados de alguma forma, seja pelos designers ou por práticas de usuários anteriores: esta foto é de um rosto humano, esta foto não é; este resultado de pesquisa foi selecionado por muitos usuários em resposta a esta consulta, este não. O algoritmo é, então, executado nesses dados para que ele possa “aprender” a emparelhar consultas e resultados considerados satisfatórios no passado, ou “aprender” a distinguir imagens com ou sem rostos. (4 n.p., tradução nossa)³⁵.

Por tudo isso, é possível construir um novo entendimento de texto. Compreendemos o texto plataformizado como imanente e contingente, que se constrói de modo hipertextual pela natureza do meio e do processo de leitura e também em atenção ao desempenho para a visibilidade. E quando nos referimos à visibilidade, destacamos que a conquista dela não é um caminho desprezioso ou intuitivo, mas uma modulação de materialidades das plataformas e dos textos para performar. Para complementar essa ideia, trazemos a definição de Thompson (2005) sobre “visibilidade mediada”. No texto *The New Visibility*, o autor faz uma releitura de sua obra *A mídia e a modernidade: uma teoria social da mídia* (THOMPSON, 1998), refletindo sobre quanto o desenvolvimento das mídias comunicacionais trouxe novas formas de visibilidade:

Nessa nova forma de visibilidade mediada, o campo da visão não está mais restrito às características espaciais e temporais do aqui e agora, ao invés disso é moldada pelas propriedades distintivas das mídias comunicacionais, por uma gama de aspectos sociais e técnicos (como angulações de câmera, processos de edição e pelos interesses e prioridades organizacionais) e por novos tipos de interação que as mídias possibilitam. Ela é moldada também pelo fato de que, na maioria das mídias comunicacionais, a visualidade não é uma dimensão sensorial isolada, mas vem

³⁵ Tradução nossa para: “This brings us to the second value-laden element around the algorithm. To efficiently design algorithms that achieve a target goal (rather than reaching a known answer), algorithms are ‘trained’ on a corpus of known data. This data has been in some way certified, either by the designers or by past user practices: this photo is of a human face, this photo is not; this search result has been selected by many users in response to this query, this one has not. The algorithm is then run on this data so that it may ‘learn’ to pair queries and results found satisfactory in the past, or to distinguish images with faces from images without.”

geralmente acompanhada pela palavra falada ou escrita – trata-se do audiovisual ou do texto-visual. (THOMPSON, 2005, p. 35)³⁶.

O texto plataformizado, portanto, movimenta-se na complexidade do sistema “produção, circulação e leitura”. Fica claro como a plataformização se deu perante promessa das empresas provedoras de que se configurariam um empoderamento da produção textual e a recuperação de dados para construção de conhecimento coletivo. Na primeira promessa, plataformas cederiam estrutura tecnológica para as pessoas se conectarem, ampliarem seus acessos e se comunicarem com mais facilidade. Na segunda promessa, dariam condições para reunir todos os históricos de interações, a fim de proporcionar entendimentos globais que poderiam sugerir inteligência coletiva. No entanto, o que aconteceu, em alguma medida, foi uma apropriação, por parte desses provedores, de dados externos para seus próprios bancos de dados, traço fundamental para o sucesso de seu modelo econômico, além de uma profunda transformação no sistema “produção, circulação e leitura”.

2.2 CULTURA DIGITAL PÓS-PLATAFORMAS

A discussão da reconfiguração de práticas de produção de texto nos remete ao debate levantado por Ribeiro (2018, p. 12), quando relativiza os termos cultura escrita, cultura impressa e cultura digital. O primeiro conceito serve como guarda-chuva: “[...] uma cultura baseada na palavra, no texto, em algum tipo de código, alfabético ou não, inscrita em algum material, que propicia e provoca certas práticas [...]”. Já os outros dois conceitos são parte da cultura escrita, porém ancorados individualmente em suas materialidades, modulações, formatos e letramentos, ou seja, em elementos de caráter técnico e tecnológico que as diferenciam. Destaca-se o fato de que, para Ribeiro (2018, p. 12), a cultura escrita “[...] conforma-se às contingências, às práticas sociais, etc.”, como discutimos anteriormente ao construir a noção de texto plataformizado.

Vamos adentrar a cultura digital e, sem impor uma visão reducionista e considerando todas as grandes revoluções da história da escrita e da leitura (CHARTIER,

³⁶ Tradução nossa para: “In this new form of mediated visibility, the field of vision is no longer constrained by the spatial and temporal properties of the here and now but is shaped, instead, by the distinctive properties of communication media, by a range of social and technical considerations (such as camera angles, editing processes and organizational interests and priorities) and by the new types of interaction that these media make possible. It is also shaped by the fact that, in most communication media, the visual is not an isolated sensory dimension but is usually accompanied by the spoken or written word – it is the audio-visual or the textual-visual. [...] We can bring out the significance of this new form of visibility by focusing briefly on one area where its implications were particularly profound – the complex and changing relations between visibility and political power.”

1999)³⁷, temos a impressão de que a atual reconfiguração de práticas na produção de texto é muito radical, porque nos leva a um cenário de simbiose quase irreversível. Podemos pensar em uma outra cultura digital, talvez pós-plataformas, pois aquela cultura digital baseada na emergência do ciberespaço e dos hipertextos não é suficiente para olhar criticamente para o que está acontecendo. É preciso considerar a interferência dos meios e suas materialidades. Além disso, a ideia de pós-plataforma se alinha às noções de pós-verdade³⁸, que emergiu a partir de um olhar crítico para as implicações do atravessamento de algoritmos e plataforma em nossa vida social.

Reconhecendo o legado de todos e todas que estudaram as situações comunicativas de nosso tempo, identificando expansões de linguagens, potencialização do compartilhamento, forte cooperação, apropriação de bens simbólicos, reconfiguração de formatos midiáticos, diluição de polos emissor/receptor – como nas definições de cultura da participação (JENKINS, 2008), cultura da conexão (JENKINS, 2014), cultura da convergência (JENKINS, 2008), cultura dos acessos (JUANALS, 2007), cibercultura (LÉVY, 2010) –, afirmamos ser o termo “cultura digital” aquele que engloba todas as práticas sociais e demarca melhor “contiguidades e tensões” (RIBEIRO, 2018, p. 11). Por isso, faz sentido falar em cultura digital pós-plataformas para guiar uma visão com mais enfoque na materialidade.

Rogers (2016) argumenta que estudiosos do ciberespaço – apesar de terem feito pesquisas essenciais ao entendimento da internet – geraram uma visão de desespacialização intimamente ligada ao termo “virtual”, em que haveria uma libertação dos espaços dos Estados-Nação e suas formas divergentes de tratar os fluxos de informação. O discurso da ubiquidade que esvazia possibilidade de compreender as interações não se sustenta, pois é possível seguir rastros dos usuários e seguir as plataformas *online*. Atualmente, sabemos que a geolocalização, por exemplo, é um dos dados mais potentes usados pelas plataformas. Vejamos o caso do *Google*, que usa os endereços de *Internet Protocol* (IP)³⁹ ou a localização do Sistema de Posicionamento Global (GPS, na sigla em inglês) como critério de entrega de respostas no buscador. Se um leitor procurar por “lasanha” na Itália, receberá respostas

³⁷ Roger Chartier relativiza as mudanças técnicas, morfológicas e materiais nos textos do século XX, comparando a revolução digital com outras grandes revoluções da história da escrita e da leitura, como a invenção do códex e da prensa.

³⁸ Conceito que denota circunstâncias em que os fatos objetivos são menos influentes na formação da opinião pública do que os apelos à emoção e às crenças. No mundo pós-verdade, os fatos já não importam e a verdade parece irrelevante, especialmente se for contra as crenças das pessoas, e há inúmeros exemplos deste comportamento, especialmente nos alinhamentos políticos partidários de eleitores sacramentados pela circulação de conteúdos em plataformas online.

³⁹ Um endereço de Protocolo da Internet, do inglês Internet Protocol (IP) address, é um rótulo numérico atribuído a cada dispositivo conectado a uma rede de computadores que utiliza o Protocolo de Internet para comunicação. Disponível em: https://pt.wikipedia.org/wiki/Endere%C3%A7o_IP. Acesso em: 06 jun. 2023.

completamente diferentes de alguém que faz a mesma busca no Brasil, porque o motor de busca identifica o local de onde surge a pergunta e associa a informações contextuais (nesse caso, gastronômicas ou culturais), que o ajudam a modular as possibilidades de respostas.

Quando compreendemos as interações digitais de forma mais materializada, percebemos que nem todas as pessoas vão se apropriar da mesma forma das funcionalidades disponíveis nesses ambientes. Como argumenta d'Andréa (2020), são muitos os usos táticos, lúdicos ou políticos possíveis nas diferentes plataformas. Gillespie (2018) recupera a noção de domesticação, de Silverstone (1994), sugerindo que à medida que as tecnologias entram em nossas casas, transformamo-nas em propriedade, incorporando-as em nossas rotinas e acrescentando significados que os desenvolvedores não puderam antecipar, como o exemplo de plataformas de busca:

Algoritmos de informação pública certamente são importantes para o modo como os usuários encontram informações, comunicam-se uns com os outros e percebem o mundo ao seu redor. Mas, mais do que isso, os usuários expressam preferências por suas ferramentas de busca favoritas, opinam sobre as recomendações de um site como sendo incômodas, intuitivas ou precisas. Alguns usuários confiam em uma determinada ferramenta, enquanto outros desconfiam dela, usando-a com cuidado ou nem a utilizando de todo. (GILLESPIE, 2018, p. 112).

Para o autor, os usuários, perante as assimetrias de poder da coleta de dados, ordenação da informação e vigilância *online*, desenvolveram uma série de táticas para escapar de lógicas pré-determinadas dos algoritmos, portanto, não ficam somente sob influência dessas tecnologias e, sim, as reconstróem. Gillespie (2018) considera que os algoritmos não são apenas o que seus *designers* fazem deles, ou o que eles fazem a partir da informação que processam, mas o que nós fazemos deles dia após dia. Vale ressaltar que o poder efetivo de modelar os algoritmos ainda está nas mãos das grandes empresas provedoras e qualquer ato mais subversivo exigiria redesenhar essas lógicas do zero, como Ochigame (2021a) vai sugerir ao defender a remodelação de motores de busca, que exemplificaremos no Capítulo 5.

Para complementar, Gillespie (2018) versa sobre a forma como usuários de tecnologias reconfiguram suas práticas para se adequar aos algoritmos dos quais dependem e como podem transformar algoritmos em espaços de disputa política, às vezes até mesmo para questionar as políticas do próprio algoritmo. O autor nomeia essa dinâmica de “entrelaçamento com a prática”. No caso dos profissionais do texto, vamos verificar que muitos produtores já dominavam certas habilidades e atitudes necessárias para a produção em ambiente digital, mas acabam por adquirir novas competências justamente para serem capazes de entrelaçar, fazer apropriações criativas, recriar e subverter ou repetir um modo quase

vicioso com vistas à performar. É por isso que só podemos saber como os algoritmos tornam a vida cotidiana diferente observando seu trabalho no mundo sob diferentes condições.

Quando os usuários empregam um algoritmo, digamos para jogar ou trabalhar, eles não estão simplesmente jogando ou trabalhando em conjunto com o algoritmo, em vez disso, eles estão “aprendendo, internalizando e se tornando íntimos” (Galloway, 2006, p. 90); como eles se comportam é sutilmente remodelado através do engajamento, mas ao mesmo tempo, o que o algoritmo faz é condicional à entrada que recebe do utilizador. (KITCHIN, 2016, p. 12).

Vejam como a pesquisadora Joana Ziller (2012; 2019) compreendeu a visibilidade e invisibilidade na circulação de vídeos no *YouTube*, a partir dessa noção dos usos. A autora partiu das concepções de tática e estratégia, criadas por Michel de Certeau (2009) ao tratar do consumo de cultura e de sua inserção no cotidiano. As táticas acontecem a partir de apropriações das regras por aqueles que não ocupam o lugar do poder. Pensando na perspectiva das plataformas *online*, são usuários comuns ou “ordinários” (como denominou Michel de Certeau) que, astutamente, criam práticas diferentes (e até surpreendentes) do que estava preestabelecido. Já as estratégias surgem como disputa de forças em que uma empresa, uma organização ou uma instituição, sob ameaça de ser isolada ou vencida, postula ações calculadas de contraposição ao poder que tenta se estabelecer. Um exemplo pode ser dos veículos de comunicação se reorganizando para mapear lógicas e produzir informação mais circulável nas plataformas *online* como estratégia de retomada ou manutenção do poder de pautar a sociedade.

Defendemos a existência dessas ambivalências na produção e edição de texto, a partir de uma *Web* platformizada. As ações de domesticação, táticas e estratégias são, em grande medida, contingentes, mas que parecem se repetir em alguns casos. Não chega a ser uma repetição que estabiliza as práticas, aliás, ressaltamos que, ao observar o atravessamento das plataformas *online* em produções textuais, não nos esforçaremos em fixar essas interações ou dizer sobre a estabilidade delas. Todavia, um exemplo dessa repetição é a absorção massiva, por parte dos produtores, de rotinas de engenharia reversa, que se configuram como habilidades técnicas para deduzir os critérios algorítmicos de plataformas. Castro (2019) argumenta o quanto a estabilização de práticas pode, inclusive, facilitar o “trabalho” das plataformas:

Na medida em que consolida perfis, a performatividade está a serviço da governança algorítmica. Em princípio, a estabilização dos perfis facilita essa governança: se os interesses de um usuário são bem definidos, gerar recomendações a ele torna-se uma tarefa mais simples. Mas, não se trata tanto de um imperativo técnico, posto que os

algoritmos são capazes de captar de modo relativamente rápido as mudanças nas ações do usuário que impactam seus perfis. (CASTRO, 2019, p. 18).

Gillespie (2018) cita o caso de profissionais de *Search Engine Optimization* (SEO – Otimização de *sites*), que se engajam em tentativas de descobrir o funcionamento desses sistemas, para se apropriar de forma assertiva das plataformas *online*. As técnicas de SEO são responsáveis pelo tratamento de textos para que se tornem potencialmente mais visíveis no ambiente digital. Essa visibilidade é mediada e condicionada por algoritmos de plataformas de busca, que definem critérios de relevância, elegibilidade e prioridade para que textos sejam selecionados e distribuídos a leitores. O principal objetivo das técnicas de SEO é tornar o conteúdo mais atraente aos algoritmos de indexação dos motores de busca, como *Google e Bing*⁴⁰, sem perder de vista a pertinência ao leitor. Parte da técnica é papel dos desenvolvedores de *sites* e outra parte é de responsabilidade de produtores de texto, que inclui a hierarquização de parágrafos e a escolha de palavras-chave, como detalharemos no Capítulo 5.

Um exemplo desse ciclo de contingenciamento baseado em SEO é a criação de títulos testáveis, recursos de atração de cliques por meio dos títulos (BUENO; REINO, 2019). É possível apontar “eficiência” do título no que diz respeito à audiência na seguinte lógica: quando cliques do público engrossam as estatísticas, o produtor analisa a partir de quais palavras-chaves a audiência está acessando aquele conteúdo e, assim, cria uma engenharia reversa de usos de *tags* e outras estratégias cuja função é atrair audiência. Sobre isso, Rogers (2016) reflete criticamente:

As posições superiores das buscas são altamente procuradas; organizações fazem uso de técnicas de otimização de busca para aumentar a visibilidade de seus sites. Existem técnicas – mocinhas e vilãs – com essa finalidade, ou seja, aquelas aceitas pelos mecanismos e outras que induzem os motores para retirar os Websites dos resultados até que eles cumpram novamente com as regras do mecanismo. (ROGERS, 2016, p. 21).

As técnicas de SEO são apenas umas das práticas incorporadas que demonstram as dimensões de atravessamentos dos algoritmos na cultura escrita. Há outros elementos a se considerar, como avaliação de relevância feita pelos algoritmos de plataformas de busca. São

⁴⁰ Fazer SEO para Google pode ser diferente de fazer SEO para o Bing. Nesta pesquisa, toda vez que nos referirmos às técnicas de SEO, o enfoque é no Google. Apesar de serem funcionamentos semelhantes, os critérios (ou o peso desses critérios) e as equações para realizar o ranqueamento podem variar. Mas tudo isso é parte da caixa preta dos buscadores, sendo exatamente o que movimenta o mercado de SEO e faz com que a comunidade realize experimentos constantes. O Bing e os outros motores de busca do mercado têm uma governança própria, com documentos e diretrizes. Mas sabemos que as métricas de relevância são similares ao Google. Não as estudamos nesta pesquisa.

critérios pelos quais os algoritmos determinam o que deve ou não circular. No caso do *Google*, há mais de 200 fatores que definem a ordem das páginas apresentadas na entrega da busca, conforme detalharemos no Capítulo 4. Não sabemos exatamente quais são todos esses parâmetros e, além disso, os algoritmos que definem o ordenamento são alterados, em média, 600 vezes por ano (GOOGLE, 2022e)⁴¹. Investigamos o quanto essas escolhas do *Google* – que parecem ser apenas técnicas, mas têm imbricamentos político-sociais – remoldam práticas de escrita em CMS.

Para a distribuição de conteúdo, sabe-se que o *Google* considera data de publicação, localização do agente que faz a busca e idioma, que são fatores funcionais para identificação algorítmica. Há outros parâmetros, como transparência e originalidade do conteúdo, que são muito controversos, por colocar em xeque a promessa da objetividade algorítmica, que sabemos ser irreal. Gillespie (2018) caracteriza as dimensões dos algoritmos de relevância pública que têm valor político, ou seja, que cumprem o papel de controle da realidade. Uma das principais é, justamente, a não neutralidade algorítmica, principalmente sobre os critérios de seleção da informação, que são ocultos.

Na verdade, nenhum serviço de informações pode ser completamente isento de interferência humana ao entregar informações: embora um algoritmo possa avaliar qualquer site como o mais relevante para sua busca, esse resultado não a parecerá se for pornografia infantil, não aparecerá na China se for um discurso político dissidente, e não aparecerá na França se promover o nazismo. (GILLESPIE, 2018, p. 107).

Respondendo à crítica sobre o processo de mediação da informação social – por vezes problemática, ao permitir a circulação de desinformação –, em 2019, o *Google* anunciou que passaria a avaliar a reputação de artigos *online* a partir da tríade EAT (sigla em inglês para especialidade, autoridade e confiabilidade). Em 2022, atualizou a diretriz incluindo mais um item: EEAT (sigla em inglês para significar experiência, expertise, autoridade e confiabilidade). Isso significa que os algoritmos de busca devem reunir cada vez mais informações contextuais sobre uma empresa ou um autor que publica textos *online*. Ademais, a plataforma incluiu avaliadores da qualidade da pesquisa, que são pessoas que informam se os algoritmos geram bons resultados. A entrega da plataforma de busca, cada vez mais baseada em EEAT, leva em consideração a avaliação que empresas recebem *online* e *off-line*, a experiência de usuários e clientes em *sites* em lojas físicas, o currículo e o histórico de publicações de um autor de texto, entre outros fatores. Em complementaridade, naquele

⁴¹ Informação da própria empresa repassada pelo seu representante no *Google News Webinar for LATAM Publishers*.

mesmo ano, o *Google* lançou o *Core Update* BERT, baseado em modelos de processamento de linguagem natural (NLP). O BERT ajuda o *Google* a interpretar melhor os contextos do que é inserido na barra do buscador, em resumo, as pesquisas são interpretadas de uma forma mais parecida com o pensamento humano. Assim, a plataforma “melhora as entregas”, baseando-se na correspondência de intenção do usuário e se desconectando um pouco da lógica de busca por *tags* ou palavras-chave. Na prática, o buscador passa a compreender frases mais contextualmente, inclusive, diferenciando e considerando preposições e conjunções, conforme descreveremos no Capítulo 4.

Afinal, o que essas mudanças sinalizam? Parece ser uma tentativa de “aperfeiçoamento” do processo de mediação. Produtores de texto que usaram estrategicamente os buscadores para circular *fake news* ou conteúdos de baixa confiabilidade obrigaram o *Google* a se recriar. A ironia desses *updates* é o fato de que eles criam mais e mais camadas controversas na função de *gatekeeper* do *Google* e parecem contingenciar outros produtores e editores de texto a alterar as dinâmicas de engenharia reversa. É “aperfeiçoar” para perpetuar uma lógica hegemônica dessa plataforma que, aliás, tem 92%⁴² de fatia de mercado no setor de buscas.

Rogers (2016 p. 24) questiona: “Que sites são mais visitados por país e o que a visitação de sites diz sobre a cultura informacional de um país?”. No Brasil, os *site* mais visitados⁴³ são: google.com, youtube.com, facebook.com, globo.com, instagram.com, uol.com, whastapp.com, xvideos.com, twitter.com, google.com.br. A presença dos *sites* pertencentes às *Big Five* demonstra um cenário de grande intervenção em fluxos informacionais que incitam a mobilização de categorias profissionais no Brasil, a exemplo dos jornalistas. A Federação Nacional dos Jornalistas (Fenaj) sugere que essas companhias drenam recursos das empresas de mídia nacionais, contribuindo para o fechamento de veículos e a demissão em massa de jornalistas.

Ademais, a Fenaj tem se engajado em exigir a taxaço das plataformas *online* com a criação de uma Contribuição de Intervenção no Domínio Econômico (Cide), um imposto especial que permite sua destinação a determinado fim, neste caso, ao Fundo Nacional de Apoio e Fomento ao Jornalismo (Funajor). Em perspectiva mais ampla, o Comitê Gestor da Internet no Brasil (CGI)⁴⁴ vem discutindo modos de regulação das plataformas *online*,

⁴² Disponível em: <https://gs.statcounter.com/search-engine-market-share>. Acesso em: 18 mar. 2022.

⁴³ Disponível em: <https://www.similarweb.com/pt/top-websites/brazil/>. Acesso em: 03 jan. 2023.

⁴⁴ Organização multissetorial que congrega diversos setores ao reunir sociedade e governo para estabelecer diretrizes estratégicas relacionadas ao uso e desenvolvimento da internet no Brasil. Foi preponderante os

envolvendo debates sobre moderação de conteúdo, transparência, modelo de negócio e concorrência, em um esforço que envolve cientistas, Agência Nacional de Telecomunicações (Anatel), Conselho Administrativo de Defesa Econômica (Cade), Ministério Público Federal (MPF), associações de empresas de tecnologia, representantes das *Big Five* no Brasil, entre outros atores.

Em 2023, o Comitê publicou diretrizes⁴⁵ para regulamentação futura, entre elas, respeito à Lei Geral de Proteção de Dados Pessoais, promoção de trabalho decente e desenvolvimento de algoritmos que respeitem os Direitos Humanos. As discussões se inspiram no marco regulatório europeu – a partir do *Digital Market Act* (DMA), que já citamos, e do *Digital Services Act* (DSA) – e buscam soluções para o contexto brasileiro. Em complementaridade, há o Projeto de Lei n° 2.630/2020 (conhecido como *PL das Fake News*), que trouxe à cena uma queda de braço entre poderes Executivo e Legislativo no Brasil e as plataformas *online*.

Em meio a esses debates, o *Google*, em franca campanha⁴⁶ contra a votação do PL, chegou a ser investigado por privilegiar o ranqueamento de conteúdos contra o projeto, inclusive desinformativos, em detrimento de outros textos. O *Google* deverá explicar às autoridades⁴⁷ a lógica de seleção e visibilidade de conteúdos para o termo-chave “PL da Censura”, bem como os critérios de impulsionamento do *site*, em desacordo com suas próprias regras de autorregulação. Pela defesa do *Google*⁴⁸, tudo indica que a empresa usou estrategicamente a lógica da busca para ranquear conteúdos de seu próprio *blog* para o termo “PL 2.630”, uma combinação – até aquele momento – de menor tendência de procura. Afinal, o grande volume de busca se concentrava em “PL das Fake News”, em que no ranqueamento predominavam conteúdos jornalísticos, pois a imprensa denomina o projeto dessa forma. Ao que tudo indica, os usuários que optaram por buscar “PL 2.630” encontraram, prioritariamente, textos da empresa do Vale do Silício.

debates sobre o Marco Civil da Internet (Lei 12.969) e a Lei Geral de Proteção de Dados Pessoais (Lei 13.709).

⁴⁵ Disponível em: <https://cgi.br/publicacao/acoes-e-diretrizes-para-a-regulacao-de-plataformas-digitais-no-brasil/>. Acesso em: 17 fev. 2023.

⁴⁶ Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/maisdebate2630/>; <https://blog.youtube/intl/pt-br/inside-youtube/pl2630-impacto-criadores-de-conteudo/>; <https://blog.google/intl/pt-br/novidades/iniciativas/pl2630-precisamos-debater/>; <https://blog.google/intl/pt-br/novidades/iniciativas/pl-26302020-deixou-de-ser-sobre-combater-fake-news/>. Acesso em: 06 jun. 2023.

⁴⁷ Disponível em: <https://www1.folha.uol.com.br/poder/2023/05/moraes-manda-pf-ouvir-dirigentes-de-big-techs-apos-ofensiva-contr-pl-das-fake-news.shtml>. Acesso em: 06 jun 2023.

⁴⁸ Disponível em: <https://www1.folha.uol.com.br/opiniaio/2023/05/sobre-a-credibilidade-da-busca-no-google.shtml>. Acesso em: 06 jun. 2023.

Figura 5 – Página inicial do *Google* em 1 de maio de 2023

Fonte: Google, 2023.

Os debates da regulação encontram muita resistência – por conta de argumentos baseados nos modelos de negócios – das plataformas *online*. Nem mesmo a participação da ciência como mediadora das discussões está ilesa. A construção intelectual do conceito de ética em IA e as discussões noções de regulação de sistemas de busca e mídia vêm sendo absorvidas e financiadas pelas *Big Five*, conforme relatou Ochigame (2021a). De acordo com o pesquisador, as empresas estão, inclusive, manipulando a academia sob a máscara de estarem investindo no debate, quando na verdade estão minando a discussão com vieses de interesse comercial. Ochigame (2021a) desvela os entremeios do *lobby* das empresas principalmente com publicações de documentos demagógicos que demarcam seus “princípios éticos” ou “práticas responsáveis” – dotados de expressões como “algoritmos justos”, “viés”, “proteções” –, quando, na verdade, esses documentos protegem as empresas de regulações legais. O auge é a tentativa de essas companhias, juntas, proporem uma estrutura de “autorregulação”, na qual seria falseada a representação da sociedade civil. As *Big Five* estão investindo mesmo em criar algoritmos livres de regulação legal ou livres de problemas de regulação.

Tudo isso reforça que a tecnologia não é isenta, imparcial ou objetiva, pelo contrário, está coberta de ideologias e expressões de poder, além da natureza ambivalente. Cupani (2001, p. 168) nos lembra que “[...] vivemos numa sociedade na qual a tecnologia é apresentada como algo neutro e como algo positivo (na medida em que acena com a possibilidade de uma vida melhor)”. Olhar com criticidade nos ajuda a compreender que os algoritmos e as plataformas *online* são manejados com intencionalidades políticas e econômicas, além de serem apresentados em caixas-pretas. Isso dificulta (mas não impede) movimentos de contrafluxo ou ressignificações da produção que se sobreponham à lógica

algorítmica, em especial, quando pensamos nas plataformas de busca. “O Google não é mau, mas também não é bom do ponto de vista moral. Tampouco é neutro – longe disso” (VAIDHYANATHAN, 2011, p. 18). Furtado (2003), citando Roncaglia (2001), reforça que as características de interfaces usadas para “produção, circulação e leitura” não são de nenhum modo neutras e não deixam de influenciar nem a estrutura textual, nem os modos de fruição do texto. Algoritmos interferem nos processos que realizam, conforme pontua Jurno (2020, p. 106).

É justamente por serem atores ativos na rede de ações que se dá na plataforma que os algoritmos agem e fazem outros atores agirem a partir da sua própria ação. [...] Quando assumimos que os algoritmos são entidades que agem, assumimos também que eles modificam a ação e influenciam os outros atores que compõem uma determinada rede. Portanto, eles não podem ser tomados como neutros.

Os algoritmos são carregados de vieses e automatizações equivocadas. Manter aparência de tecnicidade ajuda os provedores a fugir de acusações de manipulação de resultados. Imperdoável é ignorar que os algoritmos são programados por seres humanos, com suas escolhas subjetivas ou guiadas por instituições (GILLESPIE, 2018). Quando se trata de experimentos tecnológicos das *Big Five*, certamente, a preocupação é maior com o funcionamento que os mantém detentores de poder (político e econômico) do que com o conhecimento coletivo.

Sendo assim, reiteramos a intenção em definir uma cultura digital pós-plataformas pensando ser esta mais atravessada pela comunicação mediada por algoritmos e pela tradução de tudo em dados numéricos, conforme alertou Ribeiro (2018), citando Felinto (2006). Segundo a autora, ao lermos autores entusiastas da tecnologia e ciberespaço, talvez tenhamos ficado com a impressão de que a cultura digital era tudo que sempre sonhávamos: “[...] um mundo de compartilhamento, cooperação, diálogo, leituras móveis, andanças, liberdade e troca” (RIBEIRO, 2018, p. 15). No entanto, percebemos que essa cultura digital pós-plataformas tem como fator central embeber nossas vivências sociais com uma experiência tecnológica financeirizada pelas *Big Five*.

Faz sentido dizer sobre outra cultura digital porque a dinâmica que se estabelece não é nova, é somente outra. Afinal, a financeirização também permeava a cultura digital do início dos anos 2000, haja vista a explosão da *bolha.com*⁴⁹. Ela também atravessava cultura

⁴⁹ Em março de 2000, cerca de 500 empresas – que prestavam serviços na recém-popularizada internet – quebraram, depois de perderem muitos índices na bolsa de valores. Isso aconteceu pouco depois de essas companhias triunfarem de modo especulativo no mercado. Despençou-se um cenário de grande confiança do mercado nessas empresas, o que ficou conhecido como estouro da *bolha.com*. Disponível em:

impressa, pois sabemos do grande poderio de poucos conglomerados de comunicação e do mercado editorial que vigorou por mais de um século. Há, portanto, uma tensão constante para troca desse lugar de poder temporário. Em todos esses momentos, houve efeitos em nossas práticas de escrita e leitura, com culturas em disputa, impressas e digitais.

2.3 CADEIA DE VALOR DO TEXTO PLATAFORMIZADO

O texto plataformizado se insere no que argumentamos ser a cultura digital pós-plataformas; portanto, funciona com fluxos de produção particulares, alterados pelas dinâmicas dos meios, em especial, pelas plataformas de busca. A ideia de fluxo se refere ao texto em construção e que possibilita intervenções geradas, em cadeia, por seu autor ou não. Recorremos a alguns pesquisadores, em particular Thompson (2013) e Furtado (2003), que se dedicaram a refletir sobre essa cadência produtiva do texto (impresso e digital).

Jonh B. Thompson, em *Mercadores de cultura*, detalhou essa noção de cadeia (ao se referir à cadeia editorial), como o conjunto de atividades e papéis desempenhados por esses agentes almejando o objetivo comum de produzir, vender e distribuir a mercadoria livro. Thompson (2013) desenhou o fluxo editorial dos livros considerando que há uma cadeia de suprimentos e uma cadeia de valores, formatadas por diferentes atividades. Desde a criação de um texto até o acesso de leitores, por meio do mercado varejista, o produto passa pelas mãos de agentes (ou apenas um agente), que modificam e agregam valor, em etapas suplementares. Cada elo fornece, de modo funcional ou simbólico, algo substancial ao trabalho de produção e, apesar da relação de interdependência entre agentes, a cadeia editorial é dinâmica, ou seja, mudanças estruturais, como crises financeiras ou questões tecnológicas, influenciam o todo, podendo alguns dos intermediários deixarem de existir ou nova função ser acrescida.

A representação do fluxo é uma fotografia possível de ser feita em outros campos da produção de texto, em modo impresso e em ambiente digital. Seria possível enxergar esses fluxos na produção jornalística, em conteúdos para marcas, conteúdos de entretenimento, na divulgação científica, entre outros. Essa ideia nos inspira a pensar sobre uma cadeia de produção de texto plataformizado, quando incluímos as várias camadas de mediação e condicionamento que compõem o fluxo de trabalho na produção de conteúdos em/para plataformas *online*.

Thompson (2013) detalhou as mudanças na lógica do campo editorial após revolução digital e listou nove aspectos (QUADRO 1) em que as tecnologias podem permitir que produtores e provedores de conteúdo agreguem valor real a seus negócios. Partindo dos aspectos apresentados por Thompson (2013), dialogamos com Furtado (2003), que argumenta sobre a existência de uma cadeia de valor da informação e propõe uma mirada exclusiva para a “produção, circulação e leitura” de textos em ambiente digital. Conforme o autor, em comparação à produção impressa, cria-se uma economia específica para a edição e a distribuição de conteúdo, pois se multiplicam os formatos, meios e canais de distribuição no ambiente digital. Isso repercute em dimensões culturais e sociais de escrita, alterando o modo como as pessoas interagem com textos. Furtado (2003) propõe, então, uma (re)concepção de cadeia de valor da informação e, para ele, é fundamental pensar em seis elementos-chave que integram o fluxo produtivo da informação (QUADRO 1).

Quadro 2 – Elementos da cadeia de valor da informação e do conteúdo

Thompson (2013)	Furtado (2003)
Facilidade de acesso	Seleção
Capacidade de atualização	Acesso
Escala	Agregação
Capacidade de pesquisa	Navegação
Portabilidade	Autoridade
Flexibilidade	Desenvolvimento
Preços acessíveis	-
Intertextualidade	-
Multimídia	-

Fonte: Elaborado pela autora com base em Thompson (2013) e Furtado (2003).

Todos os elementos têm implicações relevantes na circulação de textos, porque determinam a tênue relação entre espaço produtivo e espaço de leitura/consumo. No contexto da plataformização da *Web* e da produção cultural, esses critérios ganham camadas de mediação e são subvertidos de muitos modos. Em um texto produzido anteriormente (SOARES; CRUZ, 2020), demonstramos, por exemplo, como esses elementos-chave funcionam para os textos de ciência em ambiente digital, gerando controvérsias midiáticas e impulsionando o potencial de propagabilidade dessas textualidades. Portanto, vamos articular os elementos apresentados para argumentar o atravessamento das plataformas *online* e seus

algoritmos nessa nova cadeia de valor da produção textual, pensando, especialmente, em como os processos de edição se reconfiguram.

Quadro 3 – Articulação de elementos da cadeia de valor da informação e do conteúdo

Thompson (2013) e Furtado (2003)
Seleção
Acesso, facilidade de acesso e preços acessíveis
Agregação, intertextualidade e flexibilidade
Navegação, escala e capacidade de pesquisa
Desenvolvimento, capacidade de atualização, portabilidade e multimídia
Autoridade

Fonte: Elaborado pela autora com base em Thompson (2013) e Furtado (2003).

2.3.1 Seleção

O processo de seleção se refere às escolhas de conteúdos que serão circulados; no caso dos textos plataformizados, algoritmos acabam por “dividir” espaços com *gatekeepers* tradicionais, como jornalistas, escritores e editores, para determinar quais textos terão visibilidade. Essa seleção sempre foi uma atividade não neutra, dotada de subjetividades e capaz de reconfigurar agendas sociais, portanto, controversa. Torna-se ainda mais passível de discussão quando grandes empresas provedoras de informação – autodescrevendo-se como empresas de tecnologia – dominam expressivamente essa função mediadora e são condicionantes. Afinal, a seleção pode representar uma ação onipotente de disseminação e imposição de um conjunto de ideias aos públicos.

A seleção é ação fundadora de qualquer cadeia editorial, aliás, uma componente marcante em fluxos de produção, sejam eles tradicionais ou plataformizados (esses que dependem de plataformas para existir). Recorremos a Medeiros (2009) para compreender as atividades editoriais – em suas componentes de mediação e prescrição de sentido –, apontando as relações entre as atividades de produtores/editores de textos e o uso de *softwares*. Por muitos anos, conduziu-se uma visão sustentada do “[...] editor como árbitro do gosto público (Altbach, 1995), modulando as preferências e os modos de consumir” (MEDEIROS, 2009, p. 134), função agora tensionada pelo uso das plataformas como parte da cadeia produtiva. Muitos desses profissionais, especialmente aqueles produtores sujeitos a

mecanismos de busca, plataformas de mídia social, sistemas de recomendação e bases de dados, fazem usos exacerbados da lógica de seleção algorítmica, ferindo princípios éticos em nome da performance. Aliás, Nieborg e Poell (2018), como já citamos, abordam essa questão ao apresentarem o conceito de “mercadoria cultural contingente”, considerando que as produções cada vez mais pautadas em rastros (dados) deixados pelos leitores/consumidores sejam sinais de que a lógica algorítmica toma o lugar da lógica editorial. Assim, é mais valorizada uma produção orientada por dados, no qual o conteúdo é constantemente alterado para otimizar a distribuição e monetização da plataforma. Furtado (2003) afirma que o ambiente digital permite que canais habituais sejam ignorados, o que diminui a importância da seleção humana e implica um acréscimo do valor da navegação.

Jurno (2021) corrobora essa avaliação, em suas pesquisas sobre a plataformação do jornalismo, afirmando que, enquanto as escolhas jornalísticas se baseiam em escolhas de especialistas, de profissionais capacitados para sua realização, a seleção algorítmica depende das escolhas procedimentalizadas de máquinas. Gillespie (2018) acrescenta que são escolhas projetadas por operadores humanos para automatizar alguma representação do julgamento humano ou desenterrar padrões por meio de traços sociais coletados.

Se voltarmos a pensar nas camadas do processo de edição de texto (representadas na FIGURA 1), vemos todas elas – humanas e não humanas – compondo um conjunto de “árbitros do gosto público” contemporâneos. As plataformas *online*, a partir das técnicas de ordenação da informação, canalizam a prática humana continuamente e se tornam árbitras generalizadas das possibilidades da vida real (RIEDER, 2020). É verdade que, se não fossem os processos de seleção da informação, “[...] leitores submergiriam a um dilúvio” (MEDEIROS, 2009, p. 135) de conteúdo. Nos séculos XIX e XX, a imprensa exerceu bastante esse papel seletor, o que colocava os conglomerados de mídia na mira das críticas, por exercerem a atividade pautada em interesses privados. Isso ocorre atualmente com as lógicas estabelecidas por plataformas, mas demorou um tempo para que a sociedade enxergasse essa função seletora.

Assim como os motores de busca, os jornalistas desenvolveram táticas para determinar o que é mais relevante, como reportá-lo e como garantir sua relevância - um conjunto de práticas que são relativamente invisíveis para sua audiência; uma meta que eles admitem ser mais difícil de seguir do que parece; e um princípio que ajuda a deixar julgamentos de valor e políticas pessoais de lado, mas não os elimina. (GILLESPIE, 2018, p. 108)

Outros *gatekeepers* tradicionais da sociedade também perdem espaços de poder para as plataformas *online*, a exemplos de médicos e advogados, que detinham vozes do conhecimento e agora são questionados e confrontados a partir de uma busca no *Google*, que pode fabricar falsos especialistas. Considerar que os “árbitros do gosto público” são necessários ao dinamismo da vida social é aceitar que sempre haverá camadas de interesse privado na produção do conhecimento humano. Mesmo com lógica da cauda longa (ANDERSON, 2004)⁵⁰ se estabelecendo – abrindo possibilidades de circulação de conteúdos especializados, de nicho, de autopublicação, da inovação de estilos e propostas editoriais –, ainda há grande seleção de conteúdos feito por grupos hegemônicos.

Ademais, a filtragem algorítmica rotineiramente suprime algumas perspectivas políticas e promove outras, independentemente da intervenção editorial humana. Ochigame e Holston (2016) ressaltam que essa seleção pode fazer o mesmo papel, porém de forma implícita, da censura aberta da internet, que ocorre, por exemplo, na remoção de servidores, apreensão de nomes de domínio e negação de serviço. Para os autores, não é necessária essa censura explícita para controlar o fluxo de informações para fins políticos, sendo possível fazer isso somente com filtragem algorítmica contínua e dotada suas lógicas de promoção e supressão exercida por muitos atores (cientistas da computação, linhas de código, corporações privadas e preferências do usuário). É por isso que nas discussões sobre regulação das plataformas *online*, a sociedade civil está cobrando transparência das empresas sobre critérios de seleção e moderação de conteúdo, mas as companhias resistem a abrir diálogo nessa área sob o escudo do argumento de “segredo de negócio”.

Enfim, esse processo complexo de seleção, classificação e categorização que vamos explorar mais detalhadamente, ao esmiuçar o funcionamento do *Google* no Capítulo 4, limita sistematicamente a diversidade de vozes *online* e, em muitos casos, invisibiliza certos tipos de fala. O pesquisador Sérgio Amadeu Silveira (2022)⁵¹ afirma que a tão aclamada IA das plataformas *online* se baseia em operações de classificação e extrações de padrões de uso, o que rotineiramente suprime outros tipos de inteligências e reduz “tecnodiversidades”.

⁵⁰ Conceito criado por Chris Anderson (2004) e que foi apropriado para vários setores. No contexto que aplicamos aqui, refere-se à possibilidade de conteúdos de nichos fazerem tanto sucesso quanto conteúdos populares ou “hits de sucesso”, o que pode ser crucial para estratégias de conteúdos especializados, por exemplo.

⁵¹ Palestra realizada pela Escola de Ciência da Informação, da Universidade Federal de Minas Gerais (UFMG), em 2022. Disponível em: <https://shre.ink/a8sl>. Acesso em: 13 fev. 2023.

2.3.2 Acesso, facilidade e preços

Thompson (2013) argumenta que, diferentemente de bibliotecas físicas, o conteúdo *online* não é governado pelas restrições espaciais e temporais, sendo os dispositivos de acesso (computador, celular) a porta de entrada para um volume de conteúdo. Quando reflete sobre acesso, Furtado (2003) questiona em que espaço digital se hospeda um texto e como é possível acessá-lo, o que nos remete novamente aos “espaços de escrita” (BOLTER, 1991). Essa disponibilidade dos textos depende, atualmente, de várias plataformas *online*, especialmente, das mídias sociais (espaços de publicação) e das plataformas de busca (espaços de encontrabilidade). A facilidade de acesso do texto digital parece nos libertar das limitações inerentes a uma localização espacial – como no caso da biblioteca física –, tornando-se espaço fértil para a dominação de serviços de indexação e pesquisa, como os oferecidos pelo *Google*. No entanto, como já argumentamos anteriormente, não é exatamente uma libertação.

Ademais, alteram-se os aspectos logísticos do acesso ao texto e cria-se nova ordem de chegada ao público, com alterações, inclusive, nos modelos de negócios das organizações produtoras. Na perspectiva de quem produz, é preciso pensar em como um texto será disponibilizado gratuitamente e se há formas de monetização dentro das plataformas *online*. Há também uma sobreposição de funções, pois o mesmo profissional ou empresa pode criar, editar e distribuir conteúdo sem que esses processos estejam partilhados com outros *players*. Já na visão de quem consome, o acesso quase sempre é franqueado pela troca de dados no ato de concordância dos “Termos de Serviço e Privacidade”. Trazemos, novamente a reflexão de Gillespie (2018) sobre públicos calculados em que ele afirma: “[...] somos levados – por algoritmos e por nossa própria preferência pelos que pensam de forma semelhante – para dentro de “filtros bolha”, onde encontramos apenas as notícias que esperamos encontrar e as perspectivas políticas que já nos são caras” (GILLESPIE, 2018, p. 114).

O acesso “gratuito” complexifica a noção de posse sobre o texto, diferentemente de quando se pega um livro emprestado em uma biblioteca com o compromisso de consumo temporário ou de quando se compra um livro levando o objeto como “patrimônio individual”. Quando produtores/editores passam a criar e circular conteúdos gratuitamente, o que se vende ali não são exatamente essas mercadorias culturais/intelectuais, o que está em disputa é ganhar tempo de atenção dos públicos. Vejamos um exemplo de influenciadores digitais que praticam produção de conteúdo para marcas: o criador raramente está vendendo produtos a

partir dos conteúdos que cria, mas precisa muito conquistar tempo de atenção, porque isso gera valor para as marcas. Quanto mais tempo públicos estão expostos a conteúdos, mais valorosos se tornam os negócios para todos envolvidos: influenciadores, marcas e plataformas *online*. Estas, por sua vez, beneficiam-se muito do tempo gasto pelos públicos dentro dos ambientes criados, o que ajuda a sustentar seus pilares (programabilidade, popularidade, conectividade e datificação). Sob essa perspectiva, podemos argumentar que toda produção cultural e textual plataformizada concorre entre si por tempo de atenção dos públicos. E um dos elementos que faz essa engrenagem girar é o acesso “gratuito”.

2.3.3 Agregação, intertextualidade e flexibilidade

A agregação diz respeito à relevância embutida no texto durante a criação. No caso da *Web* plataformizada, agregar significa dar novos contornos ou encaixar o texto em *templates* porque esse produto será entregue ao leitor e aos algoritmos de plataformas de busca. Furtado (2003) discute o quanto incrementar uma publicação com hipertextualidade, intertextualidade e conectividade pode gerar autoridade, o que nos remete àquele pilar que já sugerimos, na seção 2.1, de “[...] escrever para *Web* hoje: planejamento e execução de uma produção mais hipertextual possível”. A agregação pode ser pensada pelo viés da relevância ao leitor, quando ele encontrará as possibilidades de construir múltiplos caminhos com o texto e de consumir uma informação de fontes confiáveis; ou pelo viés da relevância algorítmica, quando os códigos são programados para ler a conectividade entre textos e atribuir autoridade àqueles que se juntam de forma consistente, fortalecendo uma rede de informações.

Thompson (2013) argumenta sobre o caráter dinâmico da função referencial do texto em ambiente digital, pois há uma possibilidade enorme de fazer referência a outros materiais (em uma construção intertextual e hipertextual do conteúdo). Vejamos o exemplo do processo de *link building*, uma técnica de *SEO one page* e *off page* muito popular entre jornalistas e produtores de conteúdo. Ao organizar um texto para publicação em um CMS, produtores planejam a construção de *hiperlinks* no decorrer da edição, levando o leitor a outros *sites* que podem ampliar sentidos de leitura e trazer informações extras. No entanto, essa articulação de “para quais *sites* vamos apontar nossos *hiperlinks*” passou a ser parte da estratégia de relacionamento das empresas e veículos de comunicação. Algoritmos do *Google* enxergam de forma positiva essa construção de rede entre *sites*, quando uns fazem referência a outros, atribuindo pontos de autoridade a artigos que se comunicam enredados. Com as ações de *link building*, empresas começaram a construir relacionamentos fora do digital para

angariar *links* em seus textos *online*. Tão importante quanto *linkar* para o *site* de um parceiro, é que ele também referencie seu *site*. Essa dinâmica de *lobby* por *links* fez emergir a necessidade de profissionais especializados, que são os *Digital PRs* ou Relações Públicas Digitais. Geralmente, são jornalistas ou profissionais de relações públicas que ampliam a função de assessoria de imprensa e passam a articular esses relacionamentos. Aliás, a construção de imagem que sempre pautou a atuação desses lobistas é agora legitimada (ou não) pelo *Google*, conforme veremos adiante, ao apresentar a ideia de autoridade.

Ainda sobre essa questão dos *links*, Rogers (2016) convida a uma abordagem metodológica que muito nos interessa: como os mecanismos de busca tratam os *links*? Como a reputação de um ator poderia ser caracterizada pelos tipos de *hiperlinks* estabelecidos e recebidos? Em estudos anteriores, o autor observou que *sites* podem ser perfilados não apenas pela quantidade de *links* recebidos, mas pela análise de quais *links* eles estabelecem, ou seja, é exatamente essa lógica de construção de rede de reputação da qual um *Digital PR* precisa dar conta. O que ninguém consegue responder, em consenso, é quais marcadores as plataformas de busca usam para medir o impacto desses laços que se formam e atribuir autoridade.

Por fim, Thompson (2013) aponta para um processo de produção e leitura com flexibilidade, em que é possível fazer mais idas e vindas no texto e compartilhar conteúdos. Esse modelo de agregação, segundo Furtado (2003), privilegia a escrita e a leitura orientadas para ação – clicar, assistir, ouvir, mover, abrir janela, fechar – e para pesquisa.

2.3.4 Navegação, escala e possibilidade de pesquisa

A dinâmica descrita na seção anterior já nos leva à ideia de navegação, termo que Furtado (2003) usa para descrever o modo como leitores/consumidores podem encontrar a informação que procuram. Recapitular os movimentos de interação de públicos com o texto em ambiente digital nos ajuda a perceber grandes mudanças na navegação. Em um primeiro contato com a *Web*, há mais de 30 anos, quando as vidas ainda não estavam completamente atravessadas pelas telas, o leitor parecia flunar em ambiente digital para compreender seus usos e conhecer funcionalidades. Logo depois, começou efetivamente a navegar com objetivos mais claros e, posteriormente, pautou-se no processo de busca. Quando desembocamos na “era da busca”, demos conta do quanto o movimento de interação entre públicos e textos está plataformizado, assim como as cadeias de produção desses textos. Aliás, no trecho a seguir, Furtado (2003), citando Sellen e Harper (2002, p. 207), categoriza

várias atividades digitais que acabaram sendo ponto de partida para criação de serviços de plataformas *online*:

As tecnologias digitais encarregar-se-ão cada vez mais das actividades para que os media electrónicos estão mais talhados: “pesquisa em larga escala e recuperação de informação e documentos; mensagens curtas em comunicações internas/externas; análises de dados; produção e finalização de documentos; processamento de dados referentes a negócios e a transacções; disseminação e transmissão de documentos em grande escala; e elevada capacidade de armazenamento a longo prazo de documentos. As tecnologias digitais também predominarão na produção e distribuição de novos tipos de media, tais como documentos não lineares e pesquisáveis (dicionários, manuais de referência, enciclopédias), e géneros multimedia (vídeos, música e conteúdo multimedia interactivo)”. (FURTADO, 2003, p. 50).

Thompson (2013) aponta a larga escala ou a capacidade de prover acesso a grande quantidade de conteúdo com uma característica preponderante para provedores em ambiente digital. Para o autor, “[...] o todo é mais que a soma das partes” (THOMPSON, 2013, p. 370), e a abrangência de coletâneas é valorizada em meio ao grande *continuum* textual. Ele pondera que não somente quantidade é importante, mas a relevância do que circula, considerando que o leitor consome o que acha pertinente ao seu sistema de expectativas.

É a capacidade de economia de escala da internet que tem impulsionado a agregação de negócios online. Surgiram inúmeros intermediários com o objetivo de agregar grandes quantidades de conteúdo e depois vendê-lo [...] Parte do que é atrativo aos usuários finais é saber que podem encontrar o que estão procurando em um único site – um balcão único. (THOMPSON, 2013, p. 370).

Ainda segundo Thompson (2003), quanto maior a escala, mais valioso é ter um modo eficaz de busca nos bancos de dados. Segundo ele, a possibilidade de pesquisa no acervo digital usando mecanismos de busca, baseado em palavras-chave, é uma maneira de agregar valor complementar à provisão de escala, de alguma forma, exigida por ela. Vale relacionar essa previsão do autor com a lógica de funcionamento do *Google*, que, por muitos anos, se baseou em buscas por *strings*⁵². A mudança desse modelo de classificação de conteúdos se concretizou no *Core Update BERT*, em 2019, tornando a pesquisa por palavras-chave menos relevante e privilegiando o contexto no ato da busca.

⁵² São sequências de caracteres, geralmente utilizadas para representar palavras, frases ou textos de um programa ou em um algoritmo, como o do *Google*. O BERT realiza pesquisas realmente sobre a correspondência de intenção, em vez de correspondência pura de *strings*.

2.3.5 Desenvolvimento, capacidade de atualização, portabilidade e multimídia

Ao se referir ao desenvolvimento, Furtado (2003) apresenta os processos de escolhas textuais e imagéticas que compõem o fazer/editar texto em ambiente digital, ou seja, mais uma esfera alterada por plataformas *online*. Podemos completar com o que Thompson (2013) nomeia de “capacidade de atualização”, ou seja, a facilidade de mudanças no texto em muitas etapas da cadeia editorial em ambiente digital, se comparamos com a lógica de produção do conteúdo impresso. Aliás, Camila Gonzaga-Pontes (2012) observou um exemplo dessa dinâmica, a partir das noções de reescrita e retextualização, quando estudou o (micro)fluxo de produção da webnotícia (apuração > produção da notícia > publicação da notícia). A pesquisadora descreveu o que ocorre com atualização constante na medida em que os fatos são apurados em “tempo real”; assim, concluiu que o texto vai sendo encorpado, em pequenas fases da informação, como se passasse por uma linha de produção que se revela na versão final.

Thompson (2013) também identifica a portabilidade como um valor agregado, porque deixa o conteúdo versátil, transferível e passível de armazenamento em dispositivos diversos, desde que não tenha formato exclusivo. Outro ponto de valor é a possibilidade de produção multimídia, em que uma variedade de meios pode suplementar o texto com conteúdos oferecidos em várias formas visuais. Thompson (2013) considera que pode ser caro produzir, mas pelo menos existe essa possibilidade de desenvolvimento.

2.3.6 Autoridade

Por fim, o elemento autoridade que ganha certo protagonismo na (re)concepção de cadeia de valor da informação. No contexto das plataformas *online*, importa muito qual *brand* está produzindo o texto e essa autoridade é fator de peso para as plataformas de busca, haja vista o lançamento dos pilares EEAT pelo *Google*, em 2019. Nesse ponto, enxergamos um paradoxo, pois o modelo de circulação de textos – em *continuum* abundante –, com maior liberdade conferida aos indivíduos produtores de conteúdo e com seleção algorítmica, causou uma precariedade institucional por ser um ambiente com textos em plena construção e sem garantia de autoridade e autenticidade.

Justamente no momento em que escapamos aos mecanismos e dispositivos tradicionais de institucionalização dos textos, uma plataforma de busca, como o *Google*, passa a medir, com critérios encaixapretados, os itens Especialidade, Autoridade e Confiabilidade.

Recentemente, em um webinar⁵³ oferecido pela *Google/Alphabet Inc.*, o jornalista e *Public Liaison for Search* Danny Sullivan disse que um dos principais critérios para formação de EEAT seria a originalidade de artigos publicados. Ele não conseguiu responder, no entanto, a centenas de jornalistas, sobre quais critérios a plataforma utiliza para classificar um artigo como original.

2.3.7 Desdobramentos de uma nova cadeia de valor

Ao apresentar os elementos da cadeia de valor dentro do contexto da plataformização da *Web* e da produção cultural, conseguimos vislumbrar novos desenhos, mais dinâmicos e complexos. Há um atravessamento das plataformas *online* e seus algoritmos, inclusive com uma subversão da importância de cada processo em cadeias editoriais. A navegação e autoridade ganham novos *status*, diferentemente de quando *gatekeepers* tradicionais definiam o que circularia e os públicos leitores se valiam dessas escolhas ou dessa seleção. E, nesses casos, sabemos da influência dos motores de busca e o modo como algoritmos percebem a autoridade de quem produziu, valorizando parâmetros além da linguagem e da informação em si.

Aliás, de todas as especificidades do ambiente digital, a introdução dos algoritmos na cadeia produtiva é a que mais complexifica a circulação de textos, porque tem ramificações políticas, sociais e simbólicas. As camadas descritas como hipótese na Figura 1 tensionam as funções da cadeia editorial. Percebe-se, com isso, que o produtor/editor parece dilatar suas atribuições. Por isso, nunca foi tão necessária a lucidez intelectual e técnica sobre o poder transformador dos algoritmos na escrita em ambiente digital. Esse produtor/editor raramente era cobrado, por exemplo, a ter conhecimento de técnicas de otimização, o que hoje já não se considera conhecimento acessório, é uma dinâmica naturalizada como parte da função.

Sabemos que as tecnologias não são marcapassos sociais, e que o mundo é muito mais complexo do que os deterministas tecnológicos gostariam, por isso é fundamental discutir as práticas de edição de textos tensionadas pelas várias camadas de mediação. O produtor/editor – pensando aqui especialmente nos humanos⁵⁴ – acaba se envolvendo com as lógicas de *softwares* de maneiras desconcertantes. Vale lembrar que nem todos os usuários vão se apropriar da mesma forma das funcionalidades disponíveis em plataformas *online*. A

⁵³ Ver nota 13.

⁵⁴ Vamos ponderar que a função produtor/editor é exercida muitas vezes por robôs, conforme vem demonstrando, em trabalho contínuo, a pesquisadora Silvia Dalben (2018), que cartografou casos de redação automatizada de notícias.

partir de nossas vivências na indústria da produção de conteúdo para ambiente digital, relatamos abaixo algumas situações que podem ocorrer com esse produtor/editor e, em quase todas elas observamos algum nível de precarização do trabalho:

- a) automático: sente-se obrigado a obedecer às dinâmicas das plataformas de publicação e de busca, fazendo isso de forma automatizada, viciosa e, às vezes, acrítica;
- b) performático: faz um uso exacerbado dessas dinâmicas, transcendendo questões éticas da produção textual para privilegiar somente a performance;
- c) cercado: é pressionado institucionalmente na instância de publicação para que altere brutalmente suas estratégias na instância de edição e, mesmo questionando sobre esses resultados, não encontra alternativas para subverter o condicionamento algorítmico;
- d) desconhecedor: não conhece a lógica das plataformas e ignora a necessidade de reconfigurar práticas de produção;
- e) contrafluxo: opta por usar taticamente a lógica das plataformas, sob pena da invisibilidade, porque sua produção não exige performance como resultado ou o faz isso em forma de protesto;
- f) subversivo: elabora modelos resistentes de produção e circulação que fazem os textos independerem ao máximo das lógicas de plataformas. Ou seja, alcança visibilidade com expertises que não são obedientes aos algoritmos.

A descrição dessas situações dialoga com o espectro criado por Vecchio (2022), ao pesquisar as projeções (imagens de si) de autores que publicam livros em plataformas *online*, como *Kindle Direct Publishing* (KDP), da *Amazon*. Nas análises, emergiram quatro *ethé*: o *ethos* do autor empreendedor, que se submete às lógicas das plataformas de autopublicação, investindo na divulgação de suas obras com base no funcionamento de algoritmos em busca de performance; o *ethos* do autor diletante, em que a escrita aparece como uma necessidade e para quem a publicação em plataformas seria um desdobramento dessa necessidade e, portanto, uma realização humana obtida pela apropriação da tecnologia; o *ethos* do autor independente, aquele que, mesmo publicado por plataformas, tem uma história com a autopublicação e com a literatura que vai muito além das plataformas; e o *ethos* do autor-editor, como um agente do mercado editorial que se forma editor por meio do acúmulo de experiência de suas atividades de escrita e publicação de livros próprios, iniciando sua trajetória como autor e editor de si mesmo e, paulatinamente, capacitando-se para editar terceiros. Vecchio (2022) organizou o espectro representando uma posição de

maior ou menor adesão às lógicas algorítmicas, sendo a primeira categoria aquela em que “[...] emergiram mais posicionamentos que indicavam maior tendência de submissão ao capitalismo de plataforma e à curadoria algorítmica e a última um avizinhamento maior da apropriação, subversão e independência em relação às plataformas” (p. 211).

Este é um exemplo do mercado de autopublicação de livros, mas o mundo da produção e edição de textos em ambiente digital abarca uma pluralidade de campos. Cada um desses campos tem suas particularidades e não se pode generalizá-los, apesar de que, na perspectiva de incidência das lógicas das plataformas *online*, eles podem ser semelhantes, porque seguem a dinâmica do capitalismo de plataforma. Quer dizer que, independente do tipo de texto produzido em ambiente digital – jornalístico, científico, conteúdo para *marketing*, literário –, a mediação algorítmica poderá incidir reconfigurando práticas. Chartier (1999, p. 145) já alertava: “[...] a revolução eletrônica, evidentemente, acelera as concentrações”. Os “quase” monopólios exercidos sobre a informação e o patrimônio textual que, aliás, anda junto com as dominações linguísticas, não param de se ramificar. Essas dominações que invalidam conhecimentos e instituições fazem surgir o desafio maior do nosso presente: a subversão.

Querido, Carlos d'Andrea,

Você sabe, né? Minha mãe, aos 63 anos, é trabalhadora de plataforma online. Ela é anfitriã do Airbnb e recebe pessoas de várias cidades do mundo. Por um lado, envolver-se com esta atividade fez um bem danado a ela, que vivia no apartamento sozinha e, por vezes, solitária. Em contrapartida, começou a trabalhar muito para atender as demandas contingentes de hospedagem, afinal, toda a preparação e a arrumação pós-visitantes é feita por ela mesma. Portanto, hoje minha mãe vive como dona de seus dias, mas escrava de suas horas.

É meu irmão quem faz a gestão dos aluguéis de temporada pela plataforma online, aliás, ele gerencia não só o apartamento da minha mãe, mas dezenas de imóveis. Ele é “empreendedor de si”, mas vem fazendo usos táticos do Airbnb e expandindo as perspectivas de anfitrião e trabalhador de plataforma, para um lugar de “empresário do setor imobiliário e do turismo”. Parece ser uma subversão inteligente, mas só porque ele veste conscientemente, politicamente e ideologicamente a camisa de sujeito neoliberal. Talvez, ele performe bem.

Por que estou escrevendo sobre isso? As entranhas das plataformas online estão muito próximas de todos nós, em várias medidas. Não bastasse nossos textos, trabalhos, comunicações e relações estarem plataformizados, é bem estranho ver as pessoas que a gente ama com suas vidas capturadas pela informatização, operada por gramáticas algorítmicas, definidas por softwares e suportadas por computadores.

É esquisito, por exemplo, ver os esforços da minha mãe gamificados por notas, estrelas e comentários elogiosos (ou não). A medida do desempenho só vem depois que a plataforma publica a avaliação de um hóspede. Até lá, a anfitriã vive na taquicardia de uma pré-vestibulanda às vésperas do gabarito. Parece estressante, mas é uma realidade não diferente de jornalistas, motociclista entregadores, autores de livros, editores de blogs, músicos, motoristas, donos de restaurantes, podcasters, influenciadores digitais, redatores, cientistas e divulgadores da ciência, professores e alunos, revisores... Tem muita gente esperando por métrica, pontuação, resultado, classificação e posição no ranking!

Obrigada por contribuir com nossas reflexões sobre plataformas online. Vou continuar refletindo por aqui...

Abraços,

Luana.

3 METODOLOGIAS

Neste capítulo, reunimos estudos que dão sentido ao nosso entendimento da materialidade das plataformas *online* e nos ajudam a apurar o olhar “[...] na direção da construção da consciência de ignorâncias nossas” (HISSA, 2012, n.p.). Essas possibilidades de diálogo foram fundamentais para o desenho metodológico, ou seja, não repetiremos exatamente métodos já aplicados por outros pesquisadores, mas partiremos de experiências anteriores de muitos deles. Acreditamos que isso também é conceber metodologias e pesquisas que colaboram com o desenvolvimento científico e comunicam com o mundo.

3.1 INSPIRAÇÕES METODOLÓGICAS

Dois ações são mais importantes do que constatar o cerco das plataformas *online* às nossas práticas sociais, em especial, à escrita em ambiente digital: 1) observar os usos que fazemos desses aparatos tecnológicos agora intrínsecos às atividades rotineiras, inclusive, destacando os usos incomuns – viciosos demais ou subversivos; e 2) estudar processos de produção, incluindo, veementemente, as materialidades das plataformas e se afastando da tendência de olhar para a mensagem e não para o processo. Com foco nessas duas ações, vamos respaldar nosso olhar na “[...] compreensão da relação entre mídia e tecnologia como um fenômeno complexo e sociomaterial, pensando nas formas que a materialidade, a prática e as políticas estão necessariamente interligadas” (JURNO, 2020, p. 27), buscando amparo nos Estudos de Plataformas (*Platforms Studies*), Estudos de Algoritmos (*Critical Algorithm Studies*) e Métodos Digitais (*Digital Methods*), que têm matrizes conceituais originárias dos Estudos de Ciência e Tecnologia (*Science and Technology Studies – STS*). Recorremos também aos Estudos de Busca na Web (*Web Search Studies*), que, por sua vez, se alimentam de ideias do campo de Recuperação da Informação (*Information Retrieval – IR*).

Esse conjunto de inspirações anda com o anseio de estudar o texto plataformizado como forma registrada, no entanto, reconhecendo-o como emaranhando irreconstituível ou quase impossível de se enxergar o todo. Por isso, como fundamento-mestre, está a Linguística Textual associada às questões das tecnologias, que embasou a discussão desenvolvida no Capítulo 2 desta tese. Tudo aqui é ponto de partida, compreensivamente limitado, porque não há meios definitivos de estudar a produção total de textos em ambiente digital, nem representar processos, dinâmicas técnicas e sociais de circulação e recepção por inteiro. Mckenzie (2018, p. 25), completa esse argumento destacando que a pesquisa sobre texto deve

“[...] se preocupar em demonstrar que as formas afetam o significado”, descrevendo não apenas os processos técnicos, mas os processos sociais da circulação.

Em várias “formas físicas” as textualidades podem ser vistas, na construção de uma história do texto, por meio de seus processos de transmissão, percepções de significado e efeitos sociais. Qualquer história do texto que excluísse o estudo das motivações sociais, econômicas e políticas da publicação, as razões pelas quais os textos foram escritos, lidos desta ou daquela maneira, o porquê de terem sido reescritos e redesenhados ou deixados morrer, e degeneraria em uma frágil e depressiva lista de livros jamais se tornaria uma história legível. (MCKENZIE, 2018, p. 25).

As plataformas *online*, realmente, desafiaram a epistemologia, principalmente para pesquisadores interessados em condutas, e não conteúdos. Os meios e ambientes se mostraram duradouros como categorias, criando suas próprias tradições de pesquisa, que não estão contempladas em estudos de televisão, rádio, telecomunicações e comunicações mediadas por computador (PLANTIN et al., 2016). Por isso, considerando a dimensão material da “produção, circulação e leitura” de textos plataformizado, enxergamos as implicações da inserção da lógica das plataformas de busca como um objeto de pesquisa em construção, provisório, reeditável, assim como são as próprias plataformas e os algoritmos que as regem.

3.1.1 Estudos de Plataformas e Estudos de Algoritmos

A grande concentração de dinheiro nas cinco empresas – *Google/Alphabet Inc.*, *Amazon*, *Apple*, *Facebook* e *Microsoft* – está no centro das discussões dos Estudos de Plataformas, principalmente ao gerar produções científicas que debatem regimes de poder, práticas e materialidades dos produtos e serviços ofertados por essas companhias.

A *Forbes Global 2000*⁵⁵ classifica as maiores empresas do mundo usando quatro métricas: vendas, lucros, ativos e valor de mercado. A partir desses critérios, as componentes do grupo *Big Five* figuravam nas primeiras posições, em 2022, seja na lista geral de empresas ou no recorte pelo setor de tecnologia. Outro modo de enxergar o poderio do grupo é pelo *ranking* mundial de marcas mais valiosas, feito pela empresa especializada *Kantar*⁵⁶. As marcas são, normalmente, os maiores ativos que as empresas possuem, agregando enorme

⁵⁵ Disponível em: <https://www.forbes.com/lists/global2000/?sh=11c0d6755ac0>. Acesso em: 26 ago. 2022

⁵⁶ Disponível em: <https://www.kantar.com/en-cn/inspiration/brands/2022-kantar-brandz-top-100-most-valuable-global-brands>. Acesso em: 26 ago. 2022.

valor ao balanço patrimonial. O *ranking* quantifica a contribuição das marcas para o desempenho financeiro do negócio.

Quadro 4 – Posição das *Big Five* no mercado mundial

	Posição no ranking de maiores empresas do mundo	Posição no ranking de maiores empresas de tecnologia do mundo	Posição no ranking mundial de marcas mais valiosas	Valor de mercado	Valor da marca
<i>Apple</i>	7	1	1	US\$ 2,6 trilhões	US\$ 947 bilhões
<i>Amazon</i>	6	*	3	US\$ 1,46 trilhões	US\$ 705 bilhões
<i>Google/Alphabet Inc.</i>	11	2	2	US\$ 257,5 bilhões	US\$ 819 bilhões
<i>Microsoft</i>	12	3	4	US\$ 367,3 bilhões	US\$ 611 bilhões
<i>Meta/Facebook</i>	34	6	8	US\$ 499,9 bilhões	US\$ 186 bilhões

Fonte: Elaborado pela autora com base em *Forbes Global* (2022) e *Kantar* (2022).

Legenda: **Amazon* não é classificada pela *Forbes* como empresa de tecnologia.

Essas empresas só ocupam tal lugar de importância mercadológica⁵⁷ porque conseguiram se entranhar na vida humana a partir de produtos e serviços projetados e criados como sistemas sociotécnicos, aos quais damos o nome de plataformas. Como vemos no Quadro 4, elaborado com informações do Departamento de Justiça dos Estados Unidos, por meio da Divisão Antitruste, essas companhias foram capazes de firmar presença de maneira muito diversificada, especialmente *Google/Alphabet Inc.*, que têm produtos em todas as áreas apontadas:

Quadro 5 – Concorrência entre empresas de tecnologia

Produtos	<i>Amazon</i>	<i>Apple</i>	<i>Google</i>	<i>Facebook</i>	<i>Microsoft</i>
Sistema operacional	x	x	x		x
<i>Browser</i>		x	x		x

⁵⁷ Há sinais de uma desaceleração do poder das *Big Five* em 2023, quando a alta dos juros nos EUA resultou na redução de gastos com publicidade por boa parte do empresariado estadunidense, o que afeta economicamente as *Big Five*. *Meta*, *Amazon*, *Microsoft* e *Alphabet (Google)* demitiram mais de 60 mil funcionários neste ano. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/03/14/big-techs-demissao-em-massa-em-5-meses.ghtml>. Acesso em: 31 mar. 2023.

Produtos	Amazon	Apple	Google	Facebook	Microsoft
Ferramentas de produtividade		X	X		X
E-Mail e mensagem		X	X	X	X
E-Books	X	X	X		
Smartphones e dispositivos	X	X	X		X
Distribuição de música e vídeo	X	X	X		
Streaming de vídeo	X		X	X	
Videoconferência		X	X	X	X
Serviços de entrega em domicílio	X		X		
Buscador de uso geral			X		X
Buscador de uso especial	X	X	X		X
Assistente digital	X	X	X	X	X
Rede social			X	X	
Plataforma de publicidade	X		X	X	X
Mapa		X	X		X
Serviço de nuvem	X		X		X

Fonte: Elaborado pela autora com base em *The United States Department of Justice. Antitrust Division (2022)*.

Por dois motivos, os Estudos de Plataforma, como uma área interdisciplinar, ajudam a construir o sentido de nomear essas estruturas como plataformas, e não manter outras nomenclaturas, a exemplos de “rede social” e “sites”. O primeiro motivo é o olhar para a questão infraestrutural desses artefatos tecnológicos, que se mostram cada vez mais centrais nos fluxos financeiros e informacionais, além de centrais nas práticas comunicacionais e de trabalho nas ambiências digitais. As infraestruturas das plataformas são “[...] constituidoras dos processos” (D’ANDRÉA, 2020, p. 33) e articulam dependências interativas “[...] que criam potenciais para o caos social durante episódios de colapso” (PLANTIN et al., 2016, p. 5, tradução nossa)⁵⁸. Basta pensar que qualquer falha nos serviços do *Google*, por exemplo, interromperia substancialmente a vida diária e o trabalho.

As infraestruturas são “aprendidas como parte da adesão” em comunidades, nações e mundos da vida (Bowker e Star, 1999; Edwards, 2002). Esse fluxo de estudos de infraestrutura enfatiza o papel crítico dos elementos humanos da infraestrutura, como práticas de trabalho, hábitos individuais e cultura organizacional, bem como as maneiras pelas quais uma infraestrutura pode excluir estruturalmente algumas

⁵⁸ Tradução nossa para: “[...] which create potentials for social chaos during episodes of breakdown”.

peças (por exemplo, pessoas com deficiência auditiva, visual ou cadeirantes) de serviços supostamente “universais” (Lee et al., 2006; Ribes e Bowker, 2008; Ribes e Finholt, 2009). (PLANTIN et al., 2016, p. 5, tradução nossa)⁵⁹.

Ainda com o exemplo do *Google*, Plantin et al. (2016) explica que um aplicativo como o *Google Maps* pode ser considerado uma plataforma programável na qual usuários e desenvolvedores podem construir novos objetos digitais. Já o serviço do *Google* de pesquisa na *Web*, que se tornou tão onipresente, pode ser visto como uma infraestrutura robusta, acessível e essencial. “O que é o *Google*, então: uma plataforma? Uma infraestrutura? É sequencialmente ou simultaneamente ambos?” (PLANTIN et al., 2016, p. 3, tradução nossa)⁶⁰.

Na verdade, segundo definição dos autores, o *Google* é uma plataforma infraestrutural, que detalharemos no Capítulo 4, pois concentra recursos materiais que controlam o acesso à internet e nos fazem esquecer usos fora de seus leques de serviços. Plantin et al. (2016) se dedicaram a comparar os Estudos de Plataformas com outra vertente conceitual, chamada Estudos de Infraestrutura, para construir as noções de “plataformização das infraestruturas” e “infraestruturização das plataformas”, argumentando uma certa consolidação de serviços particulares que articulam a “vida online” das pessoas.

Segundo esses pesquisadores, as tecnologias digitais possibilitaram alternativas de menor custo, mais dinâmicas e mais competitivas às infraestruturas monopolistas governamentais ou quase-governamentais, em troca da transferência de riqueza e responsabilidade para empresas privadas. Além disso, empresas como o grupo *Big Five* aprenderam a explorar o poder das plataformas – que trazem benefícios inegáveis para usuários e desenvolvedores de aplicativos menores e independentes – para ganhar pontos de apoio, como os equivalentes da ferrovia, telefone e energia elétrica que reinaram final dos séculos XIX e XX (PLANTIN et al., 2016).

Retomando, o segundo motivo pelo qual os Estudos de Plataforma ajudam a construir o sentido dado às plataformas *online* é o olhar para a existência de recursos que incluem programabilidade, conexão de atores heterogêneos e acessibilidade de dados e lógica por meio de interfaces de programação de aplicativos (APIs). Para os Estudos de Plataforma,

⁵⁹ Tradução nossa para: “As a result, infrastructures are “learned as part of membership” in communities, nations, and lifeworlds (Bowker and Star, 1999; Edwards, 2002). This stream of infrastructure studies emphasizes the critical role of infrastructure’s human elements, such as work practices, individual habits, and organizational culture, as well as the ways an infrastructure can structurally exclude some people (e.g. deaf, blind, or wheelchair-bound individuals) from purportedly “universal” services (Lee et al., 2006; Ribes and Bowker, 2008; Ribes and Finholt, 2009).”

⁶⁰ Tradução nossa para: “What is *Google*, then: a platform? An infrastructure? Is it sequentially or simultaneously both?”.

o termo plataforma é usado para nomear os serviços organizadores de um certo ecossistema diversificado de parceiros, permitindo aos usuários programar indo além do projeto inicial dos *designers*. De acordo com Plantin et al. (2016), os construtores de plataformas não buscam internalizar seus ambientes por meio da integração vertical, em vez disso, suas plataformas são projetadas para serem estendidas e elaboradas de fora, por outros atores, desde que esses atores sigam certas regras. Esses desenvolvedores não são mais meros construtores de sistemas, porque alavancam a programabilidade e a interconexão para obter controle.

Faz parte da essência das plataformas a interrelação contínua entre usuários e os fins lucrativos da empresa detentora. Por isso, novas formas de poder são geradas pelas plataformas, porque elas rompem sistemas tradicionais, a exemplo da lógica de anúncios que sustentou o jornalismo por séculos e que o *Google*, por exemplo, subverteu ao se tornar um ambiente de anúncios *online* direcionados a usuários individuais.

Os pesquisadores de Estudos de Plataforma exploram como modularidade e poder são negociados entre uma unidade central com baixa variabilidade e componentes heterogêneos de alta variabilidade. Sua perspectiva é cultural, econômica e crítica, formando um *continuum* que vai dos estudos culturais à economia política. Coletivamente, eles destacam como as *affordances* das plataformas simultaneamente permitem e restringem a expressão, bem como as preocupações técnicas, sociais e econômicas determinam a estrutura, função e uso das plataformas. (PLANTIN et al., 2016, p. 7, tradução nossa)⁶¹.

Bogost e Montfort (2009) desfazem equívocos sobre os Estudos de Plataforma, além de construírem uma concepção que rechaça críticas infundadas. Para os autores, os Estudos de Plataforma se opõem ao determinismo tecnológico e convidam a continuar a abrir a caixa-preta da tecnologia de maneira produtiva; incluem estudos sobre todas as plataformas de computação nas quais um trabalho criativo tenha sido feito; conectam detalhes técnicos à cultura; mostram como a compreensão técnica pode levar a novos tipos de *insights*, sem expulsar outros tipos de pesquisas sobre mídia digital. Enfim, os autores postulam que os Estudos de Plataforma podem enfatizar diferentes aspectos técnicos ou culturais e se basear em diferentes abordagens críticas e teóricas, mas estarão unidos por investigarem profundamente os sistemas de computação em suas interações com criatividade, expressão e cultura.

⁶¹ Tradução nossa para: “Platform studies scholars explore how modularity and power are negotiated between a core unit with low variability and heterogeneous components of high variability. Their perspective is cultural, economic, and critical, forming a continuum ranging from cultural studies to political economy. Collectively, they highlight how platforms affordances simultaneously allow and constrain expression, as well as how technical, social, and economic concerns determine platforms’ structure, function, and use.”

Em complementariedade, estão os Estudos de Algoritmos que destacam a necessidade de olhar para os algoritmos ultrapassando os elementos técnicos, sob o risco de nos agarrarmos a um “imaginário algorítmico” (BUCHER, 2017) ou uma “abstração mistificada” (GILLESPIE, 2016) ou, ainda, a uma obscurecência dos valores sociais e políticos que se inserem no conceito geral de algoritmos ao nos referimos às plataformas *online*. Por isso, são estudos “críticos”, que tratam os algoritmos como um complexo conjunto sociotécnico e pretendem desmistificar. Afinal, seria muito mais fácil nos fixarmos acriticamente sobre o atraente sentido único de algoritmos, mas não é isso que pretendemos.

Todos os problemas clássicos que enfrentamos ao tentar descompactar uma tecnologia, o termo empacota para nós. Torna-se muito fácil tratá-lo como um único artefato, quando nos casos em que estamos mais interessados, raramente é um algoritmo, mas muitas ferramentas funcionando juntas, às vezes ferramentas diferentes para usuários diferentes Também tende a apagar as pessoas envolvidas, minimizar seu papel e distanciá-las da responsabilidade. (GILLESPIE, 2014, n.p., tradução nossa)⁶².

Gillespie (2016) argumenta que diversas áreas do conhecimento usam o termo algoritmo de maneiras diferentes, por isso é melhor que estejamos cientes dos múltiplos significados do termo. Para o público mais amplo, algoritmos nomeiam algo inatingível e complexo. Para os engenheiros de *softwares*, algoritmos são uma série lógica de etapas para organizar e agir em um corpo de dados para alcançar rapidamente um resultado desejado, ou seja, uma receita composta em passos programáveis. Já os cientistas sociais estão mais dispostos a proclamar as implicações dos algoritmos do que a dizer o que são, por isso se importam muito com os valores embutidos no uso social feito de ferramentas geridas por algoritmos; preocupam-se em ver como os aspectos sociológicos atravessam os aspectos técnicos, inserindo mais do que um problema a ser resolvido por um programa previamente modelado ou operacionalizado.

Há também uma “aura” que envolve o termo algoritmo como sendo uma cartada de legitimação técnica a processos antes executados por humanos e que hoje estão aparentemente modernizados por sistemas. Frases como “isso aqui é calculado por um algoritmo” são usadas como talismã para celebrar a confiabilidade do cálculo automatizado. Ouvimos também “é culpa do algoritmo”, para justificar ou explicar erros e resultados indesejados e defender os papéis cada vez mais significativos que empresas privadas

⁶² Tradução nossa para: “All the classic problems we face when trying to unpack a technology, the term packs for us. It becomes too easy to treat it as a single artifact, when in the cases we’re most interested in it’s rarely one algorithm, but many tools functioning together, sometimes different tools for different users. It also tends to erase the people involved, downplay their role, and distance them from accountability.”

desempenham na vida pública. Atentemos-nos ao fato de que algoritmos não são entidades separadas, eles são, na realidade, materiais manejados por pessoas.

Ademais, o termo é usado para representar o “dono do negócio”, quando vamos nos referir a escolhas e intervenções feitas por uma corporação em nossas práticas sociais, querendo atribuir alguma responsabilidade, como em: “O algoritmo do Google escolhe respostas”. Aqui na tese, por vezes, usamos o termo nesse sentido, mas sem negligenciar que é a companhia *Google/Alphabet Inc.*, dotada de poder político e econômico, “[...] agindo por meio de seu algoritmo, intervindo de forma algorítmica, construindo um negócio justamente em sua capacidade de construir modelos complexos de atividade social/expressiva” (GILLESPIE, 2014, n.p., tradução nossa)⁶³.

Falamos muito sobre o termo “algoritmo”, mas o que realmente nos preocupa quando invocamos esse conceito não é a regra programada em si, mas a inserção dela na “produção, circulação e leitura”, na formação de conhecimento humano e na experiência social, em especial, quando falamos das plataformas de busca. Rieder (2020) evoca várias análises sobre mecanismos de busca – de pesquisadores como Lucas Instrona, Nicholas Diakopoulos, Astrid Mager, Rob Kitchin, entre outros – para dizer que sempre se observou a dimensão política dos algoritmos em medidas de classificação, padronização, julgamento, agregação, ranqueamento, gestão e governamentalidade da informação. Aliás, muitos desses autores analisaram responsabilidades e ideologias envolvidas no ato de mediar direitos e privilégios, gerando inclusões, exclusões e (i)mobilidades.

Pensando criticamente sobre as pesquisas que abordam os algoritmos, Kitchin (2016) também argumenta sobre as várias maneiras que os algoritmos podem ser estudados: tecnicamente, computacionalmente, matematicamente, politicamente, culturalmente, economicamente, contextualmente, materialmente, filosoficamente, eticamente. Mas o autor alerta para o fato de que algoritmos são mais bem entendidos a partir de sua natureza contingente e performativa e quando incorporados em conjuntos sociotécnicos. Observando os desafios de se pesquisar algoritmos – pela inacessibilidade aos códigos, pelo caráter contingencial e pela complexidade de envolver pesquisas práticas e robustas epistemologicamente –, Kitchin (2016) sugeriu seis abordagens metodológicas para olhar o fenômeno como um complexo conjunto sociotécnico e ampliar o pensamento crítico emergente: 1) análise de pseudocódigo e/ou sua construção no código-fonte; 2)

⁶³ Tradução nossa para: “[...] acting through its algorithm, intervening in an algorithmic way, building a business precisely on its ability to construct complex models of social/expressive activity, train on an immense corpus of data, tune countless parameters, and reach formalized goals extremely efficiently”

autoetnografias de produção de algoritmos; 3) engenharia reversa; 4) entrevista ou etnografia com desenvolvedores; 5) descompactar o conjunto sociotécnico de algoritmos; e 6) examinar como os algoritmos funcionam no mundo. Vamos destrinchar as três abordagens que mais nos inspiram:

Quadro 6 – Abordagens metodológicas para Estudos de Algoritmos

Abordagem metodológica	Como funciona
Engenharia reversa	<p>Testar entradas (<i>inputs</i>) e saídas (<i>outputs</i>) e examinar combinações para refletir sobre o que o algoritmo entrega ou faz em cada entrada. Articular as especificações de um sistema por meio de um exame rigoroso com base no conhecimento de domínio, observação e dedução para desenterrar um modelo de como esse sistema consiste.</p> <p>Acompanhar debates em fóruns <i>online</i> de usuários sobre como eles percebem que um algoritmo funciona ou mudou, ou entrevistar profissionais de <i>marketing</i>, estrategistas de mídia e empresas de relações públicas que buscam jogar com algoritmos para otimizar resultados para um clientes.</p>
Descompactar o conjunto sociotécnico de algoritmos	<p>Investigar formas de conhecimento, legalidades, governamentalidades, instituições, mercados, finanças. Analisar documentos da empresa, material promocional/indústria, concursos públicos e estruturas legais e de padrões. Examinar práticas, estruturas e comportamento das instituições; e olhar para biografias dos principais atores e as histórias dos projetos, o que ajudará a revelar como os algoritmos são imaginados e narrados.</p>
Examinar como os algoritmos funcionam no mundo	<p>Saber como os algoritmos tornam a vida cotidiana diferente, observando seu trabalho no mundo sob diferentes condições. Realizar etnografias de como as pessoas se envolvem com e são condicionados por sistemas algorítmicos e como esses sistemas remodelam organizações conduzem seus esforços e são estruturadas. Isto também exploraria as maneiras pelas quais as pessoas resistem, subvertem e transgridem os algoritmos, redefinindo-os e reimplantando-os para fins que não eram originalmente pretendido.</p>

Fonte: Elaborado pela autora com base em Kitchin (2016).

A abordagem “engenharia reversa” tem grande influência na etapa metodológica que denominamos “espaço simulado de escrita e edição”, porque testamos, examinamos, editamos textos para torná-los mais elegíveis a algoritmos. Portanto, simulamos a entrada de um texto na *Web* parametrizando essa produção para o que deduzimos ser o modelo mais ideal para a elegibilidade em plataformas de busca. Esse método, inspirado na “engenharia reversa”, nos ajudou a perceber como práticas de edição textual são incorporadas e alteram o resultado daquela redação em prol da performance e visibilidade em ambiente digital. Ademais, a abordagem “descompactar o conjunto sociotécnico de algoritmos” está muito presente na caracterização do *Google* como plataforma infraestrutural e no levantamento de

uptades que tiveram implicações diretas na reconfiguração de práticas de escrita e edição de textos. Por fim, a abordagem “examinar como os algoritmos funcionam no mundo” nos motivou apresentar a discussão e descrever exemplos de como produtores de texto usam taticamente, resistem, subvertem e transgridem os algoritmos do *Google*.

Usar como inspiração essas abordagens, que fornecem uma visão mais ampla e contextualizada dos algoritmos, posiciona-nos como questionadores em meio à “[...] crise moral dos sistemas atuariais” (OCHIGAME, 2021b, n.p.). Os sistemas algorítmicos se tornaram onipresentes na era da computação digital, bem como uma abordagem altamente instrumental da estatística e do aprendizado de máquinas, que o historiador Matthew Jones (2018) denomina “positivismo de dados”. Essa vertente, por sua vez, expressa-se em ações que reduzem o fenômeno complexo e ambíguo das interações humanas à noção muito mais contida de “execução de planos”. Ademais, discursivamente, o “positivismo de dados” se expressa em termos como “Big Data”, “inteligência artificial”, “mineração de dados” e “tomada de decisão automatizada”, muito usados pelas empresas de tecnologia e para os quais olhamos criticamente. Faz lembrar o quanto, durante anos, a Teoria Matemática da Comunicação – de Shannon e Weaver⁶⁴ – vigorou como modelo linear possível para entender fenômenos de interação verbal e que hoje, sabemos, o quanto a modelização era simplista.

Guiar-nos por esses fundamentos faz sentido pelo fato de os Estudos de Algoritmos e os Estudos de Plataforma estarem contidos nos Estudos de Ciência e Tecnologia (STS), um campo transdisciplinar, que tem se voltado para o estudo dos regimes de poder, das práticas situadas e das materialidades que constituem a produção científica e os artefatos tecnológicos. Os pesquisadores dessa área evitam se enredar pelo “determinismo tecnológico” e vão além da noção de “construção social da tecnologia”, ao assumir como premissa que artefatos tecnológicos e práticas sociais se coproduzem. Diz respeito a usar métodos de pesquisa para explicar como as pessoas se desenvolvem e, em seguida, usam tecnologias específicas, afinal, “[...] as tecnologias não surgem do nada; em vez disso, os humanos os criam em resposta a efeitos sociais e culturais, ideais, objetivos, problemas ou fatores relacionados” (BOGOST; MONTFORT, 2009, p. 3)⁶⁵.

É como pensar que, em meio a um complexo e assimétrico jogo de poder dessas plataformas, os usuários e as materialidades se constituem mutuamente. Isso porque

⁶⁴ Teoria que inspirou os estudos em Comunicação, que foi reproduzida e, posteriormente, refutada. Claude E. Shannon e Warren Weaver descreveram elementos básicos da comunicação: fonte, transmissor, canal, receptor.

⁶⁵ Tradução nossa para: “[...] technologies do not—and cannot—arise from thin air; rather, humans create them in response to social and cultural effects, ideals, goals, problems or related factors.”

pretendemos ir além de pensar como as tecnologias surgem, mas como as pessoas respondem a elas. Uma das linhas de pesquisa mais conhecidas do campo STS é Teoria Ator-Rede (TAR), uma abordagem semiótica-material que reúne estudos sobre inovações e controvérsias tecnocientíficas, incorporando como preocupação essencial as dimensões coletiva, performativa e negociada de um “social” que emerge de relações cada vez mais imbricadas entre diferentes actantes humanos e não humanos. (D’ANDRÉA, 2020).

Para estudar a plataformização do texto, a acomodação dessas duas lentes – Estudos de Plataforma e Estudos de Algoritmos – nos ajudam a compreender histórica e criticamente como a sociedade está se transformando. Não se trata somente de “[...] atribuir aos artefatos tecnológicos o poder de controlar as práticas contemporâneas” (D’ANDRÉA, 2020, p. 17), mas de reconhecer que eles remodelam a experimentação humana com os textos e com as linguagens.

3.1.2 Métodos digitais

Ao nos dedicarmos a pesquisar os entrelaçamentos entre “produção, circulação e leitura” de textos e as materialidades da plataforma *Google*, vamos levantar as especificidades do meio considerando-as e problematizando-as em cada etapa da pesquisa, de modo a torná-la reflexiva. Esse modo de pesquisa é inspirado nos Métodos Digitais, um conjunto metodológico que ficou conhecido entre os pesquisadores pelo jargão “siga o meio”, uma referência a Teoria Ator-Rede, na intenção de desdobrar as interações pensando nos objetos mediadores não neutros e constituintes de uma rede instável.

Meios e mediadores não deixam uma passagem neutra entre *input* e *output*, para usar aqui temos muito comuns entre programadores e entusiastas dos mecanismos de busca. Meios ou mediadores transformam, traduzem, modificam, deixando rastros que podem ser seguidos. Um modo de pesquisar ambiente digital é olhar para esses rastros exemplificando funcionamentos sociais gerais, sem tentar categorizar, fixar ou estabilizar os meios e as informações emergentes, afinal, o “[...] on-line se tornou a referência com a qual se poderia julgar a extensão de uma condição social percebida (ROGERS, 2016, p. 9)”. D’Andréa (2020) ressalta a importância de “seguir os meios”:

Isto é de desenvolver olhares analíticos e procedimentos metodológicos sensíveis às lógicas que singularizam os mecanismos de busca, modelos de ranqueamento e outros recursos online. Seguir os meios, nesse caso, é compreender, de modo crítico, as lógicas sociotécnicas de uma plataforma, ou seja, suas maneiras de classificar e

apresentar informações, o modo como os usuários são incentivados a construir seus perfis, como a interface é pensada para se coletar mais dados etc. (p. 61).

Mesmo que em nosso método de pesquisa não façamos a extração de dados da plataforma *Google*, consideramos que há inspirações nos Métodos Digitais. Os *plugins* que escolhemos trabalhar em uma das etapas metodológicas – *Yoast SEO e Rank Math* – emulam o funcionamento do *Google*, ou seja, usamos esses *plugins* como “espelho” da métrica de relevância criada pela plataforma de busca. Seria difícil dizer que nossa pesquisa usa os Métodos Digitais, porque ela é puramente qualitativa. No entanto, fomos além de posicionar instrumentalmente uma realidade que existe na “produção, circulação e leitura de textos”, portanto, demonstramos “[...] formas de saber que fazem emergir leituras e visões de mundo atrelados às lógicas de datificação, às mediações algorítmicas, às políticas de governança e a outras dimensões das plataformas online” (D’ANDRÉA, 2020, p. 62). Isso ocorre, por exemplo, quando pensamos em como Termos de Serviço e Princípios de Privacidade e Segurança do *Google* – documentos de autorregulação e governança – incidem nas condições em que textos são produzidos e circulam ou em quando refletimos no surgimento de gêneros textuais a partir do contingenciamento da produção para atender a demandas algorítmicas dessa plataforma de busca.

Rogers (2016) afirma que Métodos Digitais querem conhecer não apenas o conteúdo que circula nos meios, mas como os meios os ajudam a construir aquele conteúdo “circulante”. Aliás, ele propõe que uma das abordagens possíveis para pesquisar a plataforma *Google* é descobrir as diferentes ordenações de fontes de pesquisa, que servem para hierarquização de respostas e informações, mas essas ordenações nunca são disponibilizadas. Se nos propomos a estudar o *Google*, não podemos perder de vista como o *Google* funciona.

3.1.3 Estudos de Busca na Web

Por tudo que argumentamos até agora, a ordenação algorítmica de informações é, claramente, um meio de automatizar as decisões em ambientes digitais. No entanto, acaba sendo apresentada como uma solução para o “problema” da sobrecarga de informações. Esse é um terreno fértil para uma ferramenta de busca como *Google* se estabelecer em poderes diversos – simbólico, cultural, técnico e econômico. Rieder (2020) relaciona a disseminação da busca de informações com a indústria da recuperação da informação, posicionando-se criticamente diante do discurso neoliberal das *Big Five* de que grandes problemas da humanidade estão sendo resolvidos pelas plataformas *online*. Pesquisar, ou seja, o ato de

colocar uma consulta em um campo de formulário, tornou-se um gesto tão fundamental “[...] que uma caixa de pesquisa ausente em um site se torna uma experiência quase perturbadora” (RIEDER, 2020, p. 9, tradução nossa)⁶⁶. A busca por computador se expandiu para além das bibliotecas, arquivos e sistemas de documentação especializados aos quais estava amplamente confinada antes da chegada da *Web*. Por isso, na expressão “recuperar informação”, o termo “informação” passou a representar quase tudo, de fragmentos de conhecimento a coisas, produtos, pessoas, ideias ou experiências.

Portanto, os motores de busca se tornaram “[...] meio dominante para acessar e ordenar as massas de bits e peças digitais e datificadas que desordenam os ambientes em que habitamos” (RIEDER, 2020, p. 9, tradução nossa)⁶⁷. Conforme argumenta o pesquisador, o ato deliberado de formular uma consulta para encontrar algo – como no caso dos buscadores – é apenas uma das muitas formas em que a recuperação da informação hoje se manifesta. Personalização automatizada, localização, recomendação, filtragem, classificação, avaliação, agregação, síntese são práticas igualmente difundidas que não requerem entrada explícita do usuário para selecionar, sequenciar, organizar ou modular algum conjunto de itens digitais.

E as técnicas de recuperação não se limitam mais à produção de listas de resultados: geram pontuações, sugerem itens, descartam ou promovem mensagens, estabelecem preços, organizam objetos e pessoas em relação uns aos outros, montam textos, proíbem ou concedem acesso, fabricam interfaces e visualizações, e até mesmo dirigir objetos no mundo físico. (RIEDER, 2020, p. 10, tradução nossa)⁶⁸.

Daí a necessidade de abordagens metodológicas capazes de compreender, além das questões tecnológicas das plataformas de busca, as perspectivas políticas, éticas e culturais que modificam o acesso ao conhecimento, privacidade, concentração de mercado e negócios, e que podem limitar a possibilidade de servir ao interesse público. Os Estudos de Busca na *Web* estão centrados em pesquisas sobre motores de busca em perspectivas variadas, como fundamentos técnicos; precisão, confiabilidade e desempenho; análises de *log* de transações para aprender sobre a interação dos usuários (tempo gasto em tarefas de pesquisa, por exemplo); seleção de resultados; estudos de usuários, a fim de entender comportamento

⁶⁶ Tradução nossa para: “[...] that a missing search box on a website becomes an almost disturbing experience”.

⁶⁷ Tradução nossa para: “[...] a dominant means to access and order the masses of digital and datafied bits and pieces that clutter the environments we inhabit”.

⁶⁸ Tradução nossa para: “And retrieval techniques are no longer limited to producing result lists: they generate scores, suggest items, discard or promote messages, set prices, arrange objects and people in relation to each other, assemble texts, forbid or grant access, fabricate interfaces and visualizations, and even steer objects in the physical world.”

de busca e as necessidades; críticas políticas, éticas e culturais; análises legais que discutem a tentativa de regular ou legislar as práticas dos motores de busca.

Olhando para essas abordagens de pesquisa já existentes, Zimmer (2009) sugere quatro enfoques metodológicos complementares: 1) viés do mecanismo de pesquisa; 2) motores de busca como *gatekeepers* da informação; 3) valores e ética dos motores de busca; e 4) restrições e obrigações legais. Muitas pesquisas após 2010 seguiram por alguns desses caminhos, haja vista Rieder (2020), Plantin et al. (2016), Medina (2015), Ochigame (2021a), entre outros. Rodrigo Ochigame (2021a), por exemplo, dedicou-se a construir motores de busca alternativos a partir de uma visão crítica que demonstra consequências políticas dos modelos hegemônicos de busca. Ele produziu experimentos de sistemas e interface alternativos aos serviços comerciais (*Google, Bing, Yahoo*), nomeou de buscadores divergentes e os programou para apresentarem resultados orientados pelo interesse público e que quebram as lógicas de propaganda direcionada. Vamos destrinchar os dois enfoques de Zimmer (2009) que mais nos inspiram:

Quadro 7 – Abordagens metodológicas para Estudos de Busca na *Web*

Abordagem metodológica	Como funciona
Motores de busca como <i>gatekeepers</i> da informação	Foco na redução da opacidade sobre o funcionamento dos motores de busca na <i>Web</i> , identificando se existem funções intencionais de <i>gatekeeping</i> , como a intenção de censurar certos resultados de pesquisa
Valores e ética dos motores de busca	Preocupações com preconceitos e pontos de controle às maneiras pelas quais os mecanismos de busca na <i>Web</i> têm valor particular e implicações éticas para a sociedade. Entender quais valores estão em jogo com a pesquisa na <i>Web</i> , como privacidade, autonomia e liberdade, e como as atividades de pesquisa dos usuários realmente afetam os a forma como experimentam no mundo real.

Fonte: Elaborado pela autora com base em Zimmer (2009).

A abordagem “motores de busca como *gatekeepers* da informação” influenciou nossas investigações sobre o funcionamento e as políticas da plataforma *Google*, particularmente, a análise das principais patentes em *Core Updates* de algoritmos e tutoriais de operação do motor de busca, feita no Capítulo 4. Foi uma tentativa de tornar menos opacas as etapas de seleção e ordenação da informação feitas pela plataforma. Em complementaridade, a abordagem “valores e ética dos motores de busca” se faz presente por toda a pesquisa, na medida em que argumentamos as consequências do atravessamento das plataformas de busca na “produção, circulação e leitura” de textos, em especial, ao

analisarmos, no Capítulo 5, as implicações que a reconfiguração da escrita e edição tem para uma era de textos circulantes em ambiente digital.

Enfim, uma das contribuições de Estudos de Busca na *Web* que prezam por analisar a tecnologia como um fenômeno complexo e sociomaterial é revisar processos de recuperação da informação de modo crítico. De acordo com Rieder (2020), a definição clássica de recuperação da informação sustenta que um sistema de busca não informa ou não altera o conhecimento do usuário sobre o assunto que foi pesquisado, mas meramente avisa sobre a existência (ou não existência) e o paradeiro dos documentos relacionados ao tema. É isto que lemos nas entrelinhas da missão do *Google*:

A missão do Google é organizar as informações disponíveis no mundo e torná-las acessíveis e úteis para todas as pessoas. É por isso que a Busca facilita a descoberta de vários tipos de informação em uma variedade de fontes. Algumas informações são simples, como o tamanho da Torre Eiffel. No caso de temas mais complexos, a Busca é uma ferramenta para explorar vários ângulos e permitir que você compreenda o mundo do seu jeito. (GOOGLE, 2022d).

Porém, não enxergamos a versão da história dessa maneira. Apesar de claramente modular o conhecimento, plataformas de busca ficam confortavelmente em um espaço de recuperadores da informação, mesmo deixando sinais de que pretendem fazer os buscadores semelhantes às mentes humanas. Há um desejo quase cômico das plataformas em não se envolver de modo sério com a noção de construção do conhecimento coletivo e se ater ao uso do senso comum.

3.2 MÉTODOS DESTA PESQUISA

Entre os Estudos de Plataformas, Estudos de Algoritmos, Métodos Digitais e Estudos de Busca na *Web* há em comum a vontade de observar, descrever e analisar criticamente as plataformas *online*. Em nosso recorte, todos esses métodos ajudam a ver as plataformas de busca como forças que transformam o ambiente digital, não como meras câmeras que registram passivamente os “[...] movimentos atuais da cultura escrita” (RIBEIRO, 2018b, p. 28). As pesquisas sob essas inspirações apresentam caráter dual da *Web* como recurso democratizante e, em contrapartida, colonizado por interesses especializados. Vai ficando mais clara uma lógica de interesse público descentralizado *versus* interesse comercial centralizado, e as plataformas de busca constituem um terreno revelador dessa dualidade.

Até aqui, construímos uma mirada sobre o entrelaçamento da escrita com o uso de plataformas, procurando mapear a literatura que dá conta dessa relação e pode nos nortear. Esse movimento nos ajudou a traçar os sentidos da pesquisa e, a partir daqui, vamos conseguir olhar para documentos e experimentos de modo crítico e analítico. Os documentos que regem a governança do *Google* vão registrar as materialidades dessa plataforma nas mediações de espaços de escrita e espaços de leitura. Ademais, as experiências que faremos serão fotografias das práticas incorporadas por produtores de texto que pretendem performar e ter visibilidade a partir das mediações do *Google*.

Tudo isso sob uma perspectiva de pesquisa e análise qualitativas. De acordo com Silverman (2009a), o ponto forte da pesquisa qualitativa é a capacidade de estudar fenômenos simplesmente indisponíveis em qualquer lugar. Os pesquisadores qualitativos estão preocupados em estabelecer correlações entre variáveis, reunindo dados que ocorrem naturalmente para encontrar sequências (“como”) em que os significados dos participantes (“o quê”) são exibidos e, assim, estabelecer o caráter de algum fenômeno. Silverman (2009a) representou essa lógica assim: [os quês → o fenômeno → os comos].

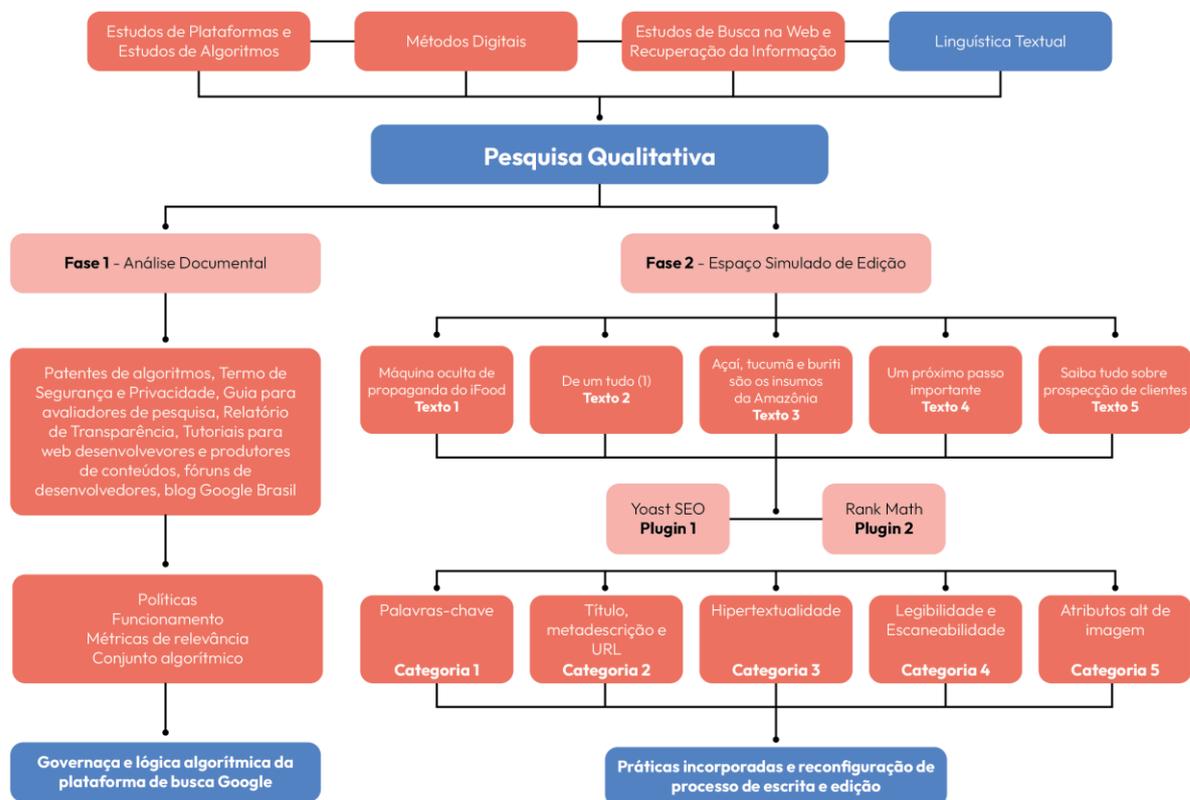
O autor ainda destaca um ponto que se alinha com a tese que construímos, ao argumentar que pesquisas qualitativas bem-sucedidas devem responder aos vários públicos eventualmente preparados para ouvir aquilo que temos a dizer (SILVERMAN, 2009b). Por isso, a escrita das análises qualitativas precisa abordar preocupações partilhadas entre esses públicos, de modo a relacionar a pesquisa com suas necessidades. Um dos nossos intentos com o desenho metodológico escolhido é justamente que a pesquisa sirva de referência para profissionais de produção e edição de texto tanto quanto pesquisadores das áreas de Linguística Textual ou Comunicação que desejem ser interlocutores das nossas abordagens.

Hissa (2012) ressalta o papel de intérprete do pesquisador nas análises qualitativas quando este organiza os dados e aplica a eles uma leitura de mundo que depende muito do lugar político de onde fala o sujeito-pesquisador. “O dado existe em razão da existência do sujeito, que constrói o seu modo de dizer o mundo. Sem o intérprete, o dado é silêncio ou inexistência” (HISSA, 2012, n.p.). Os dados da nossa pesquisa emergem em duas etapas: 1) análise em documentos do *Google* e 2) experiências em espaço simulado de escrita e edição. Mas essas duas fases só puderam acontecer por causa da pesquisa bibliográfica e revisão da literatura que nos deram condições de construir arcabouço para uma interpretação referenciada a conceitos.

A etapa 1 deu origem ao Capítulo 4 desta tese, onde está detalhadamente descrito o procedimento de análise documental. A etapa 2 deu origem ao Capítulo 5 desta tese, em que

estão os passos de criação e execução do espaço simulado de escrita e edição. Optamos por organizar o método de pesquisa com os resultados para facilitar a compreensão e dar unidade ao entendimento do leitor. A seguir, a Figura 6 traz o desenho metodológico da pesquisa, representando visualmente o percurso metodológico completo.

Figura 6 – Desenho metodológico



Fonte: Elaborado pela autora, 2023.

PARTE II: MATERIALIDADES E EXPERIÊNCIAS

A opacidade de algoritmos e do controle privado de plataformas leva os indivíduos a perceberem como uma esfera pública e aberta que, de fato, não passa de uma aparição privada e exclusiva. Como público e cidadãos não sabemos jamais se estamos vendo a mesma informação ou o que qualquer pessoa está vendo. Sem base comum de informação, pouco a pouco o debate público está se tornando impossível. (VILLAS BÔAS, 2022, p. 89).

Meu amado pai,

No grupo de Whatsapp da família, seu apelido é Google. Aliás, fora do grupo também, durante as festas e encontros só se escuta a galera te chamando pela alcunha. Você é um obcecado por informação, leitor e ouvinte voraz de notícias, um homem atento, detalhista, metódico e dono de uma invejável memória. Herdei de você a sede pelo texto. E você foge de herdar da minha avó (sua mãe, dona Rita) qualquer traço do Alzheimer que a levou aos 78 anos, afinal, a memória para você é um dom. Você costuma dizer a qualquer sinal de esquecimento: “Sai pra lá, alemão”.

No auge dos 69 anos, você é o único que lembra e faz o alerta no grupo sobre os aniversariantes do dia, incluindo aniversários de casamentos de uma galera. Costuma sinalizar, também, datas marcantes: x anos da morte de fulano, x anos que o ciclano formou em curso superior, obteve título de mestre e doutora (por enquanto, só temos uma na família. Espero ser a segunda, em breve).

“Bom dia, meu povo! Alguém deve estar pensando... o Google esqueceu. Esqueceu não. Só está meio desfusado devido ao recolhimento da quarentena. Há exatos 28 anos acontecia aquela tragédia. Coincidentemente, era um domingo. Descansem em paz”, escreveu você no dia 5 de abril de 2020, lembrando o acidente de carro que levou, de uma só vez, cinco membros da nossa família, em 1992.

Você não deixa passar em branco a contagem do tempo em que desembarcou em Belo Horizonte para começar a vida vindo de Mutum, na Zona da Mata de Minas Gerais. Faz questão de exhibir conhecimento sobre efemérides: a morte de Getúlio Vargas, o dia em que comemoramos o tetracampeonato com a seleção brasileira, a tríplice coroa do Cruzeiro, entre outros fatos políticos e futebolísticos, que são suas “editorias” preferidas.

É engraçado como a família crê no Google e clama por ele. É possível coletar frases como “o Google não erra”, “o Google não falha (ou não ‘faia’)”, “o Google não esquece”.

Há pedidos engraçados como: “Google, me diga a rua em que a tia Nair morava” ou certidão de nascimento: “Quanto anos Rodriguinho faria hoje? ou, ainda, prova de vida: “Google, o Carneirinho tá vivo?”.

O Google não pode atrasar. “Cadê o Google?”, no dia em que ele não nos avisou, antes das 6h, como de costume, sobre o aniversário da nossa saudosa matriarca, dona Rita, choveram reclamações, mas foi por uma boa causa. O Google havia saído cedo para tomar a primeira dose da vacina contra a covid-19. Teve um dia que o Google demorou a aparecer e a galera não perdoou: “Domingo Google começa mais tarde”. Em outra oportunidade, Google tava sumido, mas avisei para a família que você havia perdido o celular... Ufa!

Todavia, acontece também do horário causar estranheza. Certo dia, você mandou mensagem no grupo 4h35 parabenizando meu irmão pelo aniversário. Alguém postou às 8h17: “O Google não dorme?”. No dia em que os parabéns saíram às 0h32, o comentário foi: “O Google agora tá trabalhando com despertador!”

“Hoje, especialmente, sou grata por ser nosso Google. Ontem, conversando com você, consegui entender boa parte da minha história graças à sua memória. Por isso e muito mais te desejo vida longa, com saúde e muita energia”, escreveu uma prima, felicitando você pelo aniversário. Aliás, ela se sente privilegiada por comemorar aniversário no mesmo dia que você, pai.

Enfim, o Google está entre nós. Você está entre nós. O Google é mito, magia, registro, informação, memória, documento, conhecimento, referência, inspiração, biblioteca, cartório, fotografia, verdade.

O Google é pai.

*Um beijo carinhoso,
Luana.*

4 A PLATAFORMA DE BUSCA *GOOGLE*

Começamos com duas ressalvas. O título deste capítulo pressupõe uma argumentação que parece óbvia por tudo que desenvolvemos até este ponto da pesquisa: defendemos que *Google* é uma plataforma de busca e dotada de caráter infraestrutural. É importante demarcar essa obviedade, porque lidamos com muitos termos adjacentes, como buscador, ferramenta de busca, motor de busca e mecanismo de pesquisa, que não foram e não serão abandonados, mas estarão a serviço da causa maior, que é caracterizar a plataforma. Durante a tese, nos referimos a três “personas” entrelaçadas, que, no decorrer do texto, muitas vezes se misturam: “*Google Search*” (ferramenta de busca); “*Google* Produtos” (conjunto de produtos e serviços, além do buscador, que robustecem a plataforma) e, ainda “*Google/Alphabet Inc.*” (no sentido de empresa). Diferenciar essas “personas” não parece ser uma grande questão, pelo contrário, na literatura em português e inglês não percebemos esse melindre, porque o contexto vai dar ao leitor condição de distinguir, mas achamos que era válido o aviso, porque neste capítulo continuaremos a questionar a função de mero buscador, para apresentar uma função mais infraestrutural.

A segunda ressalva é que este capítulo não pretende ser o abrangente mergulho que Siva Vaidhyanathan (2011) fez para escrever *A Googelização de tudo*; afinal de contas, ele já o fez. Em complementaridade às pesquisas do autor, pretendemos dar luz à “especificidade do meio” (ROGERS, 2016, p. 10) que deixe clara a ação da plataforma infraestrutural na reconfiguração das práticas de escrita e edição, já que o *Google* se sustenta na programabilidade de seus algoritmos de rastreamento, indexação e classificação; reforça-se por meio do critério da popularidade, que é o gás do motor de busca; equilibra-se economicamente na conectividade de leitores que buscam, clicam/engajam, deixam seus dados e rastros de uso, instituindo um modelo de trocas e fazendo girar a lógica da propaganda direcionada; e se retroalimenta (usando aprendizagem de máquina) da coleta, mensuração e predição de dados expressos nos tipos e volumes de busca. Olhar para a especificidade do *Google* não se refere apenas a ver a subdivisão das disciplinas nos estudos de mídia, de acordo com os objetos de estudo: vídeo, rádio, televisão, *Web* etc., mas se refere ainda a ver as distinções ontológicas das mídias, investigando a natureza da existência. A especificidade do meio, proposta por Rogers (2016), considera explorar em profundidade os métodos do meio, compreendendo, por exemplo, como *hiperlinks* e palavras-chaves são manipulados, especificamente, para a medição do *Google* e aprender com as particularidades dessa manipulação.

Antes de seguir, vamos compreender o método de pesquisa que deu origem às análises apresentadas neste capítulo.

4.1 ANÁLISE DOCUMENTAL

A análise documental foi a primeira etapa da nossa pesquisa, em que olhamos para os modos como a plataforma de busca *Google* estabelece regras e negocia condutas com os usuários, ou seja, a forma como estabelece governança. Essa mirada foi importante para caracterização do *Google* como plataforma infraestrutural e, ainda, para dar visibilidade ao modelo de seleção e ordenação da informação que influencia diretamente a plataformização do texto. Para d’Andréa (2020), a governança se dá pelo conjunto heterogêneo de mecanismos e práticas de ordem técnica, política, jurídica e comercial que regulam o funcionamento de plataformas *online*. A ironia está no fato de que muitos documentos constroem a noção política das plataformas como lugares a partir dos quais se pode ler, falar, postar, clicar e ser ouvido e nem parecem ser *gatekeepers* elitistas, dotados de restrições técnicas e normativas (HELMOND, 2015).

4.1.1 Seleção do *corpus*

A autorregulação das plataformas revela o controle da circulação de conteúdos que, segundo d’Andréa (2020), é exercido de modo “[...] distribuído e não coercitivo, o que não significa que não seja, de modo crescente, controverso e conflituoso” (p. 42). Portanto, investigamos e analisamos os seguintes documentos do *Google*, em um trabalho de leitura e anotação, à luz dos fundamentos da pesquisa:

Quadro 8 – *Corpus etapa 1*

Documento	Endereço
Como resultados são gerados automaticamente	https://www.google.com/intl
Search quality evaluator guidelines	https://guidelines.raterhub.com/searchqualityevaluatorguidelines.pdf
Guia do Google de introdução à otimização para motores de busca	https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=pt-br
Como a busca do Google funciona	https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/

Relatório de transparência	https://transparencyreport.google.com/about?hl=pt-br
Guia detalhado sobre como a Pesquisa Google funciona	https://developers.google.com/search/docs/fundamentals/how-search-works?hl=pt-br
Melhorando a Busca com testes rigorosos	https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/rigorous-testing/
Como resultados são gerados automaticamente	https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/ranking-results/#meaning
Avaliação da experiência na página para melhorar o conteúdo da Web	https://developers.google.com/search/blog/2020/05/evaluating-page-experience?hl=pt-br#page-experience-ranking
Termo de serviços e política de privacidade	https://policies.google.com/terms?hl=pt-BR
Patentes	https://patents.google.com/patent/US6285999B1/en

Fonte: Elaborado pela autora, 2023

Esse *corpus* foi selecionado porque alguns documentos têm ênfase jurídica que regulam propriedade intelectual ou tentam proteger o *Google* de atitudes danosas praticadas por seus usuários. Nesse sentido, contribuíram para a coleta de diretrizes e regramentos fundamentais para compreender políticas e governança. Outros documentos dizem respeito às expectativas da plataforma sobre o que é apropriado ou não durante o uso do motor de busca e, ainda, sobre o que é ideal para alcançar melhores performances de circulação de conteúdo. Desse modo, ajudaram-nos a compreender o funcionamento e conectar esse modo de operação às reconfigurações das práticas de escrita e edição.

4.2 *DON'T BE EVIL*: A ONIPOTÊNCIA DO *GOOGLE*

O Google foi criado em 1998 por dois estudantes de ciências da computação da *Stanford University*, nos Estados Unidos, como uma empresa de prestação de serviço de busca na internet. Desde então, cresceu incrivelmente rápido. Das pesquisas em um único idioma, agora oferece *e-mail*, sistema operacional, planilhas, mapas, vídeos, armazenamento em nuvem, aplicativo de navegação, entre outros serviços. Incluem-se formas de publicidade e aplicativos para celular em vários idiomas. O que começou com dois graduandos em um quarto de universidade se transformou em milhares de funcionários e escritórios espalhados pelo mundo. Desse crescimento acelerado, nasceu o *slogan* informal *Don't be evil*, sugerido

por um funcionário. O *Google* começou a ser apontado por concorrentes como a empresa que dominaria o mundo e adotou esse *slogan* para dizer que é possível ganhar muito dinheiro sem fazer mal às pessoas.

Afinal, qual é o papel da plataforma de busca *Google* na internet? No momento da escrita deste texto, a *Google* era responsável por 92% das pesquisas na *Web* em todo o mundo, seguido por *Bing* (3,34%), *Yahoo* (1,34%) e *Baidu* (0,83%), conforme dados do *Statcounter*, site especializado em análise de tráfego da *Web*. Se recortarmos apenas pelo Brasil, as estatísticas reforçam o caráter monopolizante: *Google* tinha 96,43%, seguido por *Bing* (2,3%), *Yahoo* (1,02%) e *DuckDuckGo* (0,12%) do volume de busca. O reinado precisa ser contestado porque, conforme argumenta Rogers (2016), os mecanismos de busca são um ponto de entrada crucial na *Web*, são máquinas epistemológicas, na medida em que rastreiam, indexam, armazenam em *cache* e, em última instância, ordenam o conteúdo.

Concordamos com Rieder (2020, p. 37) que enxerga o *Google* como uma “[...] agência de ordem no sentido de que apoia e organiza sistemas sociais, políticos e econômicos de maneiras específicas”. Ao consolidar como inevitáveis seu modelo de indexação e hierarquização, a plataforma regula práticas diversas, em especial, a cultura da escrita em ambiente digital, por tensionar produtores de texto a redigir sob as amarras da relevância algorítmica. Para Rogers (2016), os efeitos do *Google* são efeitos das mídias, de uma forma tão infraestrutural para a produção de texto, que é possível pensar na “[...] morte do editor-perito e ascensão do algoritmo de *back-end*” (ROGERS, 2016, p. 4).

A googlização atinge três grandes áreas de interesse e conduta humanas: “nós” (através dos efeitos do *Google* sobre nossas informações pessoais, nossos hábitos, opiniões e juízo de valor); “o mundo” (através da globalização de um estranho tipo de vigilância daquilo que chamo de imperialismo infraestrutural); e “o conhecimento” (através de seus efeitos sobre o uso de um vastíssimo agregado de conhecimentos acumulados em livros, base de dado online e na Internet). (VAIDHYANNATHAN, 2011, p. 16).

Os serviços do *Google* atraem leitores para uma dependência cada vez mais profunda, porque parecem amplamente acessíveis numa lógica de se colocar no lugar da *Open Web*. Olhar para o *Google*, com as lentes dos estudos que detalhamos no Capítulo 3, traz em destaque a programabilidade como a base do sucesso, afinal, os sistemas do *Google* parecem ser aplicativos nativos da *Web*, nunca vendidos ou empacotados, mas entregues como um serviço. “Por projetar e controlar a pesquisa na Internet, a empresa detém enorme poder sobre

a visibilidade dos recursos baseados na Web” (PLANTIN et al., 2016, p. 18, tradução nossa)⁶⁹.

A lógica da plataforma do *Google* cria benefícios reais tanto para usuários, quanto para desenvolvedores terceirizados e empresas de *marketing*, por exemplo, que nascem e se sustentam com serviços que consistem, basicamente, em melhorar posições de conteúdos e marcas nos resultados de pesquisa do *Google*. O caráter infraestrutural ajuda na construção de um *ethos* de empresa parceira que resolve muitos problemas da humanidade, afinal, ironicamente, nem os governos têm conseguido fazer tanto. Dessa forma, o *Google* se coloca em uma posição de domínio sem paralelo, aumentando seu poder de moldar vidas e tratar o “[...] conhecimento como objeto programável” (PLANTIN et al., 2016, p. 17, tradução nossa)⁷⁰.

Vaidhyanathan (2001) propõe que os empreendimentos da *Google/Alphabet Inc.* são divididos em três grandes categorias: a) escanear e fazer *links*, a exemplo do *Google Search*, responsável por rastrear, indexar e classificar conteúdos, considerando que o buscador não hospeda essas informações, mas as ordena; b) hospedar e servir, a exemplo do *YouTube*, em que usuários são convidados a criar e carregar conteúdos nos servidores próprios da empresa; c) escanear e servir, área em que o *Google* “esquadrinha” o mundo real, digitaliza coisas reais e oferece como parte da experiência de uso de seus serviços, a exemplo do *Google Street View*. Percebe-se, portanto, que a empresa fornece aplicativos baseados na internet, incluindo os mais vitais para a maioria dos usuários.

A companhia comercializa esses serviços para empresas, escolas e universidades, em muitos casos substituindo *softwares* licenciados herdados de vários fornecedores pelo pacote do *Google Apps*. A educação se rendeu ao *Google for Education*, que oferece ferramentas educacionais *online* para professores e estudantes, como *Google Classroom* (modelos de aulas e tarefas) e *Google Meet* (aulas ao vivo). A “roda” do jornalismo digital atualmente só gira com ferramentas gratuitas, como *Google Analytics*, de gestão de tráfego para *site*; e *Google Trends*, que é um painel de termos mais populares no buscador; além de *softwares* pagos, como o *Google News Consumer Insights*, que fornece *insights* e recomendações para melhoria de SEO e estratégia de direcionamento de conteúdo (e propaganda) para os leitores. As pesquisas acadêmicas estão sob o filtro da popularidade de artigos científicos e autores mais citados, tudo isso por causa da mediação e da metrificação do *Google Scholar*, ferramenta de busca setORIZADA em documentos acadêmicos.

⁶⁹ Tradução nossa para: “Because it designs and controls Internet search, the company holds enormous power over the visibility of web-based resources.”

⁷⁰ Tradução nossa para: “Knowledge as programmable object”.

Já o setor dos livros enfrentou a mais controversa das apropriações e talvez uma das poucas em que o *Google* tenha sido parcialmente “derrotado”, conforme analisaram Darton (2009), Schiffrin (2011), Thompson (2013), Vaidhyanathan (2011) e Plantin et al. (2016). A partir de 2004, por meio do projeto *Library Google*, o *Google* fez parceria com as principais universidades – *Harvard, Stanford, Oxford* e outras – para digitalizar livros, apresentando a ideia como um ambicioso projeto de atualização de bibliotecas públicas e nacionais para o século XXI. Plantin et al. (2016) descreveu esse movimento como uma tentativa do *Google* em se tornar uma infraestrutura de conhecimento, a despeito de já ser uma plataforma infraestrutural em outras atuações. Porém, as intenções foram rapidamente vinculadas à lógica da plataforma de busca, que precisa ter ampla base de dados de documentos para aprimorar as pesquisas. Para editores, autores e escritores, o *Google* estava infringindo princípios básicos de direitos autorais, disponibilizando trechos de obras gratuitamente. Por isso, esses grupos entraram com ações na justiça, movimento que Rogers (2016) considera um pioneiro das críticas à “googlização de tudo” (VAIDHYANATHAN, 2011). Depois de um longo desenrolar, o *Google* foi derrotado na Suprema Corte dos Estados Unidos, sendo obrigado a fornecer compensações financeiras de direitos autorais para livros já digitalizados e se submeter a regras de *copyright* para qualquer plano futuro. “Os tropeços do projeto empanaram sua aura de invencibilidade” (VAIDHYANATHAN, 2011, p. 227).

[...] a herança cultural representada pelos inúmeros livros publicados anteriormente – tanto de domínio público quanto de obras órfãs – é simplesmente importante demais para ser deixada nas mãos de uma empresa privada cuja direção futura e prioridades serão decididas por acionistas, e não pelo interesse público. (THOMPSON, 2013, p. 398).

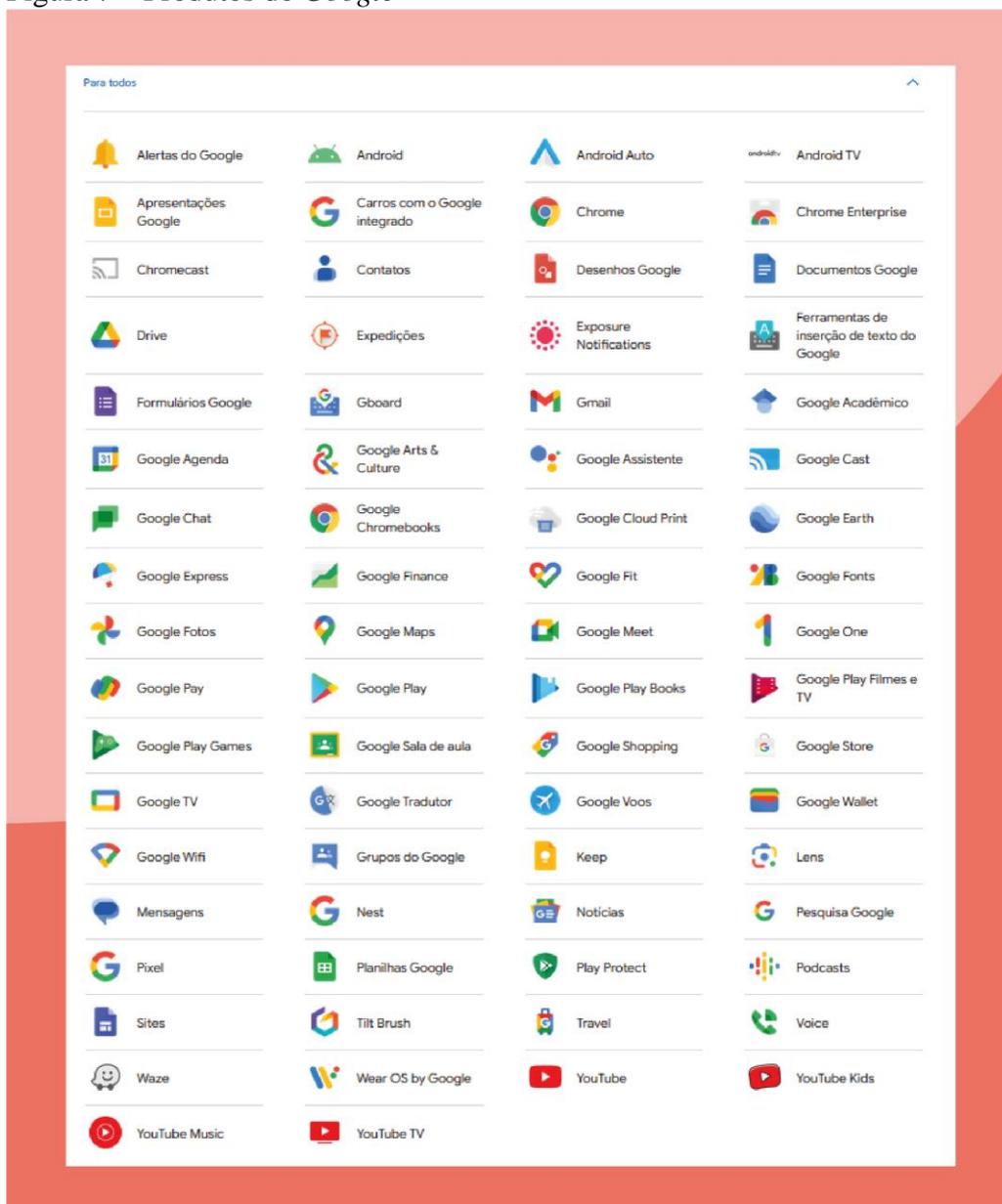
Depois do imbróglio, o que existe hoje é o aplicativo *Google Book Search* que, em acordo com editoras, trechos de livros – previstos em contratos – são disponibilizados e, ao leitor, é oferecido o *link* para comprar da obra, uma forma de *marketing digital* gratuita que serve bem às editoras. O *Book Search* continua sendo muito bom para a plataforma de busca, porque contribui para a riqueza do seu índice ou banco de dados. Entendemos que o *Google* ainda não conseguiu ser uma biblioteca digital universal – a nova Biblioteca de Alexandria –, apesar de a empresa ter certeza que alcançou esse objetivo: “[...] nosso índice de pesquisa é como uma biblioteca que contém mais informações que todas as bibliotecas do mundo somadas” (GOOGLE, 2022d).

Todavia, a plataforma concretiza parcialmente o objetivo de aumentar a base de dados com textos robustos dos livros. E, claro, o *Google* continuou as investidas. Em 2009,

por exemplo, comprou o *reCAPTCHA* (tecnologia CAPTCHA, baseada no *Completely Automated*, que significa teste de Turing público para diferenciar computadores e humanos), que aos olhos de muitos parece ser apenas uma ferramenta de segurança para combater a presença de robôs (*crawlers*) criadores de *spams*. No entanto, é um movimento do *Google* para usar a colaboração de centenas de usuários para decifrar caracteres que não poderiam ser facilmente definidos por meio de reconhecimento óptico de caracteres automatizado (OCR). Assim, o ReCAPTCHA pode ajudar a mostrar palavras escaneadas de jornais ou livros antigos e, que talvez, não estivessem muito visíveis. Toda vez que muitos usuários identificam determinada letra ou palavra corretamente, eles “ensinam” aos algoritmos do *Google* a realizar aquela leitura e tornam possível a digitalização de documentos importantes, acervos centenários e obras literárias fruto de processos de impressão mais imprecisos e precários.

Todos esses relatos só exemplificam alguns setores. Um levantamento mais sistemático de serviços, que mostra outros “braços” do conglomerado *Google*, está na página de produtos, conforme a Figura 7, que contempla produtos destinados a “todos” usuários sem recortes, por exemplo, de serviço empresarial.

Figura 7 – Produtos do Google



Fonte: Google, 2023.

Também há a atuação como sistema operacional em *smartphones*, afinal, o *Android* é do *Google*, e centraliza por *default* uma série de acessos aos aplicativos da empresa, o que gera uma vantagem concorrencial desequilibrada em relação a outros aplicativos (apps). O buscador do *Google*, por exemplo, é padrão até mesmo em *smartphones* com sistema operacional *IOS* (sistema operacional móvel da *Apple Inc.*), inclusive, a *Apple* recebe, anualmente, bilhões de dólares para manter essa prática de favoritismo. O mais óbvio seria a *Apple* ter como padrão um buscador próprio, mas isso colocaria em risco a soberania do *Google Search*. “O uso hegemônico dos mecanismos de busca da *Google* e o modo como

essa plataforma infraestrutural se consolidou como uma nova camada ‘inevitável’, ou um ponto obrigatório de passagem.” (D’ANDRÉA, 2020, p. 35).

Aliás, o domínio sem precedentes da empresa a fez acumular, desde 2010, pelo menos quatro investigações antitruste na União Europeia (os alvos são *Google Adsense*, *Android*, *Google Shopping* e *Google Search*)⁷¹ e três nos EUA (o alvo é *Google Search*)⁷². Em setembro de 2022, o Departamento de Justiça dos Estados Unidos, por meio da Divisão Antitruste, divulgou um relatório consolidado de uma ação contra o *Google*⁷³, que reúne um histórico de reclamações em desfavor da empresa. Sobre a natureza dessa ação, o Departamento afirma:

Duas décadas atrás, o Google se tornou o queridinho do Vale do Silício como uma startup determinada como um jeito inovador de pesquisar na internet emergente. Esse Google já se foi. O Google de hoje é um *gatekeeper* de monopólio da internet e uma das empresas mais ricas do planeta, com valor de mercado de US\$ 1 trilhão e receita anual superior a US\$ 160 bilhões. Por muitos anos, o Google usou táticas anticompetitivas para manter e ampliar seus monopólios no mercado de serviços de busca geral, publicidade direcionada na busca e anúncios baseados em textos - os pilares de seu império. (THE UNITED STATES DEPARTMENT OF JUSTICE, 2022, tradução nossa)⁷⁴.

No Brasil, há pelo menos três ações no Conselho Administrativo de Defesa Econômica (Cade), em que os alvos são *Google Search* e *Google Shopping*⁷⁵. Por aqui, o *Google* venceu todos os processos, que, aliás, foram arquivados pelo Cade sob alegação de “[...] indícios insuficientes das infrações” e falta de “[...] remédios antitruste disponíveis”⁷⁶ para resolução dos casos. Em todos os processos, o foco central é apurar práticas monopolistas e anticompetitivas de publicidade direcionada e ações que impedem empresas rivais se estabelecerem no mercado, como no caso dos *apps* instalados em *default* no sistema *Android*.

⁷¹ Disponível em: https://ec.europa.eu/commission/presscorner/detail/es/MEMO_17_1785. Acesso em: 30 set. 2022.

⁷² Disponível em: <https://www.wired.com/story/google-antitrust-lawsuits-explainer/>. Acesso em: 30 set. 2022.

⁷³ Disponível em: <https://www.justice.gov/atr/case-document/file/1536456/download>. Acesso em: 04 out. 2022.

⁷⁴ Tradução nossa para: “Two decades ago, Google became the darling of silicon valley as a scrappy startup an innovative way to search the emerge internet. That Google is long gone. The Google of today is a monopoly gatekeeper for the internet and one of the wealthiest companies on the planet with a market value of \$ 1 trillion and annual revenue exceeding \$ 160 billion. For many years, Google has used anticompetitive tactics to maintain and extend its monopolies in the market for general search services, search advertising and a general search text advertising – the cornerstones of its empire.”

⁷⁵ Disponível em: <https://theintercept.com/2020/11/23/cade-google-viagens-pagas-ong/>. Acesso em: 30 set. 2022.

⁷⁶ Trechos da última decisão do CADE foi em 2019, assinada pelo presidente Alexandre Barreto de Souza. Disponível em: https://sei.cade.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?DZ2uWeaYicbuRZEFhBt-n3BfPLlu9u7akQAh8mpB9yObOrdAAnkZ36Rru6H33qbFO51_fjuVWb1uid6m5S5BxJ8gFyW8xprjnuylPdYbaX3VDhhG3SAtGWLJPIqjsEDX. Acesso em: 30 set. 2022.

Além da hegemonia já descrita, há uma perspectiva tecno-material que demarca perfeitamente o funcionamento do *Google* como plataforma, que é o fato de disponibilizar uma vasta infraestrutura para a construção de aplicativos, configurando-se como um conjunto de sistemas que pode ser customizado por outros desenvolvedores externos e usuários, adaptado a incontáveis necessidades e nichos. Essa lógica, chamada por Anne Helmond (2015) de “programabilidade” – como já citamos anteriormente –, acontece por meio de interfaces que possibilitam reprogramação, as APIs⁷⁷, e ajuda a ratificar de determinados *sites* ou mecanismos como plataformas. O *Google* oferece APIs para análise de dados, ferramentas de gestão, sistemas de segurança e identidade, aprendizado de máquina, armazenamento e banco de dados, entre outras. Assim, reafirmamos que o *Google* é uma plataforma porque ele disponibiliza APIs que podem ser usadas para construir novos serviços e integrar novos *websites* e aplicativos aos dados e às funcionalidades do *Google*. Por exemplo, APIs de pesquisa do *Google* são interfaces que permitem extrair resultados de pesquisa, deixando que desenvolvedores criem *sites* e programas para acessar páginas de resultados (*Search Engine Results Page* – SERP). Um possível uso seria a construção de um sistema que permita a clientes de uma agência de *marketing* obter dados sobre classificação e posições para qualquer número de termos de pesquisa, em tempo real. É possível também que uma universidade crie seu próprio mecanismo de pesquisa programável, podendo optar por fazer com que seu mecanismo de pesquisa retorne resultados da *Web* ou de um *site* específico.

Outro aspecto tecno-material é a existência de *data centers* gigantescos que sustentam toda estrutura de *bits*. No caso do *Google*⁷⁸, são 14 unidades na América do Norte, seis na Europa, três na Ásia e uma na América do Sul. Ochigame (2021a) alerta que computadores são feitos também de átomos de ouro, prata, silício, cobre, estanho, tungstênio e muito mais. Esses centros consomem grandes quantidades de energia elétrica, geralmente provenientes de carvão, bem como milhões de litros de água para resfriamento de servidores. O *Google* vem tentando responder às críticas de ambientalistas em relação a essa infraestrutura, em especial, com uma meta de trabalhar com energia livre de carbono 24 horas por dia, em qualquer lugar e momento.

Os data centers estão transformando paisagens de Utah à Islândia. Apesar de ser imaginada como uma “nuvem” dispersa, etérea e sem lugar, a internet global requer

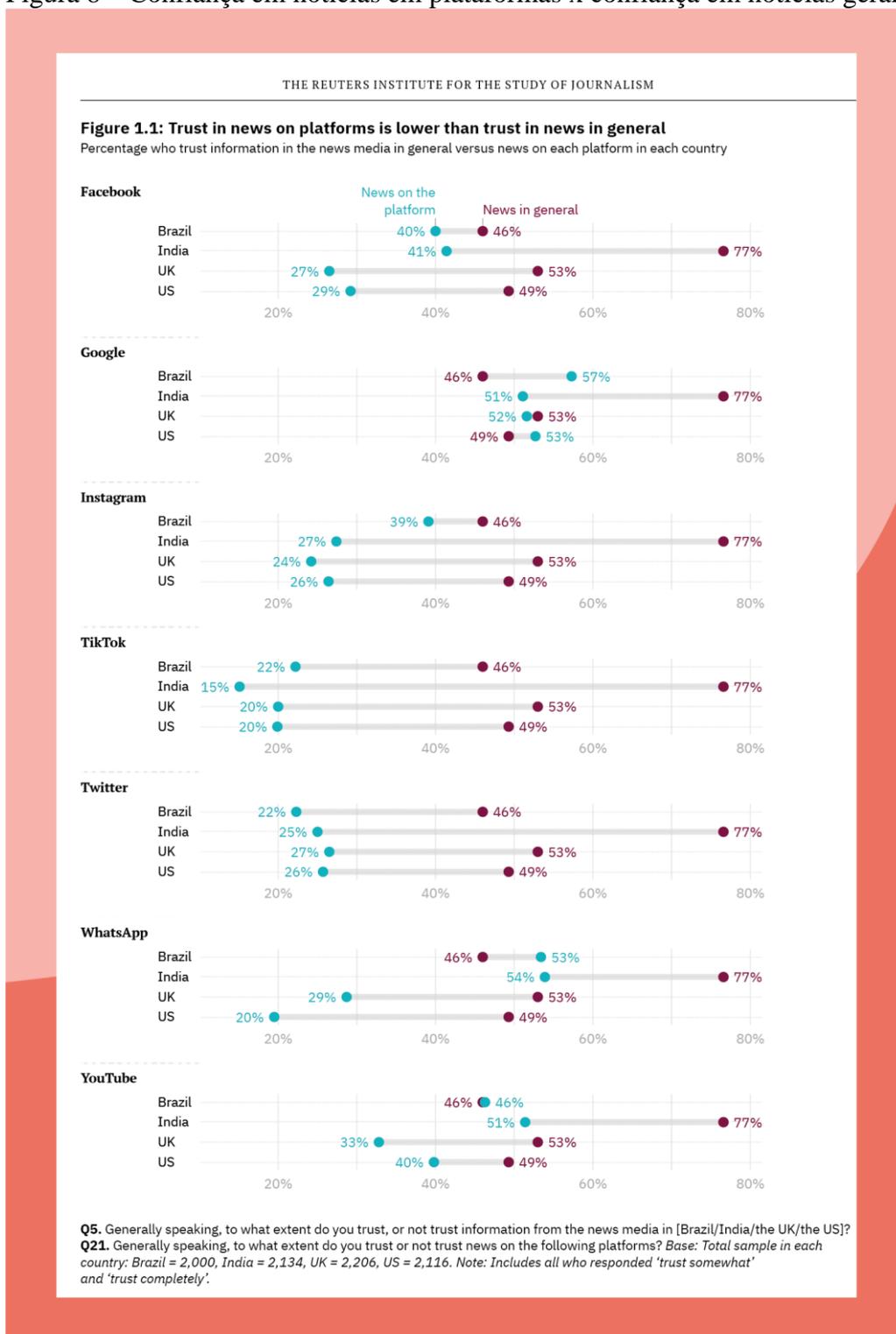
⁷⁷ “Uma API é uma interface proporcionada por um aplicativo, que possibilita aos usuários interagirem com – ou em resposta a – dados ou solicitações de outros programas, outros aplicativos ou websites. As APIs facilitam a troca de dados entre os aplicativos, possibilitam a criação de novos aplicativos, e constroem a base para o conceito de ‘web como plataforma’”. (MURUGESAN, 2007, p. 36 *apud* HELMOND, 2015, p. 56).

⁷⁸ Disponível em: <https://www.google.com/intl/pt-BR/about/datacenters/>. Acesso em: 07 jun. 2023.

infraestruturas materiais não só em terra, mas também no mar e no exterior. espaço, incluindo satélites e cabos de fibra óptica transoceânicos. (OCHIGAME, 2021a, p. 29).

Apesar das evidências, é sintomático ver a percepção das pessoas em relação à onipotência, infraestrutura, ordenação de informação e reformatação da *Web* promovidas pelo *Google*. Os dados da Figura 8 mostram a confiança das pessoas em notícias gerais *versus* notícias que circulam pelas plataformas *online*, com uma análise individualizada de cada plataforma. É um recorte sobre notícias; portanto, reconhecemos que não reflete a percepção sobre conteúdos no geral. Mesmo assim, vale observar que o *Google* é a única plataforma com maior percentual de confiança (57%) do que as notícias gerais no Brasil. Isso significa que *Google* é a única plataforma *online* em que as pessoas acreditam mais, se compararmos com fontes de notícias tradicionais. Para Metaxa e Torres-Echeverry (2017), o combustível tácito por trás do fenômeno *Google* é a confiança implícita depositada em mecanismos de busca e o uso proposital que estrategistas de texto fazem para explorar essa confiança.

Figura 8 – Confiança em notícias em plataformas x confiança em notícias gerais



Fonte: Reuter Institute/Universidade of Oxford (2022).

Notável também é o desempenho do *YouTube*, uma plataforma do grupo *Google/Alphabet Inc.* que apresentou percentual empatado (46%) na percepção de confiança sobre notícias no Brasil. O paradoxo é que os pesquisadores Fonseca e d'Andrea (2020) demonstraram que a governança e as mediações algorítmicas do *YouTube* na moderação e na

circulação de conteúdos sobre a pandemia de covid-19 potencializaram a circulação de desinformação. Para os pesquisadores, apesar de notar-se aspecto potencialmente positivo no enfrentamento da pandemia, há outras dinâmicas que privilegiam conteúdos desinformativos.

Siva Vaidhyathan (2011, p. 17) constata: “[...] devido a nossa fé no Google e em suas alegações de onisciência, onipotência e benevolência, tendemos a atribuir aos resultados de pesquisa um poder desmedido e imerecido”. A pesquisa de confiança em notícias da *Reuters* renova essa ideia de “fé cega” e traz mais perguntas do que respostas: os brasileiros percebem o *Google* como uma plataforma de ambiente e função mediadora, assim como *Facebook*, *Instagram* e *TikTok*? Existe um imaginário positivo sobre a empresa *Google* que a descola de todas “negatividades” expostas, por exemplo, sobre o *Facebook* no escândalo da *Cambridge Analytica*⁷⁹ ou pelo documentário *Dilema das redes*? Está claro como funciona o *Google*, considerando as principais lógicas algorítmicas e como elas servem à publicidade direcionada? Não fomos perguntar isso às pessoas, pois não é este o objetivo da pesquisa, mas a tese que defendemos aqui demonstra uma vitória, mesmo que temporária, da plataforma de busca em manter obscuridades sobre seu funcionamento e da *Google/Alphabet Inc.* em construir um *ethos* de empresa parceira.

4.3 LÓGICAS DOS MOTORES DE BUSCA E GOVERNANÇA DO *GOOGLE*

Os motores de busca são mecanismos que medeiam as experiências de pesquisa de informações na *Web*, portanto, concretizam as essências política e técnica para o funcionamento das plataformas de busca. Eles constituem uma poderosa fonte de acesso na *Web*, no sentido que já argumentamos no Capítulo 2, ao costurar as noções de cadeia de valor da informação de Thompson (2013) e Furtado (2003). Aliás, Introna e Nissebaum (2006), corroboram dizendo que o acesso para a *Web* é pré-configurado de maneiras sutis, mas politicamente importantes, resultando na exclusão de vozes significativas. Para os autores, o acesso envolve não apenas um computador e uma conexão de rede, mas as habilidades e *know-how* que permitem o uso eficaz: “[...] acesso implica um mecanismo abrangente para encontrar e ser encontrado” (INTRONA; NISSEBAUM, 2006, p. 181, tradução nossa)⁸⁰.

A plataforma *Google* tem uma robusta governança no entorno das essências política e técnica de seu motor de busca, em que parece negociar com profissionais de

⁷⁹ O grande escândalo político que revelou que dados de 87 milhões de usuários do Facebook foram coletados pela empresa de consultoria Cambridge Analytica, sem consentimento deles, para serem usados em publicidade política (JURNO, 2020, p. 16).

⁸⁰ Tradução nossa para: “Access implies a comprehensive mechanism for finding and being found.”

criadores de texto e desenvolvedores *Web* a gestão e regulação do sistema “produção, circulação e leitura” em ambiente digital. Bem, na verdade, a plataforma cria uma autorregulação, publiciza isso de forma extremamente detalhada – em centenas de documentos acessíveis (os quais exploramos parcialmente nesta pesquisa) – para dar a entender que é realmente aberta e democrática, ensina e qualifica o mundo para operar suas interfaces, permite o uso igualitário de instrumentos para obtenção de visibilidade das produções textuais. O modo como o *Google* faz e comunica a governança parece fluido e não coercitivo, dando a entender que facilita a vida de todos os profissionais com trabalhos atrelados à engrenagem da plataforma. Mas o regramento é parte do processo decisório do *Google* sobre o que circula ou não, por isso é controverso e conflituoso, mesmo que pareça um ingênuo conjunto de aconselhamentos. Portanto, ao compreender as lógicas dos motores de busca e a governança de *Google*, nosso foco é na redução da opacidade sobre o funcionamento das plataformas de busca, descrevendo funções intencionais de *gatekeeping*.

4.3.1 Políticas

Os motores de busca são vistos socialmente como a solução técnica perfeita para o “problema” alimentado pelos discursos neoliberais de que usuários de internet estão constantemente confrontados com muitos documentos, conteúdos, produtos, enfim, muitas “coisas” que não conseguem manusear. Esses mecanismos reduzem custos mentais e tempo, respondem a partir de uma ampla variedade de fontes. No senso comum, “[...] quanto mais adequado for o resultado apresentado pelo buscador e quanto mais rapidamente ele trouxer esse resultado, mais útil se tornará para quem faz a busca”. (GABRIEL, 2010, p. 210).

Claro que olhamos criticamente para essa percepção geral dos motores busca, apesar de reconhecermos o potencial do serviço prestado. Nunca se trata de desmerecer a utilidade e a importância dessas ferramentas para os processos de circulação e leitura em ambiente digital, criando uma aura maniqueísta, mas é o caso de questionar as intencionalidades da propaganda direcionada inseridas na fantasia de utilidade pública, pois “[...] embora a Internet e a Web ofereçam perspectivas interessantes para promover o bem público, os benefícios são condicionais, baseados precariamente em uma série de fatores políticos, econômicos e técnicos” (INTRONA; NISSENBAUM, 2006, p. 170, tradução

nossa)⁸¹. Ou seja, o *Google* e outras plataformas de busca comerciais são ambivalentes, assim como toda tecnologia. “Seu serviço de busca é o motivo pelo qual o visitamos. A publicidade é o que o faz seguir em frente.” (VAIDHYANATHAN, 2011, p. 30).

E como tudo isso afeta as experiências de busca dos leitores e a natureza da *Web*? Ao destrinchar politicamente e tecnicamente os motores de busca, vamos perceber que eles são mais propensos a encontrar *sites* populares e grandes cujos *designers* têm conhecimentos técnicos suficientes para ter sucesso no jogo de classificação, especialmente aqueles *sites* cujos proprietários podem investir para melhorar o posicionamento do seu *site*. A definição do *Google* para “experiência de busca” ilustra o modo como a empresa posiciona seus serviços socialmente e que em nada tem a ver com o entendimento crítico sobre excluir e visibilizar determinados conteúdos.

A World Wide Web é uma vasta coleção de informações e conteúdos online. Os mecanismos de pesquisa na Internet fornecem uma poderosa maneira de explorar esse universo online. [...] Diferentes tipos de pesquisas precisam de tipos muito diferentes de resultados de pesquisa. Os resultados da pesquisa médica devem ser de alta qualidade, confiáveis e com autoridade no assunto. Os resultados da pesquisa para “imagens de animais fofos” devem ser adoráveis. Já os resultados da pesquisa para um site ou página da web específica devem ter o resultado desejado no topo. (GOOGLE, 2022b).

Mesmo que as empresas responsáveis nunca assumam claramente o papel mediador, os motores de busca acabam sendo espaços para colisão entre “relatos alternativos da realidade” (ROGERS, 2016, p. 21), pois classificam e apresentam, lado a lado, verdades subjetivamente construídas. Essa influência no acesso e na seleção das informações é um dos motivos pelos quais insistimos em caracterizar o *Google* como plataforma infraestrutural. Vaidhyathan (2011) entende que o *Google*, mesmo fazendo entregas individualizadas nas pesquisas, trabalha em uma lógica universalizante, tratando a circulação de textos no mundo todo sem considerar diferenças ou preocupações locais. É como se a plataforma *online* considerasse que as pessoas são iguais em toda parte, têm necessidades de informação idênticas, os mesmos valores e desejos, ou seja, o *Google* preconiza um funcionamento casado ao fundamentalismo de mercado e ao liberalismo. Esses são indícios do que Vaidhyathan (2011) denomina “imperialismo infraestrutural”, que se dá justamente no desequilíbrio dos fluxos globais de informações.

⁸¹ Tradução nossa para: “[...] is that although the Internet and the Web offer exciting prospects for furthering the public good, the benefits are conditional, resting precariously on a number of political, economic, and technical factors”.

O que é perturbador, perigoso ou mesmo interessante não é tanto a marca Google: é que, ao difundir e estruturar diferentes maneiras de procurar, encontrar, explorar, comprar, apresentar tudo isso, o Google influencia hábitos de pensamento e ação. Essas configurações predefinidas e esses empurrões são expressos de uma ideologia. (VAIDHYANATHAN, 2011, p. 124).

Rieder (2020, p. 43) corrobora a ideia de universalização quando afirma que motores de busca organizam as informações sociais de maneiras cada vez mais complicadas, sem evoluir para aleatoriedade. O autor, citando Couldry e Hepp (2016, p. 175), refere-se à noção de “coletividades midiáticas”, que podem muito bem ser montadas e ordenadas algorítmicamente, representando vetores de coagulação, mesmo que o espectro de coletividades tenha aumentado exponencialmente.

As técnicas de ordenamento da informação, que atuam como motores de ordem que modulam ativamente as relações entre usuários e unidades circulantes de vários tipos, operam em padrões e falhas existentes em sociedades diversificadas, mas desiguais. [...] Grandes e pequenas variações entre indivíduos (com dados) podem ser lidas de pontos de vista vinculados a metas de desempenho específicas, como maior tempo no local, taxas de cliques mais altas, taxas de inadimplência mais baixas, funcionários mais produtivos e assim por diante, mas também integrados em atividades mais amplas, como pesquisa de mercado, desenvolvimento de produtos ou planejamento estratégico. (RIEDER, 2020, p. 46).

Na lógica do “imperialismo infraestrutural”, os textos, signos e mensagens que fluem pelas redes globais, especialmente na direção Norte-Sul, importam menos do que o modo como se dão esses fluxos, quanta receita geram e quais empresas vão utilizá-los e reutilizá-los. Até 2010, o *Google* sustentou a posição de “mero buscador” de forma veemente, mas, a partir daquele ano, a pressão por transparência, em especial na seleção de informações, obrigou a empresa a divulgar relatórios de atividades. A partir de 2016, quando o *Google* começou a ser acusado pela sociedade civil de permitir a circulação de *fake news* e alimentar a desinformação em contextos eleitorais, ele anunciou mudanças algorítmicas, alterações técnicas e modos de divulgar ações de transparência para se proteger das críticas. Foi uma tentativa de se deslocar da posição de “mero buscador” para assumir, quase demagogicamente, o discurso de que conteúdos importam sim, afinal, já não era mais possível dissociar o desequilíbrio causado no fluxo informacional da missão da empresa.

O *Google* divulga desde 2010 seu relatório de transparência⁸², que, segundo a empresa, são “[...] dados que demonstram como as políticas e as ações de governos e corporações afetam a privacidade, a segurança e o acesso à informação on-line.” (GOOGLE, 2023a). Ao final de 2016, em diversas audiências realizadas no Congresso dos Estados

⁸² Disponível em: <https://transparencyreport.google.com/about?hl=pt-br>. Acesso em: 09 jan. 2023.

Unidos (EUA), o *Google* assumiu o compromisso de realizar ações relacionadas à transparência e à responsabilidade na publicidade eleitoral. Em 2017, reformulou seu relatório de transparência, sinalizando fornecer informações mais contextualizadas e, desde então, passou a divulgar dados sobre: solicitações governamentais de remoção de conteúdo; remoção de conteúdos por falsificação, detalhamento da publicidade política no *Google*, solicitações relativas ao “direito de ser esquecido”, atendendo à ordem do Tribunal Europeu de Justiça e permitindo que cidadãos europeus solicitassem a exclusão de informações sobre si mesmos dos resultados de pesquisa, entre outros temas que sacramentam a atuação da plataforma na ordenação de conteúdo. Em suma, quando enxergamos o *Google* como plataforma *online*, defendemos que o sistema é influente e importante como uma infraestrutura – técnica e/ou social.

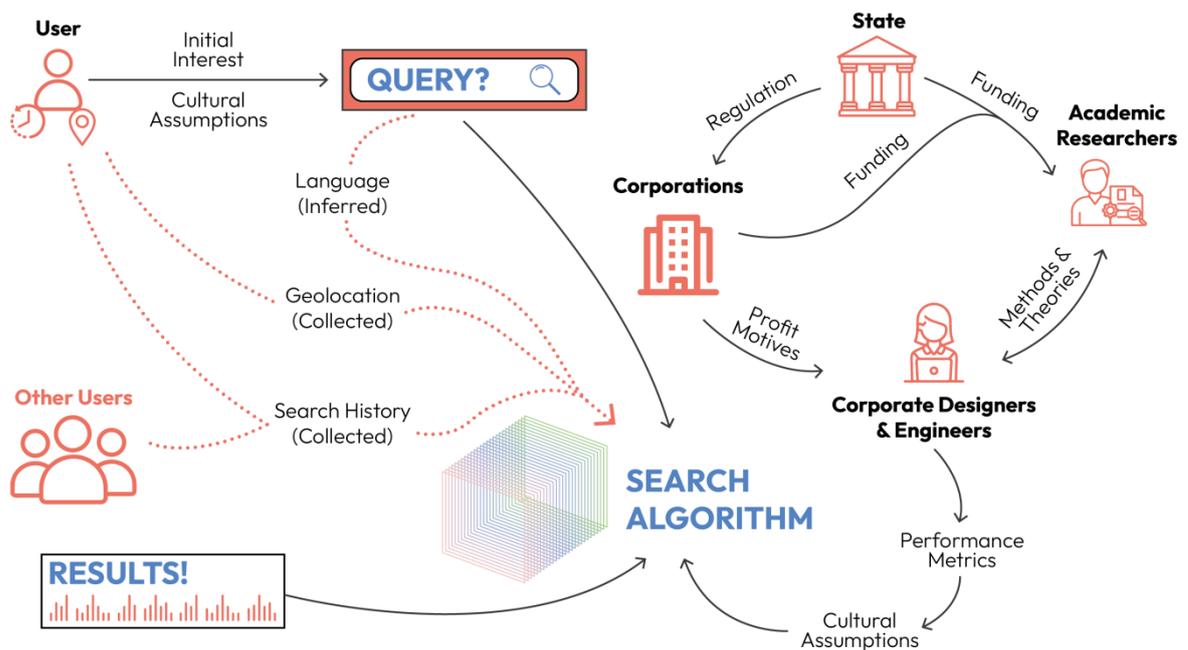
O desenho de motores de busca é uma questão política, em que os criadores fazem escolhas para determinar relevância de conteúdos ou *links*, por isso, precisamos ser céticos em relação às métricas que excluem ou visibilizam determinados *sites*. Ochigame (2021b) faz a seguinte analogia: se colocássemos cientistas da computação para classificar livros em bibliotecas físicas, provavelmente, mediriam-nos por produtividade (mais empréstimos). Atualmente, quando os especialistas em IA dizem que os algoritmos classificam conteúdos por relevância, essas afirmações nunca são inocentes, porque disfarçam prescrições interpretativas e políticas sobre o que é relevante ou irrelevante. Segundo Ochigame (2021b), essas prescrições servem para reproduzir *status quo* muito injusto e reforçam padrões excludentes. Assim como os livros em bibliotecas não devem ser descartados só porque têm baixo número de empréstimos, os textos digitais não podem ser considerados somente pelas taxas de cliques e receitas de anúncio.

Portanto, Ochigame (2021b) postula que algoritmos de plataformas de busca, como *Google*, não estão encontrando resultados mais relevantes, estão, na verdade, escolhendo definir relevância, envolvendo métodos de contar e agrupar citações. Para o autor, a escolha não é objetiva, ela é normativa e baseada em valores culturais e políticos. É que não existe uma métrica independente capaz de dizer quais seriam os resultados de pesquisa mais relevantes para qualquer busca, o que abre portas para os desenvolvedores ajustarem os algoritmos para coletar evidências e decidir qual resultado mais correto. Assim, nascem as controvérsias sobre vieses, porque não há um terreno sólido e objetivo para balancear disputas sobre avaliações algorítmicas (GILLESPIE, 2018).

O desenho minimalista do *Google* – aquela simples barra de busca – esconde todas as complexas correlações (Figura 9) que a plataforma faz em milésimos de segundo na

entrega de resultados; inclusive, esse formato pode ser visto como uma escolha política, afinal, quem irá imaginar a robustez de políticas por trás daquela fenda? A brecha está ali nos convidando a expressar anseios de leitura e o mais irônico é que a “barrinha” promete oferecer a representação total e objetiva para as perguntas/questões, quando o mais equilibrado seria mostrar uma gama mais ampla de perspectivas parciais.

Figura 9 – O que está por trás da barra do *Google*?



Fonte: Ochigame (2021a).

A Figura 9, proposta por Ochigame (2021a), destrincha a engrenagem política e econômica dos motores de busca, mostrando que os resultados da pesquisa são produtos não apenas dos anseios de leitores, mas de lutas complexas entre o Estado, as corporações e a academia. Segundo o autor, os algoritmos de pesquisa codificam as suposições culturais e as métricas de desempenho dos *designers* e engenheiros, que, por sua vez, são moldados pelo lucro corporativo, regulamentos estatais, além de métodos e teorias acadêmicas.

Por mais de duas décadas, um amplo corpo de pesquisa questionou a política dos motores de busca – Introna e Nissembaum (2006), Zimmer (2009), Vaidhyathan (2011), Plantin et al. (2016), Rogers (2016). Mesmo os mais antigos estudos, baseados em observações, já sugeriram que os motores de busca suprimem sistematicamente alguns *sites* em favor de outros, de acordo com interesses privados. Estudos mais recentes têm argumentado que os motores de busca comerciais implementam algoritmos que reforçam as

estruturas sociais existentes, particularmente racistas, sexistas e de marginalização, conforme pontuamos no Capítulo 2. “Os mecanismos de pesquisa não respondem aos interesses dos usuários com resultados objetivamente ‘relevantes’, nem são redutíveis a editores puramente subjetivos com poder ilimitado para escolher e escolher.” (OCHIGAME, 2021a, p. 208).

Para Ochigame (2021a), a construção do motor de busca do *Google*, que baseia a lógica da plataforma, é inseparável das prioridades do negócio de publicidade direcionada, um modo segmentado de as marcas se comunicarem com públicos-alvo. As plataformas de busca comerciais lucram vendendo a segmentação de grupos de consumidores, com base em variáveis como localização, idade, gênero, estilo de vida, histórico de compras, entre outras. As plataformas organizam de forma altamente escalável a oferta e a demanda, mediando a relação entre propaganda e público. Aliás, é importante compreender que o arranjo da publicidade se assemelha com o da mídia tradicional, como a televisão comercial em que anunciantes pagam pela promessa de telespectadores. Da mesma forma, as plataformas *online* atraem anúncios pagos com base na promessa de uso de pesquisa. O alto uso, presumivelmente, atrai anunciantes e, assim, as plataformas de busca vão estabelecendo uma reputação, o que ajuda a explicar um ciclo vicioso de ascensão quase monopolista do *Google*.

Ochigame (2021a) esclarece que a “indústria da propaganda direcionada” molda o campo da recuperação da informação. Mesmo quando estudamos apenas “produção, circulação e leitura” de conteúdo orgânico é fundamental entender a lógica do conteúdo patrocinado ou dos anúncios para reforçar as características do *Google* não apenas como plataforma prestadora do serviço de busca, mas sim uma infraestrutura de interesses privados. Na estratégia orgânica, as empresas investem em técnicas para melhorar ranqueamento, equilibrando a compreensão de comportamento de leitores e de curadoria algorítmica para aparecer nos primeiros resultados, o que pode levar um bom tempo. Na estratégia paga, os resultados de uma segmentação altamente especializada por trazer retorno mais rápido em termos de visibilidade da marca.

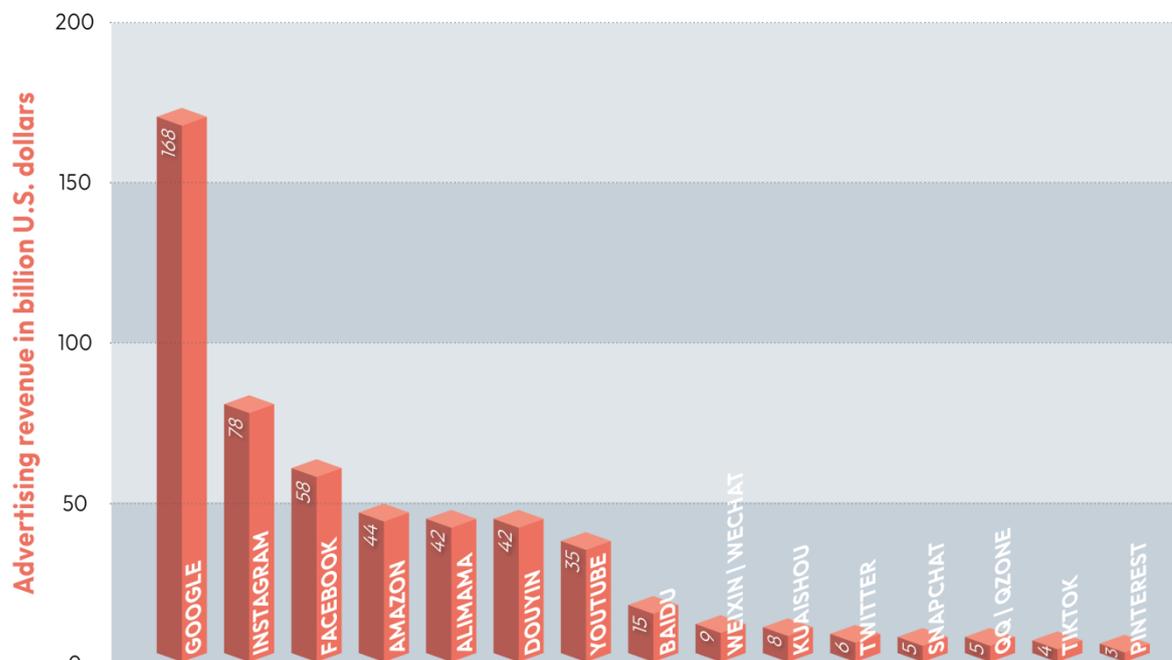
Se os departamentos de relações públicas do Vale do Silício alegarem que seus produtos simplesmente “encontram os resultados mais relevantes e úteis em uma fração de segundo, e apresentá-los em uma forma que o ajuda a encontrar o que procura” – é assim que o *Google* descreve seus algoritmos de busca em seu site – executivos e acionistas sabem melhor. Eles sabem que o objetivo do negócio é ser pago para mostrar a você coisas que você não está procurando: anúncios. (OCHIGAME, 2021a, p. 108).

Segundo o Departamento de Justiça dos Estados Unidos, por meio da Divisão Antitruste, os modelos de anúncio do *Google* se baseiam na venda de termos e palavras-

chave. Marcas, empresas e organizações pagam para que seus conteúdos ou anúncios sejam exibidos quando as pessoas pesquisarem determinadas combinações, e claro, com a demarcação de que são conteúdos pagos. Assim, esses anúncios são mostrados acima e abaixo dos resultados da pesquisa do *Google* ou na rede *display*, ao lado do conteúdo correspondente a eles, em páginas parceiras do *Google*. Por exemplo, a marca *Havaianas* pode adicionar “melhor chinelo” como palavra-chave de sua campanha no *Google*. Quando as pessoas digitam essa combinação, o anúncio da empresa será exibido na página de resultados ou poderá aparecer em *displays* de *sites* que produzem conteúdo sobre moda.

De acordo com levantamentos do *Securities and Exchange Commission dos EUA*, a receita de publicidade do *Google* nos EUA, em 2022, foi de aproximadamente US\$ 70,2 bilhões. Este é um valor líquido que inclui anúncios do *Google* vistos em *desktops*, *laptops*, *smartphones*, *tablets* e outros dispositivos, bem como vários formatos de anúncios, incluindo aqueles no *YouTube*, uma empresa do grupo *Google/Alphabet Inc*. Foi a primeira vez que a receita ultrapassou a marca de US\$ 70 bilhões naquele país e chegou a US\$ 168 bilhões no recorte mundial (Figura 10), o que distancia a *Google* de todas outras empresas que operam no mercado de publicidade digital. Os dados já coletados e as projeções indicam que as receitas do *Google* em publicidade crescem, em média, 14,8% ao ano.

Figura 10 – Receita de publicidade das principais empresas de venda de anúncios digitais em todo o mundo, em 2022. (em bilhões de dólares americanos)



Fonte: Statista, 2023⁸³.

⁸³ Disponível em: <https://www.statista.com/statistics/1202672/digital-ad-revenue-ad-selling-companies-worldwide/>. Acesso em: 22 fev. 2023.

Introna e Nissebaum (2006) afirmam que motores de busca priorizam suas seleções e listagens de acordo com o lance mais alto que terão no mercado. No exemplo da *Havaianas*, que demos anteriormente, o que poderia acontecer é a marca *Ipanema* também comprar os termos “melhor chinelo”, assim o *Google* fica responsável por decidir, por meio de um leilão⁸⁴, qual anúncio será exibido para aquela palavra-chave. Vale reforçar, sobre os resultados de busca orgânica, que o *Google* realmente não recebe dinheiro para posicionar *links*, conforme a própria empresa esclarece: “O Google nunca aceita pagamentos para rastrear um site com mais frequência. As mesmas ferramentas estão disponíveis para todos os sites para garantir os melhores resultados possíveis para nossos usuários.” (GOOGLE, 2022d). Todavia, não acreditamos na isonomia de ferramentas para todos.

No webinar⁸⁵ oferecido pela *Google/Alphabet Inc.*, o jornalista e *Public Liaison for Search* Danny Sullivan chegou a dar um exemplo pitoresco explicando que uma vendedora de doces artesanais do Norte do Brasil tem as mesmas ferramentas que qualquer grande indústria alimentícia e que o *Google* dá a essa empreendedora as condições de ocupar o primeiro lugar na página de resultados do mecanismo de pesquisa (*Search Engine Results Page – SERP*). Nossa tese é de que a lógica construtiva desses mecanismos já privilegia *sites* com mais recursos para investir nos critérios conhecidos para ranqueamento e isso é parte fundamental da engrenagem de negócios. Senão faria pouquíssimo sentido existirem empresas destinadas a cobrar para limpar a reputação *online* de pessoas, apagando informações comprometedoras, conforme exemplificamos o caso de uma multinacional⁸⁶ que prestava esse serviço a criminosos condenados no Brasil.

Por fim, ao falar de política dos motores de busca, é importante retomar as questões sobre acesso, lembrando que “[...] o modo como funcionam (os buscadores) parece estar em desacordo com a ideologia convincente da Web como um bem público” (INTRONA; NISSEMBAUM, 2006, p. 178, tradução nossa)⁸⁷. Para esses autores, ironicamente, a *Web* foi vista por muitos como meio de contornar a hegemonia da mídia tradicional, porque foi concebida como espaço de bem público. Sendo assim, a *Web* parece ser idealmente pública

⁸⁴ O processo para decidir se algum anúncio será exibido, quais anúncios aparecerão em uma pesquisa específica e a ordem de exibição deles na página. A classificação combina o lance da marca, a qualidade do anúncio, os limites mínimos de classificação do anúncio, o contexto da pesquisa e o impacto esperado das extensões e de outros formatos de anúncio. Disponível em <https://support.google.com/google-ads/answer/142918>. Acesso em 22 fev 2023.

⁸⁵ Ver nota 13

⁸⁶ Caso descrito na Introdução da tese

⁸⁷ Tradução nossa para: “[...] because of the way they function, but also because the way they function seems to be at odds with the compelling ideology of the Web as a public good”.

mesmo porque sua infraestrutura de *hardware* e *software* não deveria ser propriedade integral de qualquer pessoa, instituição ou nação. “A natureza colaborativa de grande parte da atividade na Web leva a uma sensação de que ela não é simplesmente sem dono, mas de propriedade coletiva [...] algumas das funções de outros espaços públicos tradicionais – museus, parques, praias e escolas.” (INTRONA; NISSEMBAUM, 2006, p. 180, tradução nossa)⁸⁸.

Acontece que o desenho e a função dos motores de busca nos levam a dois paradoxos que interferem no funcionamento da *Web* como “propriedade coletiva”: 1) mais acesso à informação *versus* menos chances de localizar; e 2) capacidade de rastrear *links* na *Web* *versus* tamanho do índice do motor de busca. O primeiro paradoxo revela que, com o estreitamento das opções na circulação de textos, mediados por plataformas de busca, a *Web* como um bem público deixa de existir. Há menos oportunidades para localizar vários tipos de informações, indivíduos e organizações, ou seja, uma redução da gama de capacidades deliberativas e recreativas. “Se o acesso à Web for entendido como o acesso dos requerentes a todos os desses recursos, o resultado de mecanismos de busca tendenciosos equivale a uma diminuição do acesso à Web.” (INTRONA; NISSEMBAUM, 2006, p. 180, tradução nossa)⁸⁹. Esses autores reforçam que o bem público da *Web* não está apenas em seu funcionamento como um repositório para quem busca encontrar coisas, mas como um espaço para quem tem algo a oferecer (bens, serviços, pontos de vista, ativismo político etc.). O custo de um viés de mecanismo de busca pode ser ainda maior para proprietários de *sites* que desejam ser encontrados se os mecanismos de busca restringirem sistematicamente o escopo do que os leitores podem receber e quais *sites* podem ser vistos, assim perderemos o valor geral da *Web* como um espaço público e como uma fonte de informação amplamente inclusiva.

O segundo paradoxo está concentrado no entorno de uma das maiores controvérsias⁹⁰ entre especialistas em SEO, desenvolvedores e representantes do *Google* que é, justamente, a “capacidade” do motor de busca. A polêmica se materializa no entorno de perguntas como: Quanto da *Web* está escondida? Quantas *webpages* o *Google* indexa? Quantos *sites* o *Google* é capaz de rastrear? Quantas *URLs* únicas o *Google* pode encontrar? Quantos *sites* existem na *Web*? Destrinchamos propositalmente essas perguntas, que muitas

⁸⁸ Tradução nossa para: “The collaborative nature of much of the activity on the Web leads to a sense of the Web’s being not simply unowned but collectively owned. [...] is some of the functions of other traditional public spaces—museums, parks, beaches, and schools.”

⁸⁹ Tradução nossa para: “Web is understood as access by seekers to all of these resources, then the outcome of biased search engines amounts to a shrinking of access to the Web.”

⁹⁰ Apesar de não ser um conceito central da nossa tese, vale destacar que o termo é usado aqui no sentido dado pelos Estudos de Ciência e Tecnologia para caracterizar situações de instabilidade e conflito (D’ANDRÉA, 2020). Há, claramente, disputas e incertezas em torno da capacidade do motor de busca do *Google*.

vezes se confundem, porque existirão respostas diferentes e em fontes diversas⁹¹, o que sustenta a controvérsia. Embora pareça que o *Google* indexa tudo o que imaginamos, a *Web* como um todo é algo muito maior. Especialistas calculam a capacidade de indexação do *Google*, mas a informação não é oficialmente divulgada pela companhia. A empresa suspendeu a publicização desse dado em 2005, o que fez com que especialistas criassem séries de projeções estatísticas em torno desse número. O que o *Google* (2022d) declara é a capacidade de indexar mais de 100 milhões de *gigabytes* em tamanho. No entanto, os detalhes que se escondem na controvérsia são a diferença entre as páginas que o *Google* conhece e as páginas que o *Google* realmente indexa, afinal, a internet está cheia de lixo ou *spams*. Isso significa que a quantidade de *links* encontrados ou rastreados não é igual ao tamanho do índice ou à gama de *links* que serão ranqueados. O paradoxo se reforça no fato de que o tamanho do índice não é mais o objetivo das plataformas de busca, pois elas perceberam que precisam manter o índice enxuto, garantindo que ele inclua apenas os melhores resultados. A qualidade dos resultados indexados é o que importa e, talvez, seja por isso que o *Google* parou de relatar publicamente o tamanho do seu índice.

Van den Bosch, Bogers, Kunder (2016) provaram que o tamanho dos índices do *Google* e do *Bing* variaram drasticamente entre 2006 e 2015, e que o *Google* tem, indiscutivelmente, maior capacidade de indexação. Tanta variabilidade pode ser explicada por mudanças na infraestrutura de indexação e classificação, ou seja, pelos *updates* em algoritmos responsáveis por essas funções, o que, para os autores, lança mais dúvidas sobre os mecanismos de pesquisa da *Web* e reforça que a “Googleologia é ciência ruim” (KILGARRIFF 2007, p. 147 *apud* VAN DEN BOSCH, BOGERS, KUNDER, 2016, n.p.). Para os autores, os mecanismos de busca comerciais parecem exibir variações em seu funcionamento que não se vinculam naturalmente ao *corpus* que pretendem indexar.

Enfim, dada à vastidão da *Web*, a guarda cerrada dos algoritmos e a obscuridade da tecnologia, não é tão surpreendente o resultado da pesquisa de confiança em notícias mostrada, anteriormente, na Figura 8, em que os leitores confiam no que veem mediados pelo *Google*. As pessoas que usam a barra de busca talvez não estejam familiarizadas com os mecanismos sistemáticos que conduzem a plataforma. Isso nos faz lembrar o fato de que no século XX as pessoas também não compreendiam o funcionamento da lógica comercial da

⁹¹ Algumas dessas fontes são o projeto <https://www.worldwidewebsite.com/>, da Tilbuth University; o estudo *Estimating search engine index size variability: a 9-year longitudinal study*; sites especializados em estatísticas, como <https://www.statista.com/>, <https://www.internetlivestats.com/>; e o próprio *Google* em seus relatórios anuais, páginas de apresentação do serviço e *updates* divulgados nos *blog* oficial.

grande mídia. Muitos demoraram a saber – ou ainda não sabem – que um jornal não vende notícias, mas sim espaços para anunciantes.

Por tudo isso, Introna e Nissenbaum (2006) vão discutir se é certo que o mercado de motores de busca se movimenta livremente e se autorregule, permitindo que se destaquem os “melhores” buscadores e fracassem outros. Eles defendem que não podemos deixar o mercado decidir, pois acreditam que a informação está na lista de bens que não podem ser regulados pelo mercado, porque sabemos que não há limitações éticas para as empresas provedoras das plataformas.

4.3.2 Funcionamento

A essência do funcionamento do motor de busca – que subsidia a plataforma *Google* – está no caráter contingencial, ou seja, no dinamismo que consegue imprimir ao que entrega como verdade construída para os leitores. Sem qualquer intenção de nos aprofundar, vale aqui um resgate da Teoria Pragmática da Verdade⁹² sobre a verdade ser um processo de experimentação, descoberta, ligada às experiências com as coisas do mundo com nossas conversações, aliás, reforçando que é uma construção de realidades, pois é formulada a partir de perspectivas diferentes. Motores de busca trabalham com o dinamismo da verdade em modo contingencial. Os critérios que baseiam a lógica de funcionamento dos motores de busca consideram o que Vaidhyanathan (2011, p. 75) chama de “processo dinâmico de verificação por aprovação coletiva”. Os algoritmos absorvem de modo extremamente mutante o que os leitores estão aprovando como *links* (textos e conteúdos) relevantes e, por isso, quem os produz trabalha nos esforços da “contingência da performatividade” (CASTRO, 2019, p. 21) e na produção de “mercadoria cultural contingente” (NIEBORG; POELL, 2018), conforme argumentamos no Capítulo 2.

Não podemos negar a genialidade disso para os negócios das plataformas de busca, em especial, do *Google*. Também não vamos negligenciar que há muitos problemas nessa verdade contingenciada apresentada pelos motores de busca e, muitos deles, já apresentamos no decorrer da tese, com destaque para a não neutralidade. Talvez, os teóricos da Teoria Pragmática da Verdade já soubessem que alguns vieses poderiam influenciar mais o

⁹² Estudos de pesquisadores sobre retórica, discurso e argumentação, como Wiliam James (*Pragmatismo*, 1907), Charles Peirce (*The essential Peirce: selected philosophical writings*, 1867), Richard Rorty (*Objectivity, relativism and truth*, 1991), Friedrich Nietzsche (*Assim falava Zaratustra*, n.d.), além de releituras mais recentes do tema feitas por Muniz Sodré (*A narração do fato: notas para uma teoria do acontecimento*, 2009).

consenso de verdade do que outros e são essas distorções o que os motores de busca reforçam em uma escala brutal.

Inicialmente, o *Google* funcionava a partir de sistemas de “avaliação de importância” dos *links* espalhados na *Web* e que podem ser rastreados (BRIN; PAGE, 1998). Ainda hoje, esses esquemas são fundamentais para a busca em ambiente digital, mas foram agregadas outras lógicas ao mecanismo de apresentação de respostas. O sistema pioneiro de avaliação de páginas do *Google*, o *Pagerank*, foi desenvolvido pelos fundadores da empresa Larry Page e Sergey Brin (1998). Basicamente, o sistema compreendia que uma página tinha muito valor de acordo com a quantidade e qualidade dos *links* que direcionavam para ela. O sistema era descrito como “mecanismo de pesquisa da *Web* hipertextual em larga escala”, e vale lembrar aqui o paradigma funcional da palavra “hipertextual”, como característica intrínseca à lógica da rede de hiperconectividade entre *links*, textos e conteúdos. Quando a patente⁹³ foi apresentada, em 1998, entendeu-se que o algoritmo enxergava o *link* de um *site* para outro como um voto de confiança e autoridade, portanto, quanto mais *links* (votos) apontam para uma página, mais se deve confiar nela e, conseqüentemente, seu ranqueamento deve ser mais alto. Assim, se popularizou a métrica original e única do *Google* para calcular a importância de uma página e classificá-la.

O processo como um todo pode ser resumido em: algoritmos esquadriham seu vasto índice de páginas na rede à procura de correspondências com o texto digitado por leitores na barra de busca; após encontrar as correspondências, avaliam o quesito “relevância” daqueles *links*; a relevância, por sua vez, tende a significar “valor”, mas, como já argumentamos, é um valor relativo e condicional. Durante muito tempo, era possível medir (verificando um índice de 0 a 10) o *PageRank* de uma página para saber se *link* era mais ou menos elegível aos olhos do *Google*. Não é mais possível ver essa nota, pois o recurso foi aposentado em 2016, o que não significa que os princípios de determinação de importância mudaram ou deixaram de ser fundadores. Quando a nota *PageRank* era um dado aberto ao público, ela se tornou uma métrica para usos táticos de plataformas de busca, em especial dos especialistas de SEO. Muitos desenvolvedores *Web* e produtores de texto focaram mais aumentar o *PageRank* do que em cuidar de uma experiência sólida e bons conteúdos para os leitores. Se olharmos da perspectiva do *Google*, permitir a esses especialistas conhecer o índice *PageRank* era um jeito de manter a aura algorítmica da neutralidade, afinal, aquele número parecia materializar a avaliação precisa da autoridade de uma página da *Web*. Mas,

⁹³ *Method for node ranking in a linked database*. Disponível em: <https://patents.google.com/patent/US6285999B1/en>. Acesso em: 09 jan. 2023.

Ochigame e Holston (2016) apresentam uma visão mais apurada do *PageRank* como mecanismo seletor:

Embora eles descrevam a pontuação do PageRank de uma página da web como uma medida objetiva de sua importância, é visto com mais precisão como uma medida de popularidade, uma vez que se baseia em quantas outras páginas vinculam à página, aumentando o peso daquelas páginas que são vinculadas por muitas outras. Em essência, o PageRank é uma versão mais sofisticada de uma contagem de citações acadêmicas, com a principal diferença de que as contagens de citações não consideram se os artigos citados são populares. (OCHIGAME; HOLSTON, 2016, p. 91, tradução nossa)⁹⁴.

Desde 2006, o *Google* não usa mais a versão original do *PageRank*, mas ele continua sendo um dos sistemas para identificar fontes confiáveis entre centenas de bilhões de páginas. Naquele ano, a empresa publicou uma patente substituta⁹⁵, sem uma referência direta ao algoritmo pioneiro, mas descrevendo um sistema que tenta enxergar a *Web* como grafos de *links*, compostos por *hubs* ou “sites sementes” (*seed pages*). Esses *sites* são, segundo o *Google* (2023l), identificados por sua alta qualidade, boa conectividade, diversidade para cobrir variedade de campos de interesse público, grande número de *links* úteis que ajudam a identificar outros “sites sementes”. Quer dizer que a questão da hipertextualidade continuou sendo fator preponderante, pois quanto mais próximo um *site* está de um *hub* ou de um “conjunto semente”, maior a pontuação dele no ranqueamento. A “avaliação de importância” ainda está presente, ou seja, é interessante aos olhos do *Google* que o *site* esteja conectado e recebendo “votos de confiança” de um *site* semente.

Atualmente, os algoritmos incorporam informações mais contextuais e semânticas sobre os *sites*, analisando onde eles estão hospedados, quem são os autores de conteúdo que ali atuam, além das formas como os conteúdos são publicados. Esses indicadores ajudam a estimar a importância de um *link*, isto é, podemos concordar que “[...] relevante é um julgamento fluido e carregado de significados, tão aberto à interpretação quanto alguns dos termos avaliativos sobre os quais estudiosos da mídia já se debruçaram, como ‘notável’ ou ‘popular’”. (GILLESPIE, 2018, p. 103).

O *Google* e o *Bing* já se envolveram em uma guerra de competição por indicadores de relevância. O *Google* (2022d) sempre sustentou que mantinha cerca de 200

⁹⁴ Tradução nossa para: “Although they describe the PageRank score of a web page as an objective measure of its importance, it is more accurately seen as a measure of popularity, since it is based upon how many other pages link to the page, increasing the weight of those pages that are themselves linked from many others.¹⁵ In essence, PageRank is a more sophisticated version of an academic citation count, with the main difference being that citation counts do not consider whether citing articles are themselves popular.”

⁹⁵ Disponível em: <https://patents.google.com/patent/US9165040B1/en>. Acesso em: 09 jan. 2023.

indicadores, mas quando o *Bing* anunciou, em 2010, que usava 1.000 sinais, então, o *Google* acabou publicizando que seus 200 indicadores tinham até 50 variações, levando seu total a quase 10 mil sinais para relevância. Vale ressaltar que o *Bing* e os outros motores de busca do mercado têm uma governança própria, com documentos e diretrizes. No entanto, mesmo que nossa pesquisa se dedique exclusivamente ao *Google*, sabemos que as métricas de relevância são semelhantes. Quando generalizamos nossas observações sobre plataformas de busca, consideramos o funcionamento parecido, que varia a partir do modo como cada motor de busca monta as equações de ranqueamento, o que é um segredo de negócio.

Longe de nós quisermos detalhar rigorosamente tais indicadores que são, na verdade, a grande caixa-preta do *Google*, juntamente com a personalização de resultados por usuário (geolocalização, perfil do usuário, idioma e outros itens cujo escopo nunca fica claro). Discutiremos isso no decorrer deste capítulo, mas, por hora, fundamental é compreender que os motores de busca comerciais são regidos, basicamente, por dois critérios: a) similaridade, que reúne as semelhanças, correspondências, correlações entre as ações e características dos usuários, com o objetivo de exibir itens que correspondam aos gostos, opiniões e preferências; e b) popularidade, que calcula ou contabiliza desempenho/performance para atribuir valor a *links* (textos e conteúdos). E mesmo que promovam mudanças constantemente nos algoritmos, conforme mostraremos adiante, o mecanismo básico se desencadeia a partir desses dois critérios.

Introna e Nissenbaum (2006) explicam que a natureza da conexão entre os motores de busca e a *Web* se dá em três etapas: rastreamento, indexação e ranqueamento (ou classificação). Nas duas primeiras etapas, algoritmos vão avaliar a similaridade da palavra pesquisada com os *links* (textos e conteúdos) disponíveis para serem rastreados, correlacionado a gostos, opiniões e preferências dos leitores. Na última etapa, será avaliada e calculada a popularidade do *link*. Esse critério, por sua vez, poderá ser expresso em alguns itens possíveis de serem contabilizados, como, por exemplo, quantidade e a qualidade de *links* que apontam para uma página, o que nos leva novamente ao princípio do *PageRank*, ou ainda, nas avaliações populares daquele *site* que indiquem aprovação coletiva.

Os exemplos de polissemia nos ajudam a compreender como similaridade e popularidade amarram as três etapas de busca do *Google*. Quando um leitor busca por tênis (esporte ou calçado?), *apple* (empresa de tecnologia ou cidade ou sobrenome ou maçã?), mercúrio (planeta ou elemento químico?), *windows* (empresa de tecnologia ou sistema operacional ou janela?), lima (capital do peru ou fruta ou sobrenome ou ferramenta?), os algoritmos vão se valer da similaridade para coletar os *links* de entrega, mas precisarão de

algo mais para vencer de forma, minimamente eficiente, a polissemia. Sabemos que, atualmente, muitos contextos podem ajudar o *Google* a entregar a resposta sobre cada um desses termos polissêmicos, a exemplo do histórico de busca de um usuário, conforme a própria plataforma informa (*GOOGLE, 2022d*). Mas, em princípio, os algoritmos se valerão da grande base de dados referente às muitas buscas semelhantes feitas anteriormente e identificarão quais resultados foram mais satisfatórios para a maioria das pessoas, inclusive numa complexa disponibilidade de respostas em idiomas distintos.

É um lance estatístico que o *Google* (*2022b*) chama de “interpretação dominante”: o que a grande maioria das pessoas quer dizer com um termo ou frase de busca. Por exemplo, o termo “Amazon” é Amazônia em inglês e se refere a uma região geográfica, a maior floresta tropical do mundo. No entanto, a interpretação dominante, o que a grande maioria das pessoas quer dizer ao buscar de forma genérica “Amazon” na *Web* pode estar relacionado à gigante do *e-commerce* mundial. Nesse caso, vale lembrar as palavras de Bernhard Rieder (2018, p. 138): “O Google pode ser uma empresa de publicidade em primeiro lugar, mas em segundo lugar é empresa de estatística”, o que ressalta a indissociabilidade entre as estratégias comerciais e a datificação. E ainda:

A ordenação de informações algorítmicas depende muito de técnicas estatísticas para domar a abundância de informações e em grande parte subscreve uma epistemologia e um ethos pragmáticos que buscam tornar grandes quantidades de informações não apenas visíveis e interpretáveis, mas também navegáveis, acionáveis, e (economicamente) explorável. (RIEDER, 2020, p. 39).

Na trajetória de 25 anos do *Google*, há muitas atualizações em seus processos que alimentam o crescimento da IA⁹⁶. Um dos principais objetivos da IA é construir sistemas que possam ler a *Web* e responder a perguntas complexas sobre qualquer assunto. O *Natural Questions (NQ) Project* do *Google*, por exemplo, reúne um sistema de perguntas e respostas (QA) que podem ter implicações na maneira como acessamos as informações. Além disso, responder a perguntas de domínio aberto é uma tarefa de referência no desenvolvimento da IA, pois entender o texto e ser capaz de responder a perguntas sobre ele é algo que geralmente associamos à inteligência. O *corpus NQ* contém perguntas de usuários reais e requer que os sistemas de controle de qualidade leiam e compreendam um artigo inteiro da *Wikipédia* que pode ou não conter a resposta para a pergunta. Isso demonstra que o investimento em IA

⁹⁶ Disponível em: <https://ai.google.com/research/NaturalQuestions/visualization>. Acesso em: 10 jan. 2023.

precede o BARD⁹⁷ e ChatGPT⁹⁸. O próprio *Google* argumentou⁹⁹, em meio ao *frisson* de 2023 sobre OpenAI¹⁰⁰, que o *Core Update* BERT foi revolucionário na compreensão das complexidades da linguagem humana e esse *update* ocorreu em 2019.

Entretanto, sustentamos que todo esse desenvolvimento do *Google* não descola o motor de busca da essência descrita no artigo de lançamento (BRIN; PAGE, 1998), em 1998. É como se o ferramental adquirido servisse para melhorar os modos de operar a similaridade e a popularidade, que já estavam muito presentes na descrição inicial do mecanismo de busca. Vejamos alguns trechos:

As listas mantidas por humanos cobrem tópicos populares de forma eficaz, mas são subjetivas, caras para construir e manter, lentas para melhorar e não podem cobrir todos os tópicos esotéricos. Para os **assuntos mais populares**, uma simples pesquisa de **correspondência de texto restrita a títulos de páginas da Web** tem um desempenho admirável quando o PageRank prioriza os resultados. (BRIN; PAGE, 1998, tradução nossa, grifos nossos)¹⁰¹.

Ao dizer sobre “listas mantidas por humanos”, os autores do artigo estavam se referindo à lógica de funcionamento de outras ferramentas, como *Yahoo*, que naquela época eram baseadas em diretórios de busca, e não nos “*web crawlers*” propostos pelo *Google*. Os diretórios de busca distinguem, coletam e descrevem outros *sites*, não tendo a intenção de uma entrega semântica completa de conteúdos, como os motores de busca. No mesmo texto, os fundadores apontavam para o uso, no processo de ranqueamento, de metainformações externas que podem ser inferidas sobre um documento na *Web*, mas não estão contidas nele. Novamente, aparece o critério de popularidade, conforme o trecho a seguir:

Exemplos de metainformações externas incluem coisas como reputação da fonte, frequência de atualização, qualidade, **popularidade ou uso e citações**. Não apenas as possíveis fontes de metainformações externas variam, mas as coisas que estão

⁹⁷ Serviço experimental de conversação em IA do *Google* e alimentado por LaMDA (um modelo de linguagem neural conversacional). Disponível em: <https://blog.google/technology/ai/bard-google-ai-search-updates/>. Acesso em: 23 fev. 2023.

⁹⁸ Protótipo de um *chatbot* com IA desenvolvido pela OpenAI.

⁹⁹ Um próximo passo importante na nossa jornada de IA. Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/um-proximo-passo-importante-na-nossa-jornada-de-ia/>. Acesso em: 06 jan. 2023.

¹⁰⁰ Laboratório de pesquisas em IA que criou o ChatGPT. É uma organização subsidiária e constituída pela empresa com fins lucrativos OpenAI LP, fundada por figurões do Vale do Silício, como Elon Musk. Disponível em: <https://pt.wikipedia.org/wiki/OpenAI>. Acesso em: 23 fev. 2023.

¹⁰¹ Tradução nossa para: “Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all esoteric topics. Automated search engines that rely on keyword matching usually return too many low quality matches.”

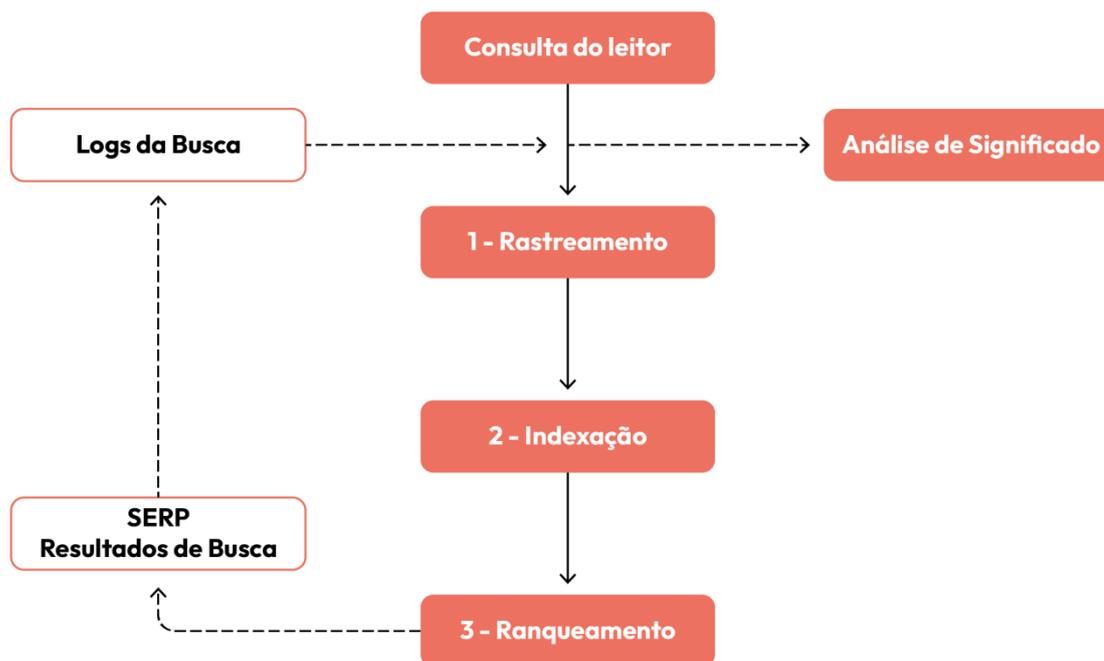
sendo medidas também variam em muitas ordens de magnitude. (BRIN; PAGE, 1998, tradução nossa, grifo nosso)¹⁰².

Ora, se o critério popularidade engloba os itens que a plataforma de busca considera “mais interessante”, “top”, “melhor”, logo, o produtor de texto trabalha em busca desses marcadores de desempenho. O contraponto, segundo análise de Ochigame e Holston (2016), é que a popularidade pode suprimir dissidências. Se o critério de similaridade permite circular conteúdos semelhantes aos mais propensos a atrair interação, em que plataformas de busca perfilam usuários com características de navegação semelhantes, o produtor de texto trabalha em busca desses marcadores de desempenho. Nesse caso, também há um contraponto e, segundo Ochigame e Holston (2016), está no fato de que a similaridade pode suprimir divergências.

Portanto, defendemos que as lógicas de produção textual pretendem atender os critérios de plataformas de busca comerciais, como a *Google*, e verificamos isso nas experimentações relatadas no Capítulo 5. Mesmo que os produtores não conheçam exatamente o funcionamento das etapas – rastreamento, indexação e ranqueamento –, nem mesmo os algoritmos que as fazem “rodar”, muitos deles já naturalizaram práticas que fazem suas produções circularem e terem (ou não) visibilidade em ambiente digital. Dessa forma, trabalham naquele espectro automático, performático, cercado, desconhecedor, contrafluxo, subversivo – que hipotetizamos no Capítulo 2.

¹⁰² Tradução nossa para: “Examples of external meta information include things like reputation of the source, update frequency, quality, popularity or usage, and citations. Not only are the possible sources of external meta information varied, but the things that are being measured vary many orders of magnitude as well.”

Figura 11 – Como funciona a busca no *Google*



Fonte: Elaborado pela autora, 2023.

São poucas as alternativas de serviços de busca que não repetem esse funcionamento. Um exemplo é o *DuckDuckGo*, que representa uma ruptura porque resguarda privacidade de dados do usuário, mas simboliza continuidade em dois sentidos: na manutenção do modelo de negócio ancorado em propaganda direcionada (o esquema de leilões de palavras-chave) e na premissa dos algoritmos baseados, fundamentalmente, em popularidade e similaridade. No *DuckDuckGo* é como se os *logs* de busca (DIAGRAMA 1), informação que retroalimenta o sistema, trouxessem menos dados privados do usuário, porém dados suficientes para entregas de resultados diferentes toda vez que usamos o mecanismo de pesquisa. É um modelo que prioriza a privacidade, ao contrário do *Google* que ganha mais coletando os dados. Criadores do *DuckDuckGo* já afirmaram em mais de uma entrevista, que o *Google* não precisaria coletar tantos dados para uma entrega de busca eficiente, mas ele o faz porque lucra muito com esse banco de informações privilegiadas. Para Zimmer (2008), uso generalizado de *cookies*¹⁰³ da *Web*, que, inicialmente, pretendia aumentar o fluxo de conteúdo protegido por direitos autorais *online*, resultou em um sistema capaz de rastrear o

¹⁰³ Arquivos que contêm pequenos fragmentos de dados – como um nome de usuário e localização – que são trocados entre o computador de um usuário.

consumo de materiais intelectuais pelos usuários, diminuindo a liberdade de consumir conteúdo, além de representar uma ameaça à privacidade dos hábitos de leitura, visualização e audição dos indivíduos.

Zimmer (2008) ainda recorre a Helen Nissenbaum para explicar as violações do *Google* no que a autora nomeia como “privacidade como integridade contextual”. Existe um fluxo de informações na sociedade regido por normas situacionais que regulam a privacidade de dados. É perfeitamente cabível que a secretária do meu médico peça o número do meu RG para iniciar meu atendimento, mas é completamente descabido que a mãe de uma criança com quem meus filhos brincam no parque, faça esse mesmo pedido para que nossas crianças possam compartilhar um brinquedo. Os tipos de informação se encaixam em uma dicotomia de público *versus* privado, mas os contextos ajudam a resolver, quase sempre, essa dualidade. Se a introdução de uma tecnologia ou prática em um determinado contexto for conflitante com as normas informacionais vigentes, uma bandeira vermelha é levantada, indicando que a integridade contextual pode ter sido violada. Parece ser esse o sinal vermelho em relação ao *Google*.

Não é mais aceitável se esconder atrás da retórica de que nenhuma informação privada é divulgada ao utilizar as ferramentas que compõem o mecanismo de busca perfeito, ou que a informação compartilhada é simplesmente a mesma fornecida em outros cenários de busca de informações. Revelar esse tipo de transgressão das normas informacionais ajuda a expor a barganha faustiana implícita na procura pelo mecanismo de busca perfeito. (ZIMMER, 2008, p. 120).

Diante de tantas camadas de complexidade – privacidade, critérios de seleção, determinação de relevância, modelo de negócio, contingenciamento –, a existência de um mecanismo de busca perfeito parece utópica. Ochigame (2021a) acredita que há espaço para iniciativas divergentes ou dissidentes, nacionais ou regionais, e que podem ser importantes para a luta social, uma vez que textos circulantes determinam a luta. Antes disso, é fundamental conhecer as lógicas de racionalidade dos sistemas de busca ortodoxas e como elas refletem interesse e premissas das *Big Five*. Afinal, como funcionam as três etapas no *Google*? Compreendê-las pode nos ajudar a clarear ainda mais o entrelaçamento entre o *Google* e as práticas de escrita e edição.

4.3.2.1 Rastreamento

Primeiramente, é importante entender o índice do *Google*. Ele é um grande banco de dados que reúne informações coletadas por algoritmos do *Google* na *Web*. Nesse conjunto de armazenagem, estão todas as páginas conhecidas pelo *Google* e que ele avaliou serem úteis para estar no *hall* de possíveis informações para ranqueamento. Para construir o índice, os algoritmos do *Google* rastreiam e encontram uma página, lêem e analisam o conteúdo dela para, enfim, adicioná-la ao índice. Esse banco de dados, que, conforme já citamos, possui mais de 100 milhões de *gigabytes* de tamanho (*GOOGLE, 2022d*), contém também mais do que o conteúdo originalmente criado na *Web*, já que informações úteis podem ser localizadas em outras fontes, como, por exemplo, livros digitalizados e que fazem parte do projeto *Google Books*. Segundo o *Google* (2022d), é como um “índice de um livro” com uma entrada para cada palavra encontrada em cada página indexada. Quando adicionam uma página da *Web*, adicionam também as entradas para todas as palavras dela.

O rastreamento é um trabalho constante de motores de busca para descobrir páginas úteis e adicioná-las ao índice. Esse processo é denominado “descoberta de URLs” e ajuda a manter um conjunto de páginas a serem visitadas para entrega de respostas. Sendo assim, quando um leitor faz uma consulta na barra de busca, o *Google* procura centenas e bilhões de páginas dentro do índice em busca de correspondências textuais (similaridade) que atendam ao “gosto” perfilado do leitor (*GOOGLE, 2022d*). Imediatamente após a entrada do termo de pesquisa (*input*) e antes de circularem os *webcrawlers*, responsáveis por rastrear as páginas, o *Google* analisa os significados, procurando por erros, compreendendo as maneiras pelas quais as pessoas escrevem palavras incorretamente e usa isso para prever o que o leitor realmente quis dizer (ou quis escrever). É como se algoritmos de análise de significado criassem modelos de linguagem para tentar decifrar como as palavras digitadas na caixa de pesquisa podem corresponder aos conteúdos disponíveis. Por exemplo, se um leitor digitar erroneamente “mochila para ecoturimo”, faltando uma letra, os algoritmos vão soletrar a expressão e perceber que a maioria das buscas nesse contexto se referem a “mochila para ecoturismo”. O *Google* (2023d) afirma ter um sofisticado sistema de sinônimos, que permite encontrar documentos relevantes mesmo quando não contêm as palavras exatas usadas. Por exemplo, se o leitor procurar por “mudar o brilho do *laptop*”, mas o fabricante escreveu “ajustar o brilho do *laptop*”, os sistemas entendem que as palavras e a intenção são correlacionadas (similaridade).

Após o processo de compreensão de significado, a palavra-chave será usada pelos *webcrawlers* percorrerem páginas buscando correspondências textuais. O nome genérico dado ao rastreador é *Googlebot*. Esses rastreadores acessam automaticamente páginas da *Web*

disponíveis ao público e seguem os *links* desses *sites*, da mesma forma que um leitor faria se estivesse navegando no mesmo conteúdo e fazendo um processo de leitura hipertextual. Cada página é avaliada e os rastreadores, se julgarem pertinente, armazenam informações sobre o que encontraram nesses *sites*. Os rastreadores descobrem a frequência das mudanças dos conteúdos acessados e os revisitam conforme necessário.

4.3.2.2 Indexação

Depois que uma página é rastreada, o *Google* tenta identificar o conteúdo dela, estágio conhecido como indexação. Essa etapa inclui o processamento e a análise do conteúdo textual e das principais *tags* e atributos do conteúdo, como elementos *<title>* (títulos) e atributos alternativos, imagens, vídeos, entre outros (*GOOGLE, 2022d*). Durante o processo de indexação, o *Google* determina se uma página é “canônica”¹⁰⁴ ou uma cópia de outra na *Web*, sendo que a canônica tem maior chance de ser exibida nos resultados da pesquisa. A escolha de páginas canônicas se assemelha muito a noção de “site semente” (*seed pages*), expressão presente na patente de ranqueamento apresentada pelo *Google*, em 2006. Para selecionar a canônica, primeiro os algoritmos agrupam as páginas encontradas que têm conteúdo semelhante (similaridade), processo que se chama *clustering*. Imaginemos que o *Google* enfileira as páginas em estantes organizadas por temas ou áreas de conhecimento, assim como se faz na organização de livros em uma biblioteca física.

Em seguida, o *Google* seleciona a que melhor representa o grupo. As outras páginas do conjunto são versões alternativas que podem ser exibidas em contextos diferentes, como, por exemplo, se o leitor estiver pesquisando uma página muito específica do *cluster*. O *Google* também coleta indicadores sobre a página canônica e o conteúdo dela: o idioma da página, o país em que o conteúdo está localizado, a usabilidade, entre outros. Todos esses itens podem ser usados na próxima etapa, em que são ranqueados e exibidos os resultados da pesquisa. As informações coletadas sobre a página canônica e o *cluster* dela podem ser armazenadas no índice da plataforma. De acordo com o *Google* (2023b), a indexação não é garantida. Já demonstramos anteriormente, ao descrever os dois paradoxos que interferem no funcionamento da *Web* como “propriedade coletiva”, que nem todas as páginas processadas

¹⁰⁴ Uma página canônica originalmente representa um tema de forma mais abrangente e consistente, servindo de modelo para determinado assunto. A definição, formação e fixação de uma página cânone acaba sendo uma sistematização de conjunto de modelos não muito diferente da referência que temos de cânone literário. Uma página canônica é a representativa em um grupo de páginas duplicadas. Por exemplo, se houver dois URLs da mesma página, como “example.com?dress=1234” e “example.com/dresses/1234”, o *Google* escolherá um deles como canônico.

pelo *Google* são indexadas. Fatores como baixa qualidade do conteúdo e o *design* do *site* podem impedir a indexação.

4.3.2.3 Ranqueamento

A etapa de ranqueamento é a que envolve maior dificuldade operacional, portanto, é a grande caixa-preta do *Google*, aquela que especialistas tentam desvendar colhendo migalhas de informações deixadas pela plataforma sobre critérios ou motivos de classificação de uma página. O algoritmo de ranqueamento não é estático. Embora o algoritmo por completo seja secreto, o *Google* nos diz quais coisas são importantes mesmo que não saibamos quão importante. Rieder (2000) afirma que à medida que a complexidade e a opacidade crescem, também aumenta o valor epistêmico e comercial de técnicas que prometem produzir descrições viáveis e decisões eficazes para ordenação da informação social. Por isso, essa etapa tem enorme valor como modelo de negócio para o *Google*. Introna e Nissenbaum (2006) argumentam que os desenvolvimentos no conhecimento técnico e na capacidade de encontrar e indexar *sites* não é correspondido pela capacidade técnica para resolver classificação de relevância, pois os desenvolvedores se deparam com o desafio de aproximar um valor humano complexo (relevância) com um algoritmo de computador. Enquanto parece que o *Google* está dominando a questão da cobertura (o quanto rastreia, o quanto indexa), continua a lutar com a questão do que exatamente extrair do grande volume de possibilidades para uma determinada pesquisa.

Depois de rastrear e indexar as páginas, deixando-as organizadas para o ranqueamento, algoritmos do *Google* começam a calcular ou dar notas ou especificar “pesos” para cada uma dessas páginas e, assim, determinar a posição no *ranking* que vai compor a SERP. São consideradas centenas de sinais contextuais, que podem incluir informações como a localização, o idioma e o dispositivo do leitor (computador ou *smartphone*), o histórico de busca, data de publicação de conteúdos, frequência de publicação, especialidade das fontes que produziram conteúdo, entre outras (*GOOGLE, 2022d*). Rieder (2020) questiona os vieses na utilização até mesmo desses sinais, pois argumenta que a forma de análise desses dados pode ser usada para prever automaticamente e com precisão uma série de atributos pessoais altamente sensíveis, incluindo: orientação sexual, etnia, opiniões religiosas e políticas, traços de personalidade, inteligência, felicidade, uso de substâncias viciantes, separação dos pais, idade e gênero.

Figura 12 – Composição da SERP

The image shows a Google search for 'lego' with the following components:

- Anúncio de Compras (Shopping):** A carousel of sponsored product listings for various LEGO sets, including Minecraft, Speed Champions, and Harry Potter. Each listing shows the product name, price, and the seller (Amazon.com.br).
- Anúncio de Texto (Text Ad):** A sponsored text advertisement for 'Mbrinq LEGO' (Loja Online - Loja certificada LEGO), featuring a red logo and promotional text about the world of LEGO.
- Resultado Orgânico (Organic Result):** A search result for 'legombrinq.com.br' (Mbrinq - Loja Certificada LEGO), including the URL, a small image of a LEGO character, and a brief description of the online store.

Fonte: Elaborado pela autora com base em Departamento de Justiça dos Estados Unidos (2022).

Há outros fatores que ajudam o *Google* a montar o *ranking*, como relevância e qualidade do conteúdo. O peso de cada fator depende da natureza do que foi pesquisado. Por exemplo, a data de publicação do conteúdo é mais importante em consultas sobre notícias recentes do que em definições do dicionário. Apesar de já termos discutido a ideia de relevância no decorrer da pesquisa, vale trazer a explicação mais técnica dada pelo *Google*, ao definir esse fator. Nesse sentido, podemos perceber os critérios da similaridade (correspondência) e popularidade (contabilidade de interações para criação de estimativas) aparecendo novamente:

O sinal mais básico de que as informações são relevantes é quando o conteúdo tem **as mesmas palavras-chave da consulta**. Por exemplo, no caso de páginas da Web, se as palavras-chave aparecem na página ou se aparecem no cabeçalho ou corpo de texto, as informações podem ser mais relevantes. Além de avaliar palavras-chave, nossos sistemas também analisam a relevância do conteúdo para a consulta de outras maneiras. **Usamos dados de interações agregados** e anonimizados para avaliar se os resultados da pesquisa são relevantes para consultas, transformando-os em sinais que ajudam os sistemas de aprendizado de máquina a **estimar melhor a relevância**. Quando você pesquisa "cachorros", por exemplo, provavelmente não quer uma

página com a palavra "cachorros" centenas de vezes. Os algoritmos avaliam se o resultado tem outro tipo de conteúdo relevante além da palavra "cachorros", como imagens de cães, vídeos ou até uma lista de raças. É importante observar que, por mais que nossos sistemas não usem esse tipo de sinal quantificável para avaliar a relevância, eles não são projetados para analisar conceitos subjetivos como o ponto de vista ou a inclinação política de uma página. (GOOGLE, 2023d, grifos nossos).

Sobre o uso de qualidade na classificação, o *Google* (2023d) afirma “[...] priorizar aquilo que parece mais útil”, por isso analisa os parâmetros EEAT (experiência, especialidade, autoridade e confiabilidade), que detalharemos a seguir. Ademais, insistimos em demarcar a presença da popularidade (contabilização de *backlinks*) na forma como o *Google* classifica conteúdos: “Por exemplo, um dos vários fatores que usamos para ajudar a determinar isso é **verificar se outros sites importantes têm links ou referências ao conteúdo**. Isso já provou várias vezes ser um bom sinal de que as informações são confiáveis.” (GOOGLE, 2023d, grifos nossos).

Enfim, classificar páginas da *Web* – e, portanto, os serviços, conteúdos e pontos de vista que elas representam – é um assunto delicado. O dilema que coloca o *Google* no jogo de ser bom para leitores e se sustentar como negócio é justamente manter uma classificação útil. Um elemento que ajuda muito na melhora do processo de classificação é o histórico de pesquisa que retroalimenta o processo de entrega de resultados – como um segundo fluxo de informações na direção inversa. À medida que as pessoas interagem com a página de resultados de pesquisa, suas ações ensinam sobre o mundo. Por exemplo, se, ao buscar por “fruta do conde”, o usuário clica em uma foto, e não em um *link* de texto verbal, ele sinaliza que a imagem foi mais útil como resultado para aquela consulta. O *Google* registra essas ações e, em seguida, algoritmos consideram e aprendem com isso para criar sistemas de avaliação de relevância para páginas. Nesse exemplo que criamos, é como se nas buscas por “fruta do conde” fotos valessem mais que texto verbal.

Rieder, Matamoros-Fernández e Coromina (2018) estudaram o conceito de “cultura do ranqueamento” ou “cultura da classificação” ao olharem para os desdobramentos de processos de hierarquização e modulação da visibilidade feitas pelas plataformas *online*. Segundo os pesquisadores, as dinâmicas que se circunscrevem nos *rankings* convocam leitores/usuários e criadores de conteúdo para uma engrenagem, descrita a seguir:

De um lado, os usuários reagem aos rankings de várias maneiras, clicando, assistindo e – no caso dos criadores de conteúdo ajustando suas estratégias ao que é considerado eficaz em determinado momento. Por outro lado, as técnicas algorítmicas adotam cada vez mais uma perspectiva probabilística e experimental, onde os parâmetros são constantemente testados em relação a um resultado desejado

e adaptados de acordo com o comportamento do usuário. (RIEDER; MATAMOROS-FERNÁNDEZ; COROMINA, 2018, p. 53, tradução nossa)¹⁰⁵.

Para Rieder (2020), o sistema de classificação e criação de listas estimula modos de pensamento hierárquicos e dá para a mente um tipo de alavanca na realidade. As repercussões desse raciocínio classificatório implicam os modos como lemos, escrevemos, construímos e nos relacionamos com o mundo ao nosso redor. A ordenação de conteúdo por plataformas de busca, portanto, raramente se resume a uma simples transposição de um método manual de classificação para a forma computacional. Isso porque é uma ordenação que processa grandes quantidades de sinais distribuídos e com um caráter epistêmico que pouco tem a ver com as práticas de conhecimento que definiram as bibliotecas, enciclopédias ou arquivos nos últimos milênios.

Ao estudar a ferramenta de busca do *YouTube*, plataforma do mesmo grupo do *Google*, Rieder, Matamoros-Fernández e Coromina (2018) observaram indicativos de que a classificação é feita a partir de métricas diretas, que são independentes da consulta, como visualizações, curtidas, tempo de exibição ou assinantes do canal, tudo isso potencialmente recortado em um período de tempo. Para os pesquisadores, todos esses fatores – ligados ao critério da popularidade – são usados para decidir se um vídeo tem mais ou menos relevância do que outros. No entanto, os autores destacam que há outros fatores de maior complexidade que se relacionam com o dinamismo e o uso que se faz das plataformas, tanto no *YouTube* quanto em outras. Configura-se um ciclo de contingenciamento baseado em *feedbacks*: quando as plataformas não especificam mais explicitamente um modelo de decisão para ranqueamento, mas se baseiam no *feedback* do usuário para gerar indutivamente tal modelo, há um problema fundamental de “interpretabilidade” que persiste mesmo quando o acesso total a um sistema é fornecido.

Defendemos que é essa “interpretabilidade” que determina muitas reconfigurações no processo de escrita e edição quando se trata de produtores de texto tentando alcançar visibilidade em ambiente digital, por meio do ranqueamento do *Google*. A técnica de classificação de buscadores é uma realização distribuída e dinâmica que inclui as contribuições de vários atores. Os componentes técnicos da plataforma de busca precisam ser considerados como parte de um contexto mais amplo, a exemplo dos cenários de contingenciamento de conteúdo com técnicas de otimização que alimentam um ciclo movente

¹⁰⁵ Tradução nossa para: “On the one side, users react to rankings in various ways, by clicking, watching and— in the case of content creators adjusting their strategies to what is considered effective at a given moment. On the other side, algorithmic techniques increasingly espouse a probabilistic and experimental outlook where parameters are constantly tested against a desired outcome and adapted according to user behaviour.”

de parâmetro de classificação. Pesquisando sobre o *YouTube*, Rieder, Matamoros-Fernández e Coromina (2018) concluíram que a função de pesquisa é altamente reativa aos ciclos de atenção para cada tema e, veremos a seguir, que a busca do *Google* também é assim.

Os níveis de “interpretabilidade” nas três etapas de busca do *Google* variam. Observando o rastreamento, indexação e ranqueamento, vemos que o processo de coleta de “descoberta de URLs” parece ser dotado de certa objetividade, considerando que a similaridade ou correspondência textual é determinante. Também na classificação de conteúdo, aspectos como localização (GPS ou IP de saída), idioma e dispositivo (computador ou *smartphone*) dos usuários são parâmetros objetivos para entrega de respostas. No entanto, há muita subjetividade no processo de inclusão de uma página no índice da plataforma de busca, porque não há um regramento claro para um conteúdo de alta ou baixa qualidade, fator que impede a indexação e que é responsabilidade de produtores e editores de texto. Mais subjetiva ainda é a medida de relevância usada no ranqueamento.

Análises de documentos do *Google* vão dar sinais sobre essa construção “subjetiva” de qualidade e relevância de conteúdo, mas só a existência desses fatores moventes de qualidade já é suficiente para que nossa crítica à não neutralidade algorítmica se confirme. O *Google* faz questão de transparecer que faz testes rigorosos para a qualidade da busca (da entrega de resultados), usando, além dos fatores mecânicos, a análise de humanos no processo. Em 2021, a plataforma promoveu 700 mil experimentos que resultaram em mais de 4 mil melhorias na busca. Essa testagem acontece com análise de listas de métricas que indica no que as pessoas clicaram, quantas consultas foram feitas, se as consultas foram abandonadas, quanto tempo demorou para clicarem em um resultado, entre outros fatores. Segundo o *Google* (2023c), dados desses experimentos passam pela avaliação de engenheiros experientes e analistas de pesquisas, assim como outros especialistas jurídicos e de privacidade, que determinam se a mudança será aprovada para lançamento.

Nessas testagens, vale destacar a atuação dos chamados “avaliadores de qualidade da pesquisa”, cerca de 10 mil trabalhadores¹⁰⁶ externos que ajudam a medir a qualidade dos resultados da pesquisa constantemente e, de acordo com o *Google*, “[...] são altamente treinados na aplicação das nossas diretrizes” (GOOGLE, 2023c). Eles avaliam, por amostragem, como os conteúdos correspondem à consulta e a qualidade dos resultados com base na especialidade, autoridade e confiabilidade do conteúdo. Não há precisão sobre a quantidade de testagens, e a empresa só informa que são centenas e dezenas de pareceres,

¹⁰⁶ Disponível em: <https://www.youtube.com/watch?v=WbSouimbjtU>. Acesso em: 01 mar. 2023.

portanto, não conseguimos definir proporcionalmente ao total de buscas. Os avaliadores recebem sempre duas versões de uma entrega de busca e vão comparar qual delas apresenta melhores resultados para ajudar a balizar a plataforma. Esses trabalhadores analisam *links* de uma página, pesquisam mais sobre autores de textos, procuram comentários sobre *sites* e outras “provas sociais de reputação”. Conforme o Google (2023c), “[...] devem dar classificações com base em diretrizes comuns, livres de opiniões e crenças pessoais”. Nesse sentido, vale a reflexão de Ochigame (2021a):

Os resultados da pesquisa são produtos não apenas do design do algoritmo, mas também do julgamento humano e do trabalho curatorial. O Google emprega subcontratados trabalhadores, como “avaliadores” que julgam a qualidade percebida dos resultados da pesquisa e “moderadores de conteúdo” que julgam se os resultados parecem inadequados ou ilícito. (OCHIGAME, 2021a, p. 210).

Segundo o Google (2023c), essas avaliações não impactam a classificação diretamente, mas ajudam a comparar a qualidade dos resultados e garantir que atendam a critérios rigorosos no mundo todo. O que há de peculiar na atuação dos avaliadores, além de serem eles trabalhadores precarizados¹⁰⁷, é que, para garantir o que o *Google* (2023c) denomina “abordagem consistente”, os contratados seguem diretrizes de classificadores de qualidade de pesquisa, um guia que orienta o trabalho de avaliação. Ainda conforme a empresa, o guia usado pelos avaliadores é também muito útil para criadores de conteúdo que buscam entender como apreciar seu próprio conteúdo para ter sucesso na pesquisa do *Google*. É nesse documento que encontramos um sem-número de marcadores de subjetividades que sinalizam o *design* político dos algoritmos da plataforma, como descreveremos a seguir. Destacamos o que disse o vice-presidente de pesquisa do *Google*, Ben Gomes, em entrevista à *Consumer News and Business Channel* (CNBC), ao se referir ao guia: “Você pode ver as diretrizes dos avaliadores como onde queremos que o algoritmo de pesquisa vá. Elas não dizem como o algoritmo está classificando os resultados, mas mostram fundamentalmente o que o algoritmo deve fazer.” (tradução nossa)¹⁰⁸.

¹⁰⁷ Avaliadores de pesquisa são pessoas que trabalham com microtarefas (ou microtrabalhos invisíveis) para qualificar as pesquisas que o *Google* está entregando, em especial, ajudando a aprimorar a IA da plataforma de busca. Essas pessoas não são funcionárias da empresa e trabalham via “plataformas parasitas”, sistemas que oferecem valores baixíssimos (menos de um centavo de dólar) para execução de tarefas, como, por exemplo, marcar textos como vitais, úteis, relevantes, ou *spam*. É uma lógica parecida com as fazendas de cliques e envolve “uberização” ou precarização do trabalhador.

¹⁰⁸ Tradução nossa para: “You can view the rater guidelines as where we want the search algorithm to go [...] They don’t tell you *how* the algorithm is ranking results, but they fundamentally show *what* the algorithm should do.” Disponível em: <https://www.cnbc.com/2018/09/17/google-tests-changes-to-its-search-algorithm-how-search-works.html>. Acesso em: 03 mar. 2023.

Isso nos dá indicativos de que os avaliadores de pesquisa trabalham para ensinar algoritmos a compreender parâmetros mais subjetivos, ou seja, a curadoria humana e as subjetividades desses milhares de avaliadores estão se juntando à subjetividade dos desenvolvedores dos algoritmos do *Google*. Preocupa o fato de que os avaliadores de pesquisa não são, necessariamente, especialistas nos assuntos que dão pareceres. Além disso, a fala do vice-presidente nos faz questionar se ao re(definir) os processos de tomada de decisão do motor de busca, a partir de apontamentos subjetivos (de qualidade, autoridade e outros), o *Google* tenta tornar os algoritmos mais “humanizados” ou tenta “matematicalizar” fatores intangíveis para que se tornem passíveis de leitura algorítmica.

4.4 COMO O *GOOGLE* CONSTRÓI “MÉTRICAS” DE RELEVÂNCIA

Ao discutir a produção textual em ambiente digital, podemos dividir a noção de relevância em dois pilares se pensarmos com os olhos de produtores de texto. Para quem escreve, é como se existem duas instâncias: a relevância para públicos, que diz respeito a um texto atender um sistema de expectativas de leitores; e a relevância para os algoritmos, que é as “métricas” criadas pelos motores de busca para escolher conteúdos que vão circular. Ambas as instâncias estão carregadas por subjetividades e vieses, por isso usamos o termo “métricas” entre aspas, considerando que não há parâmetros tão objetivos (apesar de o discurso das *Big Five* tentar sustentar a objetividade). A instância com enfoque em públicos pode esbarrar em questões de linguagem, letramento dos leitores, previsão distorcida do contrato de comunicação que se estabelece no texto, aspectos discursivos de produção de sentido, entre outros fatores. A instância de relevância algorítmica é atravessada pela não neutralidade das plataformas, que defendem seus legados comerciais e modelos de negócios ancorados na propaganda direcionada.

É comum no mercado da produção de conteúdo ouvir especialistas dizerem que não há uma divisão de relevância e que, se um autor de textos escreve bem para seus leitores, certamente estará escrevendo bem para os algoritmos. Discordamos dessa afirmação, porque há práticas incorporadas especificamente para atender à relevância algorítmica, afinal, os algoritmos enxergam uma página de um modo peculiar, o que determina fatores do processo de escrita em ambiente digital.

Algoritmos de plataformas de busca são desenvolvidos para enxergar *HyperText Markup Language* (HTML), que é uma linguagem descritiva que especifica a estrutura de uma página *Web*. Um documento HTML é um conjunto de códigos simples e estruturados

com elementos, que, por sua vez, possuem atributos e marcadores. Produtores e editores de texto, geralmente, trabalham seus processos de criação em Sistemas de Gerenciamento de Conteúdo (CMS) que transformam em interfaces amigáveis todos os elementos HTML. Portanto, os produtores não precisam, necessariamente, compreender códigos HTML, assim como leitores de um *site* que verão a página de conteúdo formatada e legível. Todavia, é importante saber que as diretrizes definidas por plataformas de busca para definir a encontrabilidade de textos vão depender de como os conteúdos estão construídos em HTML. No Capítulo 5, ao descrever categorias de análise, vamos demonstrar como cada categoria textual (palavra-chave, título, *link*) tem um nome específico no HTML. Assim, poderemos perceber como ocorre a leitura algorítmica e como isso se entrelaça com as práticas de escrita.

Por exemplo, os algoritmos do *Google* valorizam no processo de seleção o que consideram como conteúdo principal, ou seja, qualquer parte de uma página que ajuda diretamente essa página a atingir seu objetivo. É sobre esse conteúdo que serão feitas análises de padrões e relevância (*GOOGLE, 2022b*). Os algoritmos enxergam o conteúdo principal a partir do elemento HTML `<main>`, que define o tema central da página ou a funcionalidade dela. Todo o resto é considerado conteúdo suplementar, podendo contribuir para boa experiência do usuário, mas não ajudando diretamente a página a atingir seu objetivo. A Figura 12, é uma reprodução do *site* da Assembleia Legislativa de Minas Gerais (ALMG) e mostra como nós leitores enxergamos a divisão entre conteúdo principal e suplementar. A notícia sobre a mineração na Serra do Curral compõe o conteúdo principal, portanto, é exclusiva dessa página. Os outros elementos se repetem em qualquer página de notícias do *site* da ALMG, pois são suplementares. Os algoritmos do *Google* vão enxergar prioritariamente a notícia sobre a mineração na Serra do Curral, caso ela esteja dentro do elemento `<main>` no HTML.

Figura 13 – Conteúdo principal e suplementar

The image shows a screenshot of the website of the Assembleia Legislativa de Minas Gerais (ALMG). The main content area features a news article titled "ALMG prossegue discussões sobre mineração na Serra do Curral". The article text discusses the legislative process and the impact of mining on the local community. A red arrow points from the text "Conteúdo principal" to the article's main text. To the right, a blue arrow points from the text "Conteúdo suplementar" to a tweet and a video player. The tweet is from the account "Assembleia de Minas" and discusses the legislative process. The video player shows a video titled "Assembleia em Destaque".

Fonte: Elaborado pela autora, 2023.

Para saber mais sobre como algoritmos enxergam as páginas, seguimos destrinchando os documentos do *Google*; coletando diretrizes de análise dos algoritmos e dos avaliadores de pesquisa sobre qualidade e elegibilidade de páginas e conteúdos. Nas seções a seguir, analisamos as que mais influenciam processos de escrita e edição, reforçando que essas diretrizes não são indicadores de classificação. Apresentamos essas diretrizes traduzidas e retextualizadas, mas com fidelidade a alguns termos da documentação da plataforma de busca, em especial, palavras que ascendem discussões a respeito de como se configura o texto plataformizado. Os aspectos a seguir reforçam e complementam os fatores de entrelaçamento entre funcionamento do motor de busca e escrita na *Web*.

4.4.1 Páginas e conteúdos que reflitam o objetivo principal

Um *site* precisa deixar clara sua própria definição, além de publicizar elementos que auxiliem o algoritmo a entender a reputação. Por exemplo, lojas que estampam em suas páginas as avaliações positivas de clientes ou portais de notícias que divulgam os prêmios que seus jornalistas já ganharam podem ser incluídos nessa diretriz. De acordo com o *Google (2022b)*, a reputação também é avaliada a partir da coleta de opiniões e recomendações de “especialistas” externos (de fontes especializadas, sociedades profissionais, discussões em fóruns). Vejamos um exemplo: se existe um fórum de ambientalistas altamente especializados debatendo sobre a proteção da Amazônia e no decorrer das discussões são indicadas leituras de artigos do *site* da Assembleia Legislativa do Estado do Amazonas (Aleam) sobre o tema, provavelmente, algoritmos do *Google* interpretarão a circulação daquele *link* como um voto de confiança, como um ponto de reputação para o *site* da Casa Legislativa.

A subjetividade dessa diretriz está principalmente na definição de “especialistas”, algo extremamente difícil de metrificar. Imaginamos que um algoritmo não vai acionar competências de análise discursiva para identificar um enunciador especialista, a partir de uma construção ethótica desse enunciador¹⁰⁹, para compreendê-lo como capacitado a dar voto de confiança para um conteúdo. O que algoritmos fazem é, por exemplo, contabilizar quantas vezes os *links* da Aleam são citados em fóruns e comparar com outros *links* de conteúdos sobre o tema proteção da Amazônia para definir: site *A* é mais citado que site *B*, logo site *A* tem mais reputação que site *B*. Algoritmos respondem às lógicas, geralmente numéricas e predeterminadas. Nesse caso, entram em ação os avaliadores de pesquisa para sinalizar, de modo mais interpretativo, o que são especialistas, portanto, a definição poderá depender de uma curadoria humana.

Ainda sobre reputação, o *Google (2022b)* afirma que quantidade de tráfego (audiência ou *pageviews*) não é necessariamente uma evidência de reputação, mas sabemos que é importante indicador de popularidade. Se um leitor está buscando o significado da palavra “feminicídio” e pela lógica estatística da popularidade, há um ou dois *sites* mais bem colocados para essa conceituação e acabam sendo muito escolhidos para o *click* dos leitores, a reputação dessas páginas tende a aumentar. O lance estatístico prevalece para que o motor de busca formule uma interpretação dominante para o termo “feminicídio”.

¹⁰⁹ O enunciador detém o “poder” da enunciação e do discurso em si. Se ele consegue se projetar como um especialista, isso se relaciona com o *ethos* e imagem construídos diante de seus públicos. (MAINGUENEAU, 2014).

4.4.2 Boa experiência de uso

Algoritmos seletores de texto valorizam a forma como os conteúdos são criados, organizados e hierarquizados dentro de páginas e *sites*, ademais estão preparados para perceber se uma página facilita a vida do leitor. Para metrificar “boa experiência de uso”, o *Google* (2023e) criou uma espécie de “aconselhamento” para profissionais envolvidos com a produção de texto e desenvolvimento de *sites* denominada *Core Web Vitals*¹¹⁰.

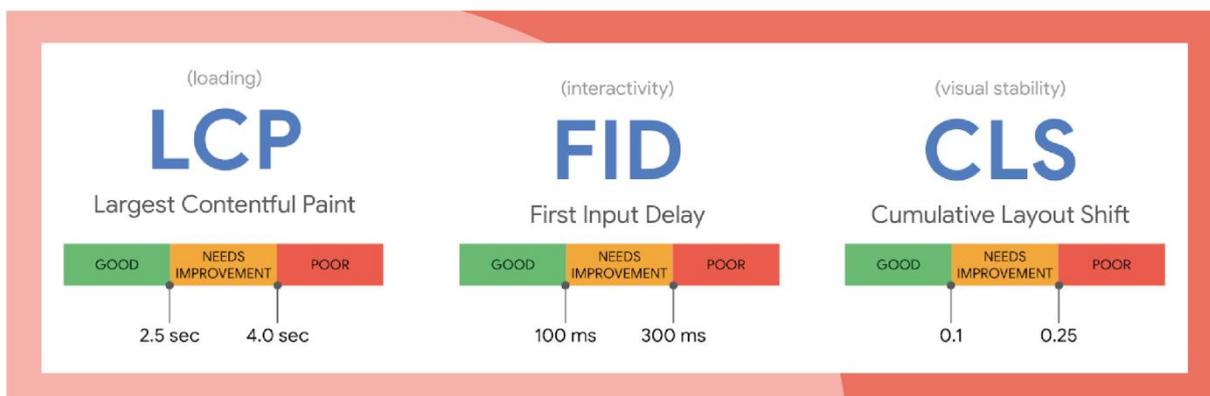
Embora todos os componentes da experiência na página sejam importantes, priorizaremos as páginas com as melhores informações gerais, mesmo que alguns aspectos da experiência sejam inferiores. Uma boa experiência na página não substitui um conteúdo relevante e de qualidade. No entanto, quando há várias páginas com material semelhante, a experiência se torna muito mais importante para a visibilidade na Pesquisa. (GOOGLE, 2023e).

O que parece ser uma iniciativa da plataforma de busca para fornecer orientações sobre os pontos de qualidade, que são considerados essenciais para dizer se uma página oferece ou não boa experiência de uso, é, na verdade, item de governança. Os sinais vitais considerados pelo *Google* são *Largest Contentful Paint* (LCP) – tempo de carregamento; *First Input Delay* (FID)¹¹¹ – interatividade com a página; e *Cumulative Layout Shift* (CLS) – estabilidade visual. A plataforma quantificou esses sinais vitais para que fosse possível parametrizar seus algoritmos, para que identifiquem essas marcas e legitimem uma página, conforme vemos na Figura 14. Mais uma vez, tenta-se objetivar o processo de seleção e elegibilidade de *sites*, criando parâmetros numéricos que ajudam a disfarçar subjetividades.

Figura 14 – Sinais vitais considerados pelo *Google* para metrificar boa experiência de uso

¹¹⁰ Disponível em: <https://blog.chromium.org/2020/05/introducing-web-vitals-essential-metrics.html>. Acesso em: 24 fev. 2023.

¹¹¹ O *Google* fez uma alteração nessa métrica em maio de 2023, passando a chamar de “Net PaiNt” (INP) e medindo o *timing* de resposta de uma ação, mas o que muda efetivamente é o modo como o *Google* mede, que continua sendo de maneira metrificada.



Fonte: Disponível em: <https://blog.chromium.org/>. Acesso em: 24 mar. 2023.

Os documentos do *Google (2022b;2023e)* descrevem que há uma preocupação com os modos como os leitores estão vivenciando a experiência de leitura. Por isso, no escopo da “boa experiência de uso” considerada pela plataforma de busca, está a usabilidade como um método que se presta a desenvolver ambientes favoráveis ao uso por seres humanos (leitura, compreensão, execução de tarefas). Nesse caso, estão incluídas as estratégias de *design*, posicionamento de *links*, transparência do sistema de navegabilidade, clareza na “percepção de clicabilidade”¹¹² (que é a medida em que leitores conseguem compreender onde está um *link* na página).

Outros dois aspectos relacionados à forma dos conteúdos são a legibilidade (*readability*), que diz respeito à facilitação de leitura (aos olhos dos leitores/humanos), e a escaneabilidade (*scannability*), que mistura uma formatação mais confortável para leitura (aos olhos dos leitores/humanos) com a varredura aos “olhos dos algoritmos”¹¹³. Veremos, a seguir, que os dois termos são “irmãos” e isso ficará mais evidente nas análises por categorias no Capítulo 5.

As descrições de legibilidade (*readability*) que aparecem nas documentações do *Google (2022b; 2022c)* são consonantes com os Estudos Linguísticos, que, por sinal, precedem a computação. Não podemos afirmar que, para construir seus algoritmos seletores, os desenvolvedores se basearam tecnicamente nos estudos sobre legibilidade (*readability*). No entanto, os aspectos considerados no *Core Web Vitals* podem ser encontrados em trabalhos da área, a exemplo do livro *É possível facilitar a leitura: um guia para escrever claro*, de Yara Liberato e Lúcia Fulgêncio (2018), que resgata tendências verificadas na leitura natural de textos e experimentos formulados por vários autores. É possível também citar consonâncias

¹¹² Conceito cunhado por Jakob Nielsen (2007).

¹¹³ Sem querer nos aprofundar nas questões de acessibilidade da leitura, que não é nosso foco, ressaltamos que a formatação – legível e escaneável – auxilia aplicativos editores de texto que transformam texto verbo-visual em áudio para pessoas com deficiência.

com os estudos de leitura feitos por Coscarelli (2002), especialmente, quando a autora descreve as preocupações de organização da informação do texto – de acordo com o grau de relevância delas ou conexão entre elas – que facilitam para o leitor construir compreensão. Ou quando Coscarelli (2017) aponta as habilidades fundamentais da leitura em ambiente digital – localizar e avaliar informações, sintetizar e integrar informações, e refletir sobre as informações – que certamente se relacionam com a estrutura de páginas, *sites*, *feeds* e *timelines*.

Legibilidade (*readability*), segundo Liberato e Fulgêncio (2018), denomina o processo que permite ao leitor entender um texto, retirando dele informações e integrando-as em um sistema de conhecimentos preexistentes na memória, de modo a construir uma paisagem mental coerente. Portanto, é um conceito centrado na interação texto e leitor. “O texto é mais legível na medida em que permite ao leitor usar maximamente as estratégias de compreensão de que dispõe” (LIBERATO; FULGÊNCIO, 2018, p. 165). Os algoritmos do *Google* consideram mais indexáveis e ranqueáveis páginas e conteúdos em que a organização do texto se apoia em conhecimentos previamente adquiridos pelo leitor, propiciando maior eficiência da leitura. É por isso que esses algoritmos vão perceber aplicações eficientes de palavras de transição, equilíbrio de uso de voz passiva e voz ativa, adequação de tamanho de frases e parágrafos, além da hierarquização e sinalização clara de tópicos do texto. Aliás, todos esses itens estão citados como elementos de legibilidade por Liberato e Fulgêncio (2018).

Já a escaneabilidade (*scannability*) é uma formatação de textos nas páginas da *Web* que promove um “passar de olhos” do leitor ou um escaneamento visual, que pode atraí-lo e envolvê-lo com o assunto tratado. Além disso, escaneabilidade (*scannability*) tenta garantir que algoritmos dos mecanismos de pesquisa visitem, indexem e classifiquem um texto, afinal, um texto escaneável possui elementos HTML bem demarcados e organizados.

Figura 15 – Escaneabilidade



Fonte: Elaborado pela autora, 2023.

Para tornar um texto mais escaneável, o *Google (2022c)* recomenda aos produtores e editores a publicação de parágrafos curtos, trechos negritados, hierarquização das informações com uma divisão em cabeçalhos denominados no HTML `<h1>`, `<h2>`, `<h3>` (*heading 1, heading 2, heading 3*) e alinhamento do texto à esquerda para promoção do que muitos chamam de leitura em F¹¹⁴. Descrevemos abaixo alguns motivos estritamente técnicos para essas recomendações:

a) o uso de parágrafos curtos e trechos negritados servem para reduzir a “taxa de rejeição” de textos e manter leitores mais na página, aumentando o “tempo de permanência”. Essas são duas métricas consideradas na elegibilidade dos motores de busca. Para produtores de texto que acompanham resultados de performance de conteúdo em ferramentas de análise, como o *Google Analytics*, é possível coletar os valores dessas duas métricas; portanto,

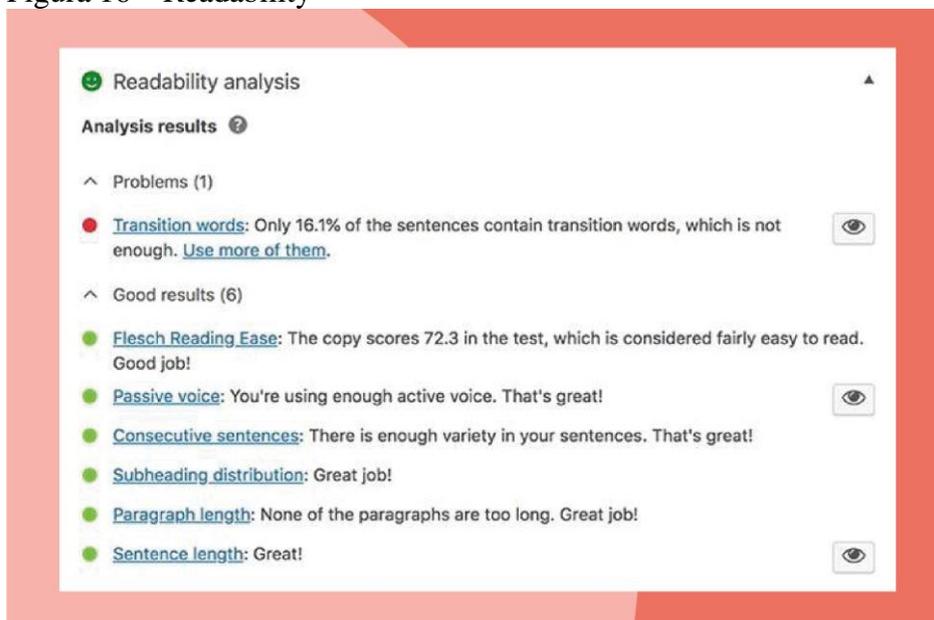
¹¹⁴ Modelo descrito pelo grupo de pesquisadores que colaboram com os estudos de Jakob Nielsen a partir de rastreamento ocular de pessoas lendo as páginas na *Web*. Os estudos mostram que leitores percorrem o título, em seguida descem os olhos pelo canto esquerdo da página, leem os intertítulos e tópicos no meio e, finalmente, miram o canto esquerdo até o final do texto, em um movimento rápido denominado “Padrão F”. Esse movimento é diferente daquele feito na leitura de materiais impressos – o correr dos olhos da esquerda para a direita em um ritmo moderado. Disponível em: <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>. Acesso em: 20 mar. 2023.

testar se determinada formação do texto altera as taxas e, por sua vez, melhora a ranqueabilidade de uma página no mecanismo de pesquisa.

b) Por natureza do meio, os componentes de páginas ou sistemas *Web* são blocos alinhados à esquerda, por isso a linguagem *Cascading Style Sheet (CSS)*, que é usada para estilizar elementos escritos em uma página HTML, tem como alinhamento padrão o “valor esquerda”. Dessa forma, há um consenso de que textos alinhados à esquerda têm maior potencial de ranqueabilidade por motores de busca, só porque seguem uma padronização mais familiar à linguagem de computador, afinal a tecnologia da *Web* não enxerga o alinhamento justificado, que muitas vezes nos seduz por deixar o texto mais próximo da visualidade de produções impressas. Conteúdos em jornais e revistas são justificados, mas fazer o mesmo para textos para *Web* não faz sentido para os algoritmos.

Quando propusemos legibilidade (*readability*) e a escaneabilidade (*scannability*) como parâmetros “irmãos” é porque algumas edições do texto podem servir aos dois. Os *plugins Yoast SEO e Rank Math*, que vamos usar na etapa 2 da pesquisa, denominam de forma única – *readability analysis* – a análise de legibilidade e escaneabilidade feita no *wordpress.com* e que determinará a potencialidade de um texto ser ranqueável. A Figura 15 ilustra a análise do *Yoast SEO* e demonstra que não há separação dos parâmetros. Mesmo assim, entendemos que é necessário compreender a legibilidade (*readability*) e a escaneabilidade (*scannability*) como diferentes, especialmente, porque essa segunda trata também da leitura de código HTML.

Figura 16 – Readability



Fonte: WordPress, 2023.

Analisemos, por exemplo, a hierarquização das informações com uma divisão em cabeçalhos (*heading 1*, *heading 2*, *heading 3*) como um parâmetro que serve como guia estrutural para o algoritmo que vai “varrer” uma página de conteúdo e como fatiador do texto em tópicos/episódios para leitores humanos.

Figura 17 – Hierarquização *heading 1*, *heading 2*, *heading 3*



Fonte: Elaborado pela autora, 2023.

Os trechos marcados como *heading* aparecem sinalizados de modo especial no código HTML¹¹⁵, conduzindo o algoritmo para compreender os assuntos destaque daquele texto. Algoritmos leem prioritariamente URLs, títulos, meta descrições, *headings* e textos negritados. Todos esses textos prioritários possuem marcas especiais no código HTML, portanto, a divisão hierárquica (*heading 1*, *heading 2*, *heading 3*) é um claro sinal de escrita sob demanda algorítmica, mas que pode servir à facilitação da leitura. Nesse caso das *headings*, vale fazer referência aos estudos de legibilidade que vão falar sobre o “fatiamento da leitura”. Liberato e Fulgêncio (2018) citam o fatiamento como um item que realmente favorece a leitura, como mostraram as análises de George Miller (1956)¹¹⁶. Essas pesquisas

¹¹⁵ HTML é linguagem de marcação utilizada na construção de páginas na Web.

¹¹⁶ MILLER, George A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, v. 63, n. 2, p. 81, 1956.

constatarem que a divisão dos textos em sete partes (com pequena margem de mais ou menos duas) aumenta a capacidade de os leitores guardarem a informação completa na memória de longo prazo ou de escolherem guardar apenas uma fatia (*chunk*) na memória de curto prazo.

Liberato e Fulgêncio (2018) descrevem “o ver” como algo episódico, ou seja, uma montagem de fatias linguísticas que compõem significado de forma rápida no processo de leitura. Ou seja, tanto leitores, quanto algoritmos estão construindo fatias do texto de forma mais facilitada a partir dessa fragmentação. Coscarelli (2002) destaca a clareza de tópicos ou assuntos no texto como característica importante para o trabalho do leitor de construir a coerência. Textos cujo tópico ou assunto central não é claro ou que mudam de tópico sem sinalizar para o leitor dificultam a construção da coerência. De acordo com a autora, durante a leitura, um conjunto de informações que podem ajudar na compreensão do texto é ativado na mente do leitor. Se o texto não indica com clareza o seu tópico, serão ativadas muitas informações desnecessárias, e o leitor pode ser levado a fazer inferências erradas e assim não compreender.

Isso poderia explicar a circulação de centenas de textos – tipo *Marketing* de Conteúdo, baseado em metodologia *Inbound Marketing*¹¹⁷ –, como demonstramos em estudo anterior (CRUZ; RIBEIRO, 2016), em que se destacam dois tipos de texto. Existem aqueles em que são elencadas numericamente vantagens, tendências, lições, estratégias, motivos ou dicas. Exemplos: “4 motivos para investir em um curso de francês”, “6 itens para deixar o seu quarto pequeno mais funcional e aconchegante”, “7 dicas e 16 exemplos de cartão de visita de psicologia”. Há outra categoria com apontamentos do que fazer, como fazer, passo a passo, entenda como funciona ou saiba como. Exemplos: “Como decorar o banheiro as crianças”, “Entenda como funciona o curso de formação de professores”, “Ansiedade infantil: entenda como você pode ajudar seu filho”. Cada um desses títulos dará origem a textos fatiados, que parece ser uma apropriação quase perfeita de legibilidade (*readability*) e a escaneabilidade (*scannability*).

Portanto, a diretriz “boa experiência de uso” se ancora na forma como os conteúdos são criados para facilitar a vida do leitor, mas serve ainda como *input* para os algoritmos compreenderem se uma página é elegível. Digamos que o *Google* realmente se preocupe em promover a “saúde dos *sites*” e uma “*Web* melhor”, porém a ideia de “*Web* melhor” já está construída sob o pilar da performance e da otimização que facilite a fluidez da

¹¹⁷ Metodologia focada em transformar leitores em consumidores fidelizados, por meio de conteúdos construídos em conceitos e contextos de marcas específicas. A metodologia é baseada em um funil para a jornada desse leitor que vai da atração, conversão, aproximação (encantamento) até a fidelização.

leitura algorítmica. Assim, o propósito que faz a engrenagem das plataformas de busca girar é um misto de textos legíveis e compreensíveis com textos passíveis de seleção dos robôs, o que torna possível o lance estatístico e todos os pilares da plataformização (datificação, programabilidade e outros).

4.4.3 Utilidade e originalidade

O *Google (2022b)* explicita que um conteúdo é útil quando o público-alvo assim o considera. De acordo com os documentos da plataforma, uma página útil precisa demonstrar a experiência no tema, ter objetivo e foco, além de ensinar para o leitor sobre o assunto que se propõe tratar. Apesar de ser um item carregado de subjetividades, talvez não seja tão difícil para algoritmos e avaliadores de pesquisa identificarem essa utilidade.

O *Google (2022b)* declara que prioriza também a originalidade, apesar de não explicitar em qualquer documento o que compreende como conteúdo “original”. Poderíamos evocar vasta literatura¹¹⁸ para questionar esse parâmetro e apontar a intangibilidade, em especial, no ambiente digital, onde há enorme *continuum* textual e grandes dinâmicas de tradução, bricolagem, transcrição (CAMPOS, 1987), remediação (BOLTER; GRUSIN, 1996), com ou sem o respeito a direitos autorais. Não ficam claros os critérios algoritmos que respondem aos questionamentos, como o original é o que foi publicado primeiro¹¹⁹? O original é o que apresenta estética e sentidos únicos? Um novo texto que cita, faz referência, retextualiza é visto como original pelo algoritmo? A definição de originalidade está completamente nas mãos de avaliadores de pesquisa que também inscrevem vieses nessa decisão?

Por exemplo, se o *site* da Secretaria de Estado de Saúde de Minas Gerais (SES-MG) divulgar um texto, feito pela assessoria de imprensa, sobre sintomas do Sarampo, mas esse conteúdo acabar reproduzido em um grande portal de notícias – aliás, rotina que se repete diariamente na prática jornalística –, não é garantido que o *Google* entregará, prioritariamente, como respostas à pesquisa por sintomas do Sarampo, a página da Secretaria. Mesmo o conteúdo original estando no *site* do órgão público, a reprodução dele em um portal de

¹¹⁸ Como a Teoria da tradução, em que Walter Benjamin (2008) discute o quanto o texto traduzido destrói o que o define como original ou toca fugazmente o sentido do texto original e, em contrapartida, o quanto o original se modifica necessariamente na sua sobrevivência, também poderíamos remeter a Haroldo de Campos (1987; 1997), que aponta o quanto é complexo pensar em originalidade quando se iniciam processos de recriação que fazem nascer novos textos com estéticas baseadas no isomorfismo, conceito que abarca união possível entre as línguas de “chegada” e de “partida”.

¹¹⁹ Publicar primeiro é uma questão extremamente esvaziada no ambiente digital, considerando que qualquer CMS permite alteração de horário de publicação, o que minou, inclusive, a noção de furo jornalístico.

notícias pode ser mais elegível aos olhos do motor e busca. É que a noção de conteúdo original isolada não é tão valiosa se todas as outras diretrizes foram ignoradas. Como acontece nesse relato da jornalista Natália Viana, fundadora da *Agência Pública*, no texto *Eu, uma escrava de robôs*:

Não seguir essas regras têm consequências graves. Por exemplo: descobrimos, em algum momento da nossa trajetória, que nosso site era penalizado pelo Google porque, como nossas reportagens são republicadas por uma rede de outros sites, gratuitamente e sob licença *creative commons*, o robzinho achava que nós estávamos copiando o conteúdo de outro lugar. O robzinho não entende a lógica do compartilhamento e da colaboração sobre a qual construímos a Pública ao longo desses 12 anos. E vejam só: se o algoritmo é burro o suficiente para não entender que a reportagem original é da Agência Pública, e portanto nosso *link* deveria estar lá no topo – bem, a culpa é nossa. O SEO matou a poesia no jornalismo, e não há nenhum remédio a não ser se curvar a ele.

Os *sites* mais comerciais têm hoje mais interesse, condições, disposição de investimento e conhecimento para investir em atender às demandas algorítmicas do que as páginas de órgãos públicos, o que poderia nos levar novamente ao debate do processo de seleção e as implicações disso para as escolhas de textos que circulam ou não. Avaliando essas diretrizes, caminhamos para concluir que, potencialmente, só vão circular textos que consigam cumprir o máximo delas e isso não é isonômico, em medida alguma.

4.4.4 Atualizações e frequência de publicações

Existe uma anedota entre especialistas de otimização e produção de conteúdo que é: “estar sempre de ‘mesa posta’ para a visita do *Google*”. Isso quer dizer ter páginas e conteúdos atualizados e preparados para o rastreamento e a indexação, mesmo que isso exija um esforço de atualização extremamente contingenciado por parte dos profissionais do texto. Aliás, esse contingenciamento foi demonstrado por Camila Gonzaga-Pontes (2012) no estudo do texto jornalístico que é atualizado constantemente durante a cobertura de um fato. Citamos este estudo no Capítulo 2, quando argumentamos a “capacidade de atualização” como novo valor da cadeia editorial do texto. O que o estudo de Gonzaga-Pontes (2012) não enfocou, mas podemos observar agora é esse processo de atualização sendo feito não somente com objetivo de levar ao leitor as novas informações, mas, especificamente, para atender uma demanda algorítmica por “frescor” de páginas.

O *Google* (2022b), por exemplo, afirma que parametriza de forma diferenciadas páginas que se identificam como “*site* de notícia”, imprimindo mais peso na seleção desses

sites quando a informação requerida pelo leitor (*query*) diz respeito a informações recentes, notícias de última hora (como um evento importante ou um desastre natural acontecendo agora), consultas de eventos recorrentes (como eleições ou eventos esportivos), consultas de informações atuais (população de um país ou feriados nacionais). Isso quer dizer que, para pesquisas com esses temas, o *Google* avalia como páginas de maior qualidade para ranqueamento os *sites* de notícias (nacionais, internacionais ou locais) que tenham atualizações sobre o tema.

Há estrategistas de conteúdo, não necessariamente da produção jornalística, que defendem a atualização de conteúdos antigos (de mais de um ano de publicação) como uma forma de melhorar a elegibilidade. Por exemplo, se um *blog* de divulgação científica tem um conteúdo robusto já publicado sobre 11 de fevereiro – Dia das Meninas e Mulheres na Ciência – pode fazer mais sentido que todo ano esse mesmo texto seja atualizado do que uma nova produção sobre a temática. Ou, ainda, a reunião de duas estratégias: a produção de novos conteúdos sobre assunto e a atualização daquele texto robusto, que pode ser considerado o conteúdo âncora da estratégia.

4.4.5 Links rastreáveis, títulos descritivos e úteis

A diretriz dos *links* rastreáveis é uma das que mais remete à ideia de que a plataforma de busca usa potencialmente características intrínsecas ao meio digital para criar lógicas de seleção e classificação, afinal, a noção de “linkabilidade” ou conectividade estava expressa na patente original do *PageRank*, conforme descrevemos anteriormente neste capítulo. Ainda é muito importante que um texto ampare outros textos e que ofereça caminhos multissemióticos ao leitor, em consequência, seja lido pelo algoritmo como um texto âncora, aquele que está recheado de votos de confiança para outros textos ou que recebeu votos de confiança de outros *sites* (GOOGLE, 2022b). Para usar taticamente essa característica da conectividade, muitos profissionais do texto se especializaram em técnicas para melhorar o modo como os algoritmos enxergam a “linkabilidade” dos textos, a exemplo do *link building*, que citamos na seção 2.3.3 e voltaremos descrever como prática incorporada no Capítulo 5.

O *Google* (2022b) incentiva produtores de texto a criarem títulos descritivos tal como as pessoas escrevem na barra de busca, ou seja, o criador do texto precisa aprender a fazer escolhas textuais para atender às tendências da plataforma de busca. Essa diretriz é o que mais explicita a importância do critério de similaridade para o funcionamento da

plataforma de busca, pois se mede a qualidade e elegibilidade de um conteúdo a partir do quanto ele consegue ser similar aos termos-chave digitados na barra de busca do *Google*.

Grandes mercados foram criados no entorno das estratégias de palavras-chave e títulos para conteúdos *online*, especialmente os textos produzidos para *Marketing* de Conteúdo, como já exemplificamos. Citamos novamente aqui as pesquisas de Bueno e Reino (2019) e Christofolletti e Vieira (2015) sobre como a luta por performance de textos interfere na escrita e edição dos títulos e reflete sobre como os jornalistas enxergam essa interposição e seu reflexo no que se entende por jornalismo de qualidade. Observa-se que, para o modelo de escrita jornalístico, a construção de “títulos descritivos e úteis”, uma diretriz documentada pelo *Google* (2022b), está diretamente ligada ao modelo de negócio das empresas de Comunicação, afinal, estar à frente da concorrência (tanto no *ranking* de busca do *Google*, quanto no alcance de diferentes públicos ou na conquista de mais assinantes digitais) significa mais rentabilidade e visibilidade.

4.4.6 Transparência e autoridade

Algoritmos do *Google* buscam compreender quem é responsável por um *site* e quem criou conteúdo (indivíduo, empresa, fundação). As políticas editoriais, descrições de equipes/expediente e informações de contato podem ajudar nessa identificação de autoria, conforme documenta a plataforma (GOOGLE, 2022b). Outros itens que ajudam nesse aspecto são a existência de páginas “sobre nós”, “entre em contato” e *blog* dedicado ao negócio. Autores podem ser notados, mas a instituição responsável por toda página é que será avaliada. Por exemplo, se o conteúdo exibido é pago, ou feito em nome de outra empresa, como no caso de *Branded Content*¹²⁰, o *Google* entenderá a autoridade a partir do *site* que hospeda, não da marca que comprou o conteúdo. Isso porque os algoritmos não parecem preparados para absorver a complexidade da discussão de autoridade, remetendo à construção de autoria como uma prática editorial coletiva e socialmente partilhada (SALGADO, 2007; 2008). Estariam os avaliadores de pesquisa prontos para esse trabalho? Em ambiente digital, especialmente nas cadeias editoriais plataformizadas (que dependem de plataformas *online* para existir), a

¹²⁰ Conteúdo feitos especialmente para marcas em projetos que expressam algum objetivo específico (lançar um produtor, posicionar-se sobre um tema ou vender um serviço etc.). Atualmente, há dezenas de modelos de produção de conteúdos para marcas em grandes empresas de comunicação, como *Uol Content Lab*, *Abril Content Lab*, *Estúdio Folha*, *Estúdio Globo*, entre outros. Nesses casos, as marcas têm conteúdos patrocinados e publicados dentro do âmbito de portais, como *Uol*, *Folha de S.Paulo*, *Globo.com*. O que valerá para o *Google* é a autoria do *site* que publica, ou seja, ele usará como métrica de relevância a autoridade do *site* publicador, e não da marca que patrocina.

autoridade pode ser uma característica muito móvel e extremamente ligada ao que Salgado (2008) denomina vitalidade textual. Os dois exemplos apresentados pelo *Google* nas diretrizes de autoridade em nada parecem considerar a vitalidade dos textos, que depende da partilha dos envolvidos no processo. A seguir, os exemplos do tipo de conteúdo jornalístico (exemplo 1) e de ciência (exemplo 2):

Exemplo 1: Notícias de EEAT alto devem ser produzidos com profissionalismo jornalístico, conter fatos cujo conteúdo é preciso. Fontes de notícias de alto EEAT publicam, geralmente, as políticas editoriais estabelecidas e seus processos de revisão robustos. (GOOGLE, 2022b).

Exemplo 2: As páginas de informações sobre tópicos científicos de EEAT alto devem ser produzidas por pessoas ou organizações com os devidos conhecimentos científicos e que apresentem um consenso científico bem estabelecido sobre questões em que esse consenso exista. (GOOGLE, 2022b).

No primeiro exemplo, salta aos olhos a subjetividade do termo “profissionalismo jornalístico”¹²¹ e nos faz pensar se estão os algoritmos do *Google* e os avaliadores de pesquisa, guiados pelo regramento da plataforma, sensibilizados para metrificar essa classificação. Luiz Gonzaga Motta (2005) tensiona essa complexa definição de “profissionalismo” colocando o fazer jornalístico em duas frentes: a) a que transforma eventos reais em histórias e b) a que articula narrativamente recomposição de acontecimentos, identificação de conflitos, construção de personagens e criação estratégias comunicativas (objetivação para construção dos efeitos de real e subjetivação para construção de efeitos poéticos). Talvez, essas duas frentes sejam subjetivas demais para os algoritmos e para os avaliadores de pesquisa, portanto, é possível que, seguindo o critério da popularidade, o *Google* entenda que a *Folha de S.Paulo* ou o *The New York Times* – jornais tradicionais – produzem mais conteúdos precisos e profissionais do que a *Revista Azmina* ou os *Jornalistas Livres* – exemplos de veículos com viés menos tradicional. Também poderia entender como profissional o conteúdo da *Jovem Pan*, que tem políticas editoriais tradicionalmente estabelecidas no ambiente digital e fora dele, mas que foi apontado como *hub* de *fake news* e conteúdos antivacina durante a pandemia da covid-19 (FONSECA, 2022).

No segundo exemplo, destacamos a noção de “consenso científico bem estabelecido” em tempos de uma produção de conteúdo em que interesses diversos tensionam a noção de ciência e consenso, em especial, em temáticas trazidas ao centro de controvérsias.

¹²¹ Vale demarcar que este é um ponto sem consenso do PL 2.630/2020, que trata de regulação e sugere a remuneração por direitos autorais de conteúdo jornalístico compartilhado pelas plataformas *online*. Críticos apontam para possível assimetria regulatória no que diz respeito ao pagamento de conteúdos de veículos tradicionais *versus* veículos não tradicionais.

Podemos exemplificar o assunto “aquecimento global”, a partir das discussões de Silva (2017), que mostraram o quanto as corporações do setor de energia dos Estados Unidos financiaram, por três décadas, campanha de relações públicas com intuito de influenciar a opinião pública e produzir controvérsias sobre o aquecimento global. Elas faziam isso, inclusive, com produção e circulação de conteúdo em campanhas voltadas para a criação de dúvidas com práticas de influência centradas na manufatura de incertezas. Não sabemos se os algoritmos e os avaliadores de pesquisa do *Google* estão sensibilizados para cercar as controvérsias ou se estão eles mesmos fomentando as controvérsias.

4.4.7 Qualidade e quantidade de conteúdo

Não sabemos exatamente como os algoritmos do *Google* podem metrificar qualidade, mas as documentações indicam que os algoritmos estão medindo qualidade e contando com a ajuda do “gosto humano” dos avaliadores de pesquisa (GOOGLE, 2022b). O *Google* (2022b) orienta que conteúdos sejam produzidos a partir de informações, relatos, pesquisas ou análises originais. Os textos devem apresentar as informações de maneira confiável, com indicação clara da fonte, evidências do uso de conhecimento especializado e dados sobre o autor ou *site* da publicação. De acordo com o *Google* (2022b), espera-se que seja um conteúdo plenamente publicável em uma revista, enciclopédia ou livro impresso, o que parece muito irônico e nos faz retomar Ribeiro (2018a), ao tensionar cultura digital, cultura impressa e cultura escrita, relativizando rupturas ou contraposições: no caso do *Google*, conteúdos impressos servem como referência de métrica para relevância digital.

Para o *Google* (2022b), conteúdo de qualidade é aquele que baseado em outras fontes, é reescrito adicionando valor e originalidade; não usa de termos exagerados ou chocantes no título principal ou da página; é confiável ou reconhecido como autoridade no assunto; é escrito por especialistas ou entusiastas que comprovadamente conhecem bem o tema; não apresenta erros factuais nem problemas de ortografia ou estilo. A seguir, apresentamos dois exemplos do que são consideradas diretrizes de qualidade para conteúdos do tipo jornalístico (exemplo 1) e de ciência (exemplo 2):

Exemplo 1: Alta qualidade: um relatório original que fornece informações que, de outra forma, não seria conhecido se o artigo não o tivesse revelado. Os relatórios originais, aprofundados e investigativos exigem um alto grau de habilidade, tempo e esforço. (GOOGLE, 2022b).

Exemplo 2: Alta qualidade: é original, preciso, abrangente, comunicado de forma clara, profissionalmente apresentado e deve refletir o consenso de especialistas. (GOOGLE, 2002b).

No primeiro exemplo, além da questão da originalidade, que já destacamos, vale discutir o nível de subjetividade no “grau de habilidade, tempo e esforço” (GOOGLE, 2022b) gasto por um produtor de texto e analisado por algoritmos e avaliadores de pesquisa. Que métrica é essa senão uma diretriz que força a entrada de parâmetros humanizados e subjetivos no processo de seleção e classificação de textos, mas ao mesmo tempo compõe uma entrega travestida de objetividade ao leitor? No segundo exemplo, o que se destaca é a ideia de “especialista”, que já discutimos anteriormente – acionando teorias da análise discursiva – para abordar quem é um enunciador capaz de se construir (ou construir um *ethos*) como uma voz especializada.

Algoritmos e avaliadores de pesquisa vão compreender a quantidade como um item variável a depender do assunto e do objetivo da página. No entanto, a diretriz geral considera que, para todos os tipos de páginas da *Web*, a criação de conteúdo precisa demonstrar, refletida na quantidade de texto, pelo menos, um dos seguintes itens: “tempo, esforço, experiência e talento/habilidade”(GOOGLE, 2022b). Considerando a extrema subjetividade da descrição anterior, o tamanho do texto é um ponto muito controverso entre especialistas de otimização e produtores de texto, porque muitos entendem que este não é um parâmetro fundamental na seleção e classificação de textos, enquanto outros acreditam que conteúdos com menos de 300 palavras não serão considerados por algoritmos, como mostraremos no Capítulo 5. A questão é que o texto deve ter o tamanho que ele merece e a escrita também precisará atender as expectativas de leitores, o que nos faz evocar, novamente, os estudos sobre legibilidade, de Liberato e Fulgêncio (2018), e sobre leitura, de Coscarelli (2002; 2017).

4.4.8 Fatores EEAT

Os parâmetros Experiência, Expertise, Autoridade e Confiabilidade (EEAT) aparecem reiteradamente na documentação do *Google* como fundamentais, e observamos que são muito interpretativos. Ao se referir à experiência, o *Google* considera páginas e conteúdos que consigam demonstrar a prática sobre determinado assunto, seja ele futebol, turismo, economia ou moda. Quando a informação é construída por alguém que já experienciou (uso real de um produto ou visita a um ponto turístico), algoritmos e avaliadores de pesquisa estão

preparados para considerar esse diferencial, por isso, atualmente, fóruns na *Web* vem sendo bastante considerados pelos motores de busca.

Já o pilar da expertise se relaciona com especialização, considerando se há indicação clara da fonte, evidências do uso de conhecimento especializado pelo autor, conteúdo sem erros factuais e que podem ser facilmente verificados. A diferença quanto ao primeiro pilar é que uma pessoa especializada pode não ter passado pela experiência da qual descreve, a exemplo de um médico obstetra que é capaz de prescrever e orientar um parto, mas traz informações diferentes de uma médica obstetra puérpera que poderá dar acréscimo de duas informações com um relato pessoal.

Em maio de 2023, ao lançar o filtro *Perspectivas*¹²², o *Google* anunciou atualizações que visam valorizar “[...] experiências humanas que tornam a Web e o mundo tão maravilhosos”:

Também trabalhamos para melhorar a forma como classificamos o conteúdo das avaliações na Pesquisa – por exemplo, páginas da Web que avaliam empresas ou destinos – para dar mais ênfase à qualidade e originalidade das informações. Agora você verá mais páginas baseadas em experiência em primeira mão ou criadas por alguém com profundo conhecimento em um determinado assunto. E como enfatizamos a importância da “experiência” como um elemento de conteúdo útil, continuamos nosso foco na qualidade da informação e em atributos críticos como autoridade, experiência e confiabilidade, para que você possa confiar nas informações que encontrar. (GOOGLE, 2023e).

Esse filtro é um desdobramento da *Search Generative Experience* (SGE), um dos projetos de IA do *Google*, que encontra informações úteis de pessoas em fóruns e *sites* de mídia social. Estão no radar *Reddit*, *YouTube*, *blogs* pessoais e outros *sites* que oferecem aos leitores “respostas de humanos reais”. O investimento nisso representa uma busca orgânica que pretende satisfazer o público sem a necessidade de clicar em *links*. Alguns jornalistas avaliaram essa novidade como ruim para o setor, considerando uma desvalorização do conteúdo jornalístico em detrimento de relatos, o que geraria queda de tráfego para os portais de notícias. Uma estratégia parecida foi usada pelo *Facebook* em 2017, com uma atualização algorítmica que culminou com a saída da *Folha de S.Paulo* da plataforma de rede social, como relatamos no Capítulo 5.

Já o quesito autoridade é sobre o reconhecimento que uma página ou conteúdo podem receber pelo que publicam, ou seja, um conteúdo pode ser considerado de maior autoridade por ser reconhecido na comunidade ou dentro de um nicho. Se buscarmos no

¹²² Disponível em: <https://blog.google/products/search/google-search-perspectives/>. Acesso em: 12 jul. 2023.

Google pelo Código Brasileiro de Trânsito (CTB), certamente, nas primeiras posições estará um *site* do governo, porque é oficial e contém a legislação original, todavia, *sites* especializados em conteúdo automobilístico poderão produzir textos especializados sobre o tema em busca de autoridade.

Por fim, para avaliar confiabilidade, o *Google* considera até que ponto uma página é “[...] precisa, honesta, segura e confiável” (GOOGLE, 2022b), o que parece bem subjetivo. Segundo o *Google*, a confiança é o membro mais importante da família EEAT e depende muito dos pilares anteriores. O próprio *Google* exemplifica: “As análises de produtos devem ser honestas e escritas para ajudar outras pessoas a tomar decisões de compra informadas (em vez de apenas vender o produto)” (GOOGLE, 2022b); ou ainda “As páginas informativas sobre tópicos *YMYL*¹²³ claros devem ser precisas para evitar danos às pessoas e à sociedade” (GOOGLE, 2022b). O termo *Your Money Your Life* (YMYL), que aparece aqui, refere-se a determinados tópicos que poderiam interferir na saúde financeira, física e mental dos leitores. Para o *Google*, páginas e conteúdos que tangenciam esses temas são mais sensíveis para o usuário e, como os algoritmos estão preparados para identificar os propósitos das páginas, vão designar aquelas que têm capacidade de afetar o usuário em suas decisões no dia a dia. Portanto, *sites* que tangenciam *YMYL* precisam garantir ainda mais os pilares de EEAT.

Um exemplo simbólico sobre a questão da confiabilidade é o *Trust Project*¹²⁴, financiado pelo *Google*, *Bing* e *Facebook/Meta*, que ouviu jornalistas, veículos e usuários de notícias nos Estados Unidos e na Europa, e desenvolveu uma lista de oito indicadores de credibilidade (termo usado na tradução do projeto em português¹²⁵). Vejamos os indicadores: 1) melhores práticas (quem financia o veículo, proprietários, sócios, patrocinadores); 2) jornalista (detalhes sobre quem produziu a matéria, incluindo expertise e outras matérias em que atuou); 3) tipo de matéria (etiquetas para distinguir opinião, análise e publicidade/conteúdo patrocinado); 4) citações e referências (para matérias investigativas ou em profundidade, maior acesso às fontes por trás dos fatos e afirmações); 5) métodos (informações sobre como os repórteres decidiram apurar a matéria e o que fizeram durante o processo); 6) apuração local (permitir que as pessoas saibam quando a matéria tem origem ou expertise local); 7) diversidade de vozes (os esforços da redação para trazer perspectivas diversas para a reportagem); 8) *feedback* acionável (estimular a colaboração do público, estabelecendo prioridades para a cobertura, contribuindo para o processo de reportagem e

¹²³ O termo *Your Money Your Life* (YMYL) refere-se a determinados tópicos que poderiam interferir na saúde financeira, física e mental dos leitores.

¹²⁴ Disponível em: <https://thetrustproject.org/>. Acesso em: 08 jun. 2023.

¹²⁵ Disponível em: <https://www.manualdacredibilidade.com.br/>. Acesso em: 08 jun. 2023.

garantindo a precisão). A partir desses oito pontos, foram criados protótipos de interfaces e programas de código aberto¹²⁶ para transmitir os indicadores no ambiente digital e torná-los identificáveis, inclusive, a algoritmos de plataformas *online*. É mais uma tentativa de sistematizar diretrizes que ajudam como indicativos para um critério carregado de subjetividades, como o caso da confiabilidade.

A implicação dos parâmetros EEAT nas práticas de escrita é a busca dos produtores e editores de textos pela mostração aos algoritmos e avaliadores de pesquisa para garantir o reconhecimento de suas páginas e conteúdos. Citamos no Capítulo 2 que o *lobby* por *links* fez emergir a necessidade do trabalho de *Digital PRs* ou Relações Públicas Digitais, mas há outras dinâmicas que acontecem em busca de reconhecimento, como esforço para que *sites* e conteúdos recebam boas avaliações de clientes ou leitores, menções na *Wikipédia* – considerada pelo *Google* como *site* confiável – e menções em fóruns. Isso reforça a ideia de que, para os algoritmos de plataformas de busca, tão importante quanto a produção do texto, é o modo como ele circula, é citado e compartilhado.

Os pilares EEAT são uma estratégia de governança do *Google* para publicizar o modelo de “como classificamos sites” sem afirmar que esses parâmetros realmente influenciam na classificação. Desse modo, as diretrizes EEAT integram o modelo de métrica de relevância que descrevemos nesta seção, porque servem como item de legitimação da plataforma de busca. Esse tipo de legitimidade para páginas e conteúdos é muito diferente daquele que se baseia na experiência subjetiva de um editor respeitado por sua experiência. Gillespie (2016) pontua que cada sistema da indústria da informática faz seu jogo de legitimidade e implica sua própria estrutura para o que é legitimidade (quantificação ou interpretação, distância mecânica ou proximidade humana).

Parece ser interessante para o *Google* que profissionais do texto e desenvolvedores *Web* estejam envolvidos em alcançar legitimidade para seus conteúdos, porque isso influencia diretamente na qualidade das páginas a serem indexadas. É mais vantajoso ainda, para a plataforma de busca, ter um índice enxuto, com conteúdos de excelência, que melhoram e facilitam o processo de ranqueamento e entrega, afinal, é dessa excelência que vive um motor de busca. Sendo assim, os profissionais do texto e desenvolvedores *Web* trabalham para si, para as empresas empregadoras e, ainda, para o *Google*, criando o que a companhia denomina “sinais vitais” ou “saúde da *Web*”. Concluimos

¹²⁶ Os protocolos estão disponíveis somente para veículos parceiros, portanto, não é possível acessá-los, mas o desenho básico, que está *online*, mostra como os indicadores seriam organizados como atributos numa HTML.

que as práticas de governança estão diretamente relacionadas às práticas técnico-comerciais das plataformas *online*.

Cabe ressaltar que o *Google (2022b)* considera como conteúdo de baixa qualidade aqueles que não cumprem as diretrizes descritas até aqui, em especial um texto criado sem tempo, esforço, experiência ou talento (habilidade) adequados, ou que violem as regras da plataforma: títulos exagerados ou chocantes; páginas que espalham ódio (conteúdos com violência contra grupos ou pessoas), que enganam os leitores (conteúdo de golpes maliciosos, incentivam danos a outras pessoas), que potencialmente desinformam usuários (conteúdo impreciso, contradiz consenso, teorias da conspiração e desmentidas); páginas sem propósito (sem conteúdo principal ou sem sentido); e páginas que não atingem seu objetivo (conteúdo copiado ou gerado automaticamente, páginas desconfiguradas ou com *spam*).

De acordo com o *Google (2022b)*, as páginas e *sites* compostos por conteúdo gerado automaticamente, sem edição ou curadoria manual, e nenhum conteúdo ou valor agregado original para os usuários são ainda considerados de pior qualidade. Irônico dessa diretriz é o fato de que muitos produtores de texto reclamam de transformarem suas produções em menos criativa, menos originais e quase automatizadas ao tentar cumprir os parâmetros de ranqueamento do *Google*, muitas vezes “forjados” em técnicas de otimização. Outra ironia é supor que o *Google* não vai ranquear os textos produzidos por assistentes de texto que funcionam a partir de IA (como *ChatGPT*, *Flair*¹²⁷, *Ink*¹²⁸ e *ThunderContent*¹²⁹, *Copy.AI*¹³⁰, *Frase*¹³¹, *Jasper*¹³²) ou por outros tipos de robôs, como nos casos do Jornalismo Automatizado, cartografados por Silvia Dalben (2018), em que *softwares* foram usados para gerar textos sobre homicídios, desempenhos de escolas, coberturas de eleições e jogos.

4.4.9 Mito da relevância

Destrinchar essas diretrizes textuais, como fizemos nesta seção, é demonstrar o quanto o regramento da plataforma de busca está, na prática, entrelaçado aos modos de entendimento do que é texto em ambiente digital e, ainda, ao entendimento do que é escrever nesse meio para alcançar visibilidade. Ademais, retomando o argumento inicial, a ideia de “métrica” (entre aspas) reflete uma narrativa simbólica documentada pelo *Google*, em todos

¹²⁷ Disponível em: <https://flair.ai/>. Acesso em: 28 fev. 2023.

¹²⁸ Disponível em: <https://inkforall.com/>. Acesso em: 28 fev. 2023.

¹²⁹ Disponível em: <https://thundercontent.com/>. Acesso em: 28 fev. 2023.

¹³⁰ Disponível em: <https://www.copy.ai/>. Acesso em: 28 fev. 2023.

¹³¹ Disponível em: <https://www.frase.io/>. Acesso em: 28 fev. 2023.

¹³² Disponível em: <https://www.jasper.ai/>. Acesso em: 28 fev. 2023.

os documentos analisados no corpus da etapa 1, para criar o mito da relevância, que envolve eloquentemente produtores de texto e desenvolvedores *Web*. A divulgação de documentos, guias, fóruns, *updates* em algoritmos e patentes serve para a manutenção do mito, que, segundo definiu Marilena Chauí (2013), é uma narrativa confiável e verdadeira que se ocupa de apresentar respostas para quem vive e convive com o assunto.

O mito da relevância trata de acomodar os profissionais do texto ao mundo da escrita em ambiente digital, configurando-se como uma estratégia aos negócios do *Google*. Todas as normativas que parecem apoiar a prática de produção de textos não têm a obrigação de explicar. A função do mito da relevância não é tornar inteligível, mas, sim, albergar profissionais do texto em uma realidade técnica e algorítmica (encaixapretada) que parece muito assustadora e que é, na verdade, dotada de subjetividades escancaradas. O mito da relevância tem vida própria porque interfere em práticas de escrita das mais diversas, e os profissionais do texto nem precisam, necessariamente, conhecer os documentos do *Google* que normatizam a relevância. O mito da relevância dá respostas, mesmo que incorpóreas e tendenciosas, a esses profissionais, assim como a mitologia grega afastou os medos e deixou o povo mais confortável diante de mudanças críticas na sociedade da época. Portanto, o mito da relevância também desempenha uma função social, fixando ritos e modelando a produção textual. Por isso, um dos papéis da nossa pesquisa é desmistificar.

4.5 MEDIAÇÃO E CURADORIA

As mediações são momentos possíveis de compreender as interações entre o espaço da produção e o espaço da recepção. Mais especificamente, ao pensar sobre mediações das plataformas de busca, estamos discutindo as interações entre espaços de escrita e espaços de leitura. No ambiente digital, as plataformas *online* são estruturas fundamentais no jogo de mediações (WINQUES; LONGHI, 2022) e complexificam qualquer tentativa de recuperar estudos de recepção do século XX¹³³, que, aliás, com a devida relativização, servem como bom instrumento de apoio à reflexão sobre pesquisas em plataformas.

Já vimos anteriormente que, no caso do *Google*, é difícil dizer que as pessoas percebem e dão sentido à mediação algorítmica, dada a natureza oculta e invisível durante o uso da plataforma. Afinal, como as pessoas experimentam e entendem suas interações com os algoritmos em sua vida cotidiana? Taina Bucher (2017) cria a expressão “imaginários

¹³³ Estudos apresentados por pesquisadores como Guillermo Orozco Gómez, Stuart Hall e Jesús Martín-Barbero, que muito contribuem para o que pensamos hoje sobre recepção.

algorítmicos” para dar conta do entendimento de produção e consumo em plataformas *online* e conclui que o modo como as pessoas (leitores, usuários) experimentam (buscam, leem, assistem) suas interações com o mundo atravessado por algoritmos torna a noção de recepção mais cíclica e multidimensional. Isso ocorre, principalmente, pela materialidade da mediação feita pelas plataformas *online*. Essa materialidade pode ser percebida, por exemplo, na análise de patentes e documentos de governança que registram esses sistemas. No caso do *Google*, o arcabouço documental, analisado na etapa 1 desta pesquisa, formaliza as possíveis modalidades de atuação dos algorítmicos no tripé “produção, circulação e leitura”, como mostramos na seção anterior.

Furtado (2003) argumenta que vivemos um paradigma em relação à mediação. Por um lado, há certa transparência ou ausência de mediação quando um dos *media* consegue se apagar, parecendo deixar os leitores (espectadores, consumidores) na presença direta com os objetos de leitura e consumo. Por outro, existe uma clara hipermediação ligada ao fato de que todo conhecimento do mundo chega até nós pelos *media* e, nesse último caso, há uma consciência das pessoas de que estão na presença de um *media* e as experiências vividas naquele espaço parecem ser experiências do real, como se ali não houvesse representação.

Acreditamos que seja preciso esquecer qualquer visão etérea e não encarnada da mediação, pois ela está materializada. É por isso também que a noção de mediação não pode negligenciar o “raciocínio mecânico” (RIEDER, 2020) colocado pelos algoritmos. Por exemplo, a recuperação e a ordenação da informação, como serviços do *Google*, desenvolveram-se em estreita relação com a estatística, conforme já analisamos. E, por causa da dimensão exponencial, do lance estatístico, a relação produção e consumo (escrita e leitura) depende da filtragem/ordenação/curadoria como ferramentas para gerenciar a conectividade e a interação em massa (RIEDER, 2020).

A curadoria algorítmica opera em plataformas *online* mesmo na ausência de regulamentação estatal, porque faz parte da própria arquitetura de *software*. Portanto, acaba sendo inevitável e, muitas vezes, invisível, como no caso dos motores de busca em que a dificuldade de enxergar a filtragem é maior do que nos *feeds* de plataformas de redes sociais. Sendo inevitável, a curadoria se torna uma fantasia do fluxo livre de informações. Para Ochigame e Holston (2016), as operações contínuas de filtragem algorítmica para organizar a visibilidade de conteúdos exigem uma nova compreensão do controle da informação como regra, e não como exceção. Na grande maioria dos casos, as plataformas não informam os usuários sobre as lógicas de filtragem que empregam – e muito menos lhes oferecem controle sobre esses filtros. As plataformas *online* apenas o fazem e que fique claro, a filtragem

algorítmica não promove ou suprime as textualidades em si, mas sua capacidade de atingir as audiências, pois a curadoria trabalha entre os pilares da circulação e leitura. É por isso que, ao serem acusados de mediação social enviesada, as empresas responsáveis pelas plataformas dizem que garantem liberdade de expressão nos seguintes moldes: “basta ter um *login* e qualquer pessoa pode produzir” ou “ao fazer uma busca é o leitor quem escolhe em qual *link* clicar”. O problema é que a curadoria algorítmica impõe restrições sobre o que se pode ouvir, a que audiência se pode alcançar e como a fala circula, como argumentam Ochigame e Holston (2016):

Se a questão crucial não é se alguém é livre para falar, mas como o discurso circula, a contestação do atual regime de controle da informação requer uma política de audibilidade mais do que de liberdade de expressão. Este último é necessário, mas insuficiente para enfrentar a regulação automatizada e implícita do que as vozes são ouvidas. A política da audibilidade deve revelar e questionar as normas que estruturam os filtros algorítmicos. (OCHIGAME; HOLSTON, 2016, p. 107, tradução nossa)¹³⁴.

Como a curadoria parece ser irremediável, é necessário ir além das exigências impossíveis de neutralidade e imparcialidade e, em vez disso, exigir estratégias que subvertam ou usem taticamente os filtros existentes para criar lógicas algorítmicas alternativas inspiradas em diferentes normas políticas (OCHIGAME; HOLSTON, 2016). Em se tratando de plataformas de busca, o próprio Rodrigo Ochigame (2021a) sugere ferramentas que privilegiem a pluralidade e a crítica, combinando abordagens curatoriais humanas e algorítmicas, como mostraremos no Capítulo 5, ao apresentar iniciativas subversivas. E aqui não estamos falando em incluir avaliadores de pesquisa realizando microtrabalho e seguindo um guia de governança, como fez o *Google*. A ideia é recriar a lógica, fugindo dos filtros baseados em popularidade, por exemplo, e inventando sistemas que elucidem melhor significados e nuances de engajamento crítico dos conteúdos. A inclusão da curadoria humana aos procedimentos matemáticos é capaz de agregar novas e inusitadas perspectivas à informação, oferecendo aos seus usuários a surpresa, o inesperado ou simplesmente aquilo que o usuário nem imaginaria existir no mundo e sobre o mundo, ampliando seu próprio entendimento de mundo (SAAD CORREA; BERTOCCHI, 2012).

¹³⁴ Tradução nossa para: “If the crucial issue is not whether one is free to speak but how speech circulates, the contestation of the current regime of information control requires a politics of audibility more than of free speech. The latter is necessary but insufficient to confront the automated and implicit regulation of what voices get heard. The politics of audibility must reveal and question the norms that structure algorithmic filters.”

4.6 MOVIMENTOS DE BUSCA

O jogo das mediações em plataformas de busca é feito também pelos leitores que misturam seus hábitos e engajam (ou não) com conteúdos, possibilitando que os algoritmos existam em suas performances, pois eles se constituem, principalmente, quando entram em “negociação” com os usuários. “Ler envolve uma situação comunicativa na qual há um autor, um leitor, um texto (ou múltiplos textos), um contexto e um ou mais objetivos para o ato da leitura.” (COSCARELLI, 2016, p. 69). Nesse contexto estão as materialidades das plataformas, agindo claramente em cada ato e em cada resultado de leitura. Por isso, é preciso saber minimamente como se dá o movimento de leitores, mesmo sem a intenção de promover estudo de recepção ou de sair do escopo desta pesquisa, que tem foco nas práticas de escrita e edição. Até aqui, ao argumentar o *Google* como plataforma infraestrutural e descrever como o funcionamento do motor de busca entrelaça processos de escrita em ambiente digital, demos visibilidade à materialidade da plataforma de busca mediando processos de “produção, circulação e leitura”. Na argumentação anterior, trouxemos a importância dos rastros deixados por leitores para o funcionamento da engrenagem do motor de busca. Agora, vamos nos dedicar, brevemente, a compreender como o *Google* monitora os movimentos de busca e como produtores de texto incorporam os rastros deixados por leitores como parte determinante para a escrita.

Nossa referência para o termo “leitor” está ancorada em estudos anteriores que se dedicaram a caracterizar e nomear o sujeito cooperador dos processos que integram a cena produtiva midiática, nos mais variados formatos e gêneros, conforme Winques e Longhi (2022) já resumiram:

A partir da ampliação dos canais de comunicação na internet, em uma tentativa de definir o sujeito conectado, surgiram diversos conceitos: prosumidor (Castells, 2003); gatewatchers (Bruns, 2003); interagente (Primo, 2007); internauta (García Canclini, 2008); leitor-produtor (Brignol, 2010) etc. Todos esses termos, em geral, apontam para o usuário/cidadão como agente participativo no processo produtivo nos mais variados níveis da construção de narrativas e informações em rede. (WINQUES; LONGHI, 2022, p. 155).

Para além dos termos que definem esse leitor mais ativo ou leitor imersivo (SANTAELLA, 2004), capaz de construir seus resultados de leitura de forma participativa, a interação dele com as plataformas de busca é ainda mais definidora. Santaella (2004) constata que observar, absorver, entender, reconhecer, buscar, escolher, elaborar, agir ocorrem em simultaneidade, ademais a complementariedade dos circuitos cibernéticos e circuitos mentais

deve produzir reorganizações das redes neurais desse leitor. Para a autora, a grande marca identificatória do leitor imersivo é a interatividade, que é onde se situam as mensagens, pois estas se formam no trânsito informacional. O buscador do *Google* pode servir a uma variedade de usos de leitura por causa das propriedades constitutivas da plataforma (*affordances*), que determinam as possibilidades de ação das pessoas e quais padrões de interação social serão dominantes (FURTADO, 2003).

Sugerimos, a seguir, olhar para quatro movimentos no processo de busca que podem constituir as leituras em interação com mecanismos de pesquisa. Nos inspiramos nos movimentos descritos por Vandendorpe (2007): seleção, associação, continuidade e estratificação. O pesquisador traça possibilidades de acesso a rotas hipertextuais a partir de como leitores escolhem elementos que desejam consultar, navegam por blocos de informação deixando-se guiar por associações de ideias, acessam o texto em fragmentação topográfica ou contiguidade associativa, ou tabulam os vários extratos de texto. Coscarelli (2017) dialoga muito com essa perspectiva, quando aponta as habilidades fundamentais da leitura em ambiente digital – localizar e avaliar informações, sintetizar e integrar informações, e refletir sobre as informações. Essas ações parecem refletir, de modo amplo, os movimentos de uso da plataforma *Google* que veremos a seguir, de forma mais localizada.

4.6.1 Intenções de busca

Nesse primeiro movimento, vamos considerar todos os atos e pensamentos que precedem a leitura-escritura, ou seja, o momento em que o leitor está decidindo sua necessidade de pesquisa e se preparando para externalizar as operações da mente (SANTAELLA, 2005). Esses propósitos dos leitores ao acessar a plataforma de busca são chamados pelo *Google* (2022b) de “intenções de pesquisa” (*user intent, search intent ou keyword intent*). Apesar de a empresa ter criado algumas tipologias, descritas no Quadro 8, a intenção de pesquisa é um aspecto muito crítico para os desenvolvedores de mecanismos de busca, inclusive se configurando como uma das linhas de investimento para IA. Percebamos o quanto é fluido e dinâmico prever o que as pessoas vão buscar e como vão buscar, ou seja, prever o pensamento humano. Na arquitetura da computação, em geral, a expressão *user intent* serve para programadores criarem possíveis caminhos de uso com soluções para cada uma dessas trajetórias, independente dos *softwares*. Programadores podem até criar intenções ou frases de treinamento, simulando possíveis propósitos de usuários e ajudando a alimentar *machine learning*.

Quadro 8 – Tipos de intenção de pesquisa

Intenção de pesquisa	Descrição	Expressões/Combinações de de busca	Exemplos
Transacional (“do”)	Atingir um objetivo ou desempenhar uma atividade como fazer um <i>download</i> , comprar ou interagir com um recurso. Representam cerca de 10% das buscas e podem incluir uma subcategoria que se relaciona com compras baratas ou com desconto. Também é conhecida como consulta de fazer. Os resultados geralmente são páginas de produto em que o próximo passo é o <i>checkout</i> ou <i>download</i> .	Comprar, encomenda, preço, barato, <i>free</i> , desconto, baixar grátis, gratuito, <i>login</i>	Fazer um <i>download</i> ou um <i>login</i> Ver um vídeo Enviar flores Fazer matrícula Comprar um telefone Instalar jogo Calcular IMC
Informacional (“know”)	Busca por informação, pretende descobrir, explorar e conhecer sobre algo. Representa 80% do volume total das buscas. Também é conhecida como consulta de saber. Pode incluir uma subcategoria de investigação comercial, onde já existe uma intenção de compra incipiente, mas o leitor precisa saber onde comprar ou comparar preços. Os resultados geralmente apresentam base para o leitor obter conhecimento	como, quando, quem, o quê, quando, por que, qual o melhor, qual é mais útil, que material é necessário para ou palavras como guia, tutorial, exemplos, dicas	Ver sintomas de uma doença Obter uma receita culinária Aprender sobre um conceito específico Conhecer um termo técnico Saber quem é uma pessoa Comparar dois ou mais produtos
Navegacional (“go”)	O leitor busca por um <i>site</i> ou uma página específica (<i>target page</i>) dentro de um <i>site</i> , mas prefere acessá-la por meio do <i>Google</i> , em vez de digitar seu endereço no navegador. Envolve a subcategoria consulta de visitar pessoalmente, em que o leitor quer encontrar estabelecimentos, seus endereços físicos e detalhes de horários de funcionamento.	[nome da marca] [nome do produto] [nome do serviço]	<i>YouTube</i> <i>Facebook</i> FAQ da Boticário Biblioteca PUC Minas Assembleia Legislativa <i>Gmail</i> Site Receita Federal Ver seção saúde <i>Folha de S.Paulo</i> Starbucks perto de mim Dentistas em BH Horários Cinemark BH

Fontes: Elaborado pela autora com base em *Google (2022b)*, *Conversion (2022)*, *SemRush (2022)*, *Power Web Marketing (2022)* *Hrefs (2022)*¹³⁵.

Em pesquisa anterior (CRUZ, 2014), relatamos o caso da busca sobre o incêndio na boate Kiss, em Santa Maria, em 2013. Tanto nessa cobertura jornalística, quanto em outras, imaginemos que as pessoas podem ter propósitos de busca diversos, procurando por fotos,

¹³⁵ Disponível em: <https://pwm.pt/tipos-de-pesquisa-nos-motores-de-busca/>; <https://www.conversion.com.br/blog/intencao-de-busca/>; <https://pt.semrush.com/blog/otimize-para-as-intencoes-nao-somente-para-as-palavras/>; <https://ahrefs.com/blog/pt/tipos-de-pesquisa/>. Acesso em: 20 mar. 2023.

informações, reportagens de portais de notícias, número de mortes, nomes das vítimas etc. Ademais, imaginemos que as pessoas podem combinar termos de modo também diverso: boate Kiss, mortes em Santa Maria, incêndio na boate Kiss, incêndio em Santa Maria etc. Naquele caso, nossa pesquisa relatou que havia um grande volume de busca no *Google* pela exata expressão “Incêndio em Santa Maria”, que pode ter impulsionado o ranqueamento de notícias que privilegiaram essa combinação de termos, mas, afinal, como um jornalista ou um produtor de conteúdo prevê a nomeação que as pessoas vão dar a um fato ou os termos que elas usarão para se referir a algum produto?

Esse é um grande desafio de quem escreve na *Web* e também de quem constrói algoritmos de mecanismos de pesquisa, como já citamos. Os padrões de busca que já se repetem para fatos sazonais – como Olimpíadas, Dia das Mães, eleições – sinalizam muito sobre os movimentos de leitura, mas não cercam todas ou novas situações. Em uma definição pouco crítica, poderíamos dizer que plataformas de busca são capazes de responder a todos os tipos de pesquisa, mas o próprio *Google* (2022b) classifica tipos de intenção de pesquisa mais comuns para se dizer preparado para essas demandas, mas talvez não esteja preparado para todas. Para os produtores de texto, estão disponíveis ferramentas – de planejamento de palavras-chave e de volume de buscas¹³⁶ – que mostram exatamente o que as pessoas procuram hoje e isso já é um grande passo para melhorar as inferências preditivas.

Em minha experiência como professora, tentando ser menos normativa possível, costumo sugerir aos alunos produtores de conteúdo que essas ferramentas sirvam apenas de apoio e que a habilidade mais importante é criação de campo semântico pertinente ao pensar um título e o texto como todo. Um texto que ofereça aos leitores múltiplos significados, mais rotas hipertextuais, possibilidades de unir sentidos, acaba dialongando com as tomadas de decisão de leitores em ambiente digital. Talvez, a habilidade do campo semântico pertinente se configure como uma ação tática dos produtores de texto em ambiente digital, pois não é coincidência que esse tipo de texto acerte mais as intenções de busca, ranqueie e performe melhor. Na seção 4.7, ao descompactar atualizações algorítmicas do *Google*, vamos compreender, por exemplo, como o surgimento dos *Core Updates Hummingbird*, *RankBrain* e, especialmente o BERT, melhoraram a capacidade da plataforma de busca em mapear os perfis de intenção do usuário.

¹³⁶ Ferramentas como *Keyword Planner*, *Keyword Tool*, *Google Trends*, *SemRush*.

4.6.2 Leitura-Escritura

A leitura-escritura das palavras-chave na caixa de texto (barra de busca) é o processo que permite ao leitor indicar a pesquisa por informação específica e serve de *input* para o motor de busca começar o rastreamento para entrega. A ação pode ser atravessada por algoritmos que ajustam erros de digitação e dão uma opção de novo texto ou por algoritmos de função “autocompletar” e preenchimento automático, em que a plataforma de busca vai oferecer previsões ou recomendações personalizadas de textos com base nas suas pesquisas anteriores.

É mais um movimento dos motores de busca tentando funcionar como a mente humana, o que Santaella (2005) chama de “modelo computacional da mente”. Os estudos de legibilidade mostram que leitores fazem naturalmente a função autocompletar ou as previsões de leitura. Liberato e Fulgêncio (2018) exemplificam: se existe um produto cuja marca é N*TURAL, provavelmente, leremos este nome como “natural”, apesar de faltar uma letra. É uma função “autocompletar” do nosso cérebro, baseada no conhecimento léxico, que inclui a palavra “natural” como já existente na língua portuguesa. Se a compreensão de texto de leitores é uma reunião entre informação visual (aquela captada pelos olhos) mais o conhecimento anterior, a tarefa do algoritmo é tentar imitar esse movimento prevendo intenções de busca com base na combinação do que foi previamente digitado (informação visual já dada no *input* da barra de busca) e no conhecimento anteriores, que nesse caso são os rastros desse leitor em ambiente digital (histórico de busca). Portanto, observando o processo de leitura, conclui-se que se o leitor não dispõe de conhecimento anterior, muito pouco do texto pode ser previsto. É assim também como a plataforma de busca que não poderá prever, caso não tenha um banco de dados estruturado. Essa justificativa auxilia a retórica do *Google* de que é preciso coletar dados para prestar um serviço relevante.

O *Google* consegue transformar em estatísticas todos os textos que entram na barra de busca e transforma esses dados em séries históricas, que sinalizam sobre como os leitores estão interagindo com o mundo da informação. Um exemplo é a ferramenta *Google Trends*, que reúne as tendências de busca de qualquer termo com recorte temporal (de 2004 até agora), local (país ou mundo), entre outros possíveis filtros. É um mecanismo de previsão do presente, que resulta de uma datificação em larga escala e transforma em dados facilmente visualizáveis, em especial, por produtores de textos que precisam apreender melhor os movimentos de busca para pautar suas escritas.

Na prática, é possível, por exemplo, saber o volume de consultas sobre “vendas de automóveis durante o primeiro trimestre de 2023” e criar *insights* úteis para prever o relatório de vendas de automóveis do próximo trimestre. Em outro caso, um divulgador científico pode comparar o volume de buscas pelos termos “BH” ou “Belo Horizonte” para saber qual deles escolher em um título de seu *post* de *blog*, objetivando atrair mais buscas para o texto, afinal, escolher usar um termo historicamente mais procurado, é uma estratégia para tornar um conteúdo mais encontrável e visível. A Figura 17 traz os termos mais buscados no *Google* em 2022, com recorte para o Brasil, e nos dá alguns *insights* sobre os movimentos de leitores usando a barra de busca.

Figura 18 – Termos mais buscados no *Google* em 2022 – Brasil



Fonte: *Google Trends*, 2022.

A expressão mais buscada no *Google* foi “eleições 2022”, trazendo a reboque a personalidade mais buscada, que foi “lula”. Por meio de ferramentas como esta,

compreendemos os anseios pela informação da vida cotidiana: “Como fazer empréstimo do Auxílio Brasil?”. Enfim, os registros dariam um potente retrato social de como leitores transportam para o ambiente digital as demandas da vida real ou como a leitura-escritura registra acontecimentos e, mais interessante ainda, como as pessoas estão nomeando esses fatos. Apesar do potencial dessa análise, interessa-nos mesmo reforçar que toda essa fotografia dos movimentos de leitura, dentro do ambiente calculável de uma plataforma infraestrutural, pode representar um grande poder de informação.

4.6.3 Leitura-Navegação

Quando um leitor se depara com uma página de conteúdos selecionados, organizados e ranqueados, é como se fossem oferecidos a ele opções de rotas a seguir. “Essas especificidades técnicas e de acesso, via interfaces programadas, franqueiam ao leitor uma organização predefinida, tornando acessível uma infinidade de possibilidades de escolhas de textos, de percursos e de fontes de informação” (JUANALS, 2007, p. 75). Por tudo isso, a leitura-navegação é o caminho para eleição de *links*, imagens, mapas, vídeos, boxes de textos ranqueados da *Search Engine Results Page* (SERP), que permitem dar saltos instantâneos para as outras páginas relacionadas ao tema procurado.

A decisão do leitor é um movimento tão monitorado e tão importante às engrenagens do *Google* quanto os que descrevemos anteriormente – intenções de busca e leitura-escritura. Isso porque o ato de clicar no conteúdo eleito ratifica duas camadas de competência: a eficiência do ranqueamento do motor de busca, ao entregar informação relevante ao leitor, e a eficiência do produtor de texto escolhido (*link* escolhido), ao cumprir relevância algorítmica e ter a chance de visibilidade para leitura. Segundo a pesquisa *SparkToro-SimilarWeb*¹³⁷, que analisou 5,11 trilhões de busca no *Google* em 2020, 33,59% dessas pesquisas resultaram em cliques nos resultados da pesquisa orgânica e apenas 1,59% resultaram em cliques em resultados de pesquisa paga, ratificando a importância de ranquear organicamente conteúdos muito mais que patrociná-los. O relatório mostra também que 66% dos cliques ocorrem nos três primeiros resultados da pesquisa, ou seja, o conteúdo estar no topo parece ser fundamental para a decisão dos leitores, afinal, a própria ideia de *ranking* influencia na opção pelos primeiros/melhores/vencedores, quando na verdade já desmistificamos aqui na tese a métrica de relevância.

¹³⁷ Disponível em: <https://sparktoro.com/blog/in-2020-two-thirds-of-google-searches-ended-without-a-click/>. Acesso em: 24 mar. 2023.

Ainda de acordo com a pesquisa, 64,82% das pesquisas no *Google* terminaram no que se chama de clique-zero (*zero-click searches*), ou seja, finalizaram no próprio *Google*, pois o leitor não escolheu um *link* que o levasse a outro *site*. O fato de um movimento de busca encerrar na leitura-navegação nos leva a pelo menos três análises: 1) o *Google* não está mais prestando o serviço de mediador/distribuidor, como historicamente fez e, assim, frustra produtores, estrategistas de conteúdo, donos de *e-commerces*, que tanto se dedicam para vencer no ranqueamento; 2) o *Google* está chegando ao almejado posto de “espaço fim de leitura”, lugar onde tudo se encontra sem precisar sair da SERP, sendo ela mesma a fonte de informação final; e 3) o *Google* pode caminhar para um “tiro no pé”, afetando drasticamente seu modelo de negócio, afinal, se o *Google* passar a oferecer na SERP a resposta perfeita para cada consulta, ninguém clicará em anúncios.

A indústria de SEO, *marketing* e produção de conteúdo criou um grande barulho no entorno do percentual de clique-zero, culpando o *Google* pelo encerramento do ciclo de leitura/consumo na SERP. O *Google* (2021), rapidamente, divulgou uma resposta pública, por meio de um texto do *Public Liaison for Search*, Danny Sullivan¹³⁸, em que ele acusava as empresas autoras da pesquisa *SparkToro-SimilarWeb* de não compreenderem como funciona a busca, ao alarmar a todos com a informação de que quase dois terços das pesquisas terminava sem cliques para direcionamento. Ademais, Danny Sullivan detalhou justamente o movimento de leitura na própria SERP, tentando convencer de que este também é um serviço prestado pelo *Google*. Vejamos os argumentos apresentados pela empresa e por que eles ajudam a compreender a leitura-navegação:

a) as pessoas reformulam suas consultas: nem todo leitor sabe formular suas perguntas quando começa a pesquisar. Alguém pode iniciar a leitura-escritura com “bolsa de couro” e perceber que, na verdade, queria encontrar “bolsa executiva de couro masculina”. Nesse caso, a primeira consulta seria considerada clique-zero, porque o leitor precisou reformar a entrada para, enfim, chegar ao *site* varejista que tem o produto desejado para finalizar a compra. Em sua defesa, o *Google* (2021) disse que, nesse exemplo, ele cumpriu a função de distribuidor tão esperada por *sites* de *e-commerce*.

b) As pessoas procuram fatos rápidos: quando o leitor procura informações como previsão do tempo, placares esportivos, conversões de moeda, conversões de fuso-horário, o *Google* geralmente entrega na própria SERP a informação final, com base em contratos de

¹³⁸ Disponível em: <https://blog.google/products/search/google-search-sends-more-traffic-open-web-every-year/>. Acesso em: 27 mar. 2023.

licenciamento ou ferramentas da plataforma que permitem curadoria legalizada desses dados. Durante as grandes coberturas jornalísticas – como Copa do Mundo e Olimpíadas –, esses serviços se evidenciam, em especial, com as agendas e calendários das competições que são entregues detalhadamente na página do buscador sem que o leitor precise clicar em *sites*. Há produtores de conteúdo, inclusive, que se especializaram na criação de textos com potencial de serem selecionados para esses campos de informação rápida. A estratégia é conhecida como conquistar as *SERP features*, ou seja, aquela posição zero em que o conteúdo será lido de forma estendida na própria SERP e que, nem sempre, converte cliques para o *site*, mas eleva aquele prestador de informação ou aquele texto à visibilidade máxima. Há algumas formas de composição de resultados, como vemos nos exemplos abaixo, que resultam no clique-zero porque, segundo o *Google (2021)*, os resultados já são suficientemente úteis para o usuário. O *rich snippet* (Figura 18) é quando a posição zero da busca é ocupada por um texto mais completo, uma tabela ou gráfico que já traz a informação final. Já o *graph search* (Figura 19) exibe a informação buscada de forma mais visual na posição zero da busca. E, por fim, o *knowledge graph* (Figura 20) a exibição de informações imediatas, sendo elas factuais ou históricas.

Figura 19 – *Rich snippet*

The image shows a Google search interface for the term "multimodalidade". The search bar at the top contains the text "multimodalidade" and shows approximately 213,000 results in 0.26 seconds. A "Rich Snippet" is highlighted with a blue box and a blue arrow pointing to it. The snippet features a purple header with the word "multimodalidade" and a yellow header with the same word. Below the purple header is a short definition: "Qualidade de multimedial, do que agrupa variados modos, féticos, aspectos, formas. Característica da comunicação que se dá simultaneamente por vários meios, formas (gesto e fala, escrita e leitura, imagem e t etc.).". Below the yellow header is a numbered list: "1. ver intermodalidade 2. coexistência de duas ou mais modalidades de comunicação, envolvendo geralmente a fala, gestos, texto, processamento de imagem, etc.". Below the snippet is a section titled "As pessoas também perguntam" with four dropdown questions: "Qual é o conceito de multimodalidade?", "O que é multimodal exemplos?", "O que é multimodalidade na educação?", and "Qual a função da multimodalidade?". At the bottom, there is a result from "ufmg.br" titled "Multimodalidade | Glossário Ceale - UFMG" with a short description: "O ensino e a aprendizagem da leitura e da escrita precisam levar em conta, atualmente, a variedade dos modos de comunicação existentes, o que chamamos de ...".

Fonte: Elaborado pela autora, 2023.

Figura 20 – *Graph search*

The image shows a Google search for "música beatles". The search results page features a "Graph Search" panel for "Os Beatles". The panel includes a grid of song cards with album covers and titles such as "Let It Be", "Come Together", "Twist and Shout", "Hey Jude", "Don't Let Me Down", "Strawberry Fields Forever", "Here Comes the Sun", "Something", and "A Day in the Life". A blue arrow points to the "Graph Search" label. Below the grid is a "Mostrar mais" button. To the right, there are sections for "Vídeos" (featuring a video thumbnail of the Beatles walking) and "Ouvir" (with icons for YouTube, Spotify, Apple Music, and Deezer). A "Sobre" section provides a brief biography of the band.

Fonte: Elaborado pela autora, 2023.

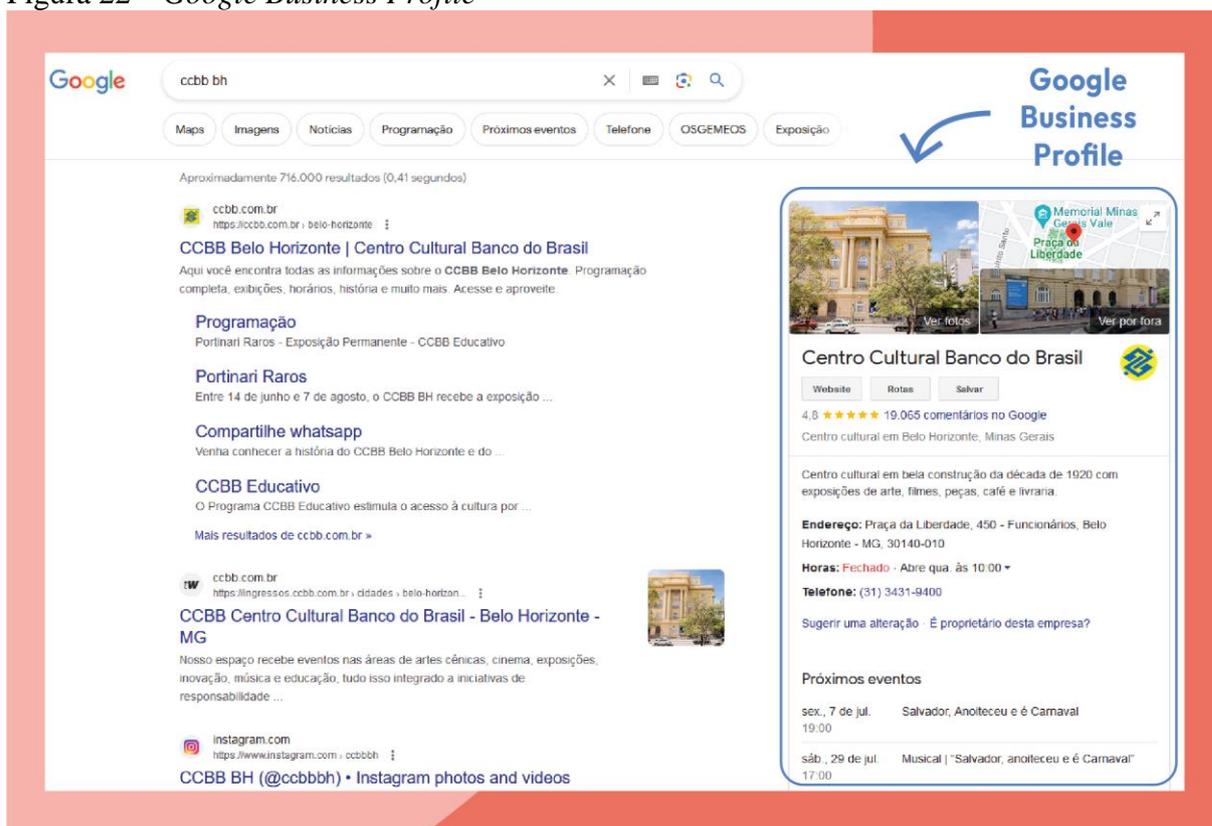
Figura 21 – *Knowledge Graph*

The image shows a Google search for "ana elisa ribeiro". The search results page features a "Knowledge Graph" panel for "Ana Elisa Ribeiro". The panel includes a grid of photos of the author, a brief biography, and links to her social media profiles (Instagram and LinkedIn). A blue arrow points to the "Knowledge Graph" label. The search results also show a list of articles and academic papers about her work, such as "Letramento digital: um terra em gêneros efêmeros" and "Letramento digital: aspectos sociais e possibilidades".

Fonte: Elaborado pela autora, 2023.

c) As pessoas se conectam diretamente com empresas: o *Google* dá a possibilidade de leitores verem informações sobre empresas diretamente na SERP, como, por exemplo, endereço, horário de funcionamento, telefone de contato. Dessa forma, acontece o clique-zero porque nem sempre o leitor precisará clicar na página do estabelecimento. Para os estabelecimentos, marcas e organizações, é preciso configurar o perfil de negócios do *Google*, preenchendo as informações básicas para atender as demandas de busca mais diretas.

Figura 22 – *Google Business Profile*



Fonte: Elaborado pela autora, 2023.

d) As pessoas navegam diretamente para aplicativos: algumas pesquisas levam os leitores diretamente a aplicativos, em vez de *sites*. Por exemplo, se um leitor pesquisar um programa de TV, verá *links* para os provedores de *streaming*. Tentamos a busca pelo programa do canal *GNT/Globo*, *Que História é esta*, *Porchat*, e o primeiro *link* da SERP foi do aplicativo *Globoplay*. Tentamos ainda a busca pela série *Stranger Things* e a primeira entrega é de um *link* “assista agora”, da *Netflix*. Se o leitor tiver esse aplicativo de *streaming* em seu telefone, esses *links* o levarão diretamente ao aplicativo. O mesmo

vale para muitos outros aplicativos, como *Instagram*, *Amazon*, *Spotify*. Portanto, não haverá direcionamento para um *site*, configurando o clique-zero.

Esse embate de especialistas em produção de conteúdo e *Google* sobre clique-zero e outros assuntos que englobam os movimentos de leitura reforçam que durante a leitura-navegação o leitor se confronta com multiplicidade de fontes, as mais heterogêneas, e formatos de texto. Segundo Juanals (2007), o leitor é levado a reconstruir o sentido a partir do conteúdo colocado à disposição ou potencialmente disponível, se não quiser se perder em um excesso de comunicação. No caso das plataformas de busca, o leitor é levado também a uma interação quase imperceptível com os algoritmos em funcionamento – aqueles que entregam respostas de busca, aqueles que contabilizam os cliques, aqueles que coletam rastros e movimentos para perfilamento desse leitor.

As interfaces veiculam uma mensagem cujo enunciado transforma a posição e o papel do leitor, que, uma vez assim interpelado, converte-se em leitor engajado numa relação, não com o enunciador, mas com o software, que cria a ilusão de dirigir-se a ele de maneira particular. Essa intervenção ativa do leitor modifica o seu papel: ele não é mais o leitor de um livro cujo discurso é para ser lido, compreendido e assimilado, mas ao qual resta basicamente o exterior (pelo menos no momento em que inicia a leitura). Ele se acha situado no centro de um texto e de um diálogo a ser construído entre ele mesmo e os dados de sua escolha, em que é solicitado a engajar-se, conforme um processo de reconstrução do sentido que ele deve desvelar. Não há mais um único percurso em direção ao conhecimento; existe daqui para frente uma multiplicidade. (JUANALS, 2007, p. 76).

Como cabe ao leitor traçar, ele mesmo, o caminho com o qual se identifica – resultado da navegação e da agregação de dados sempre singular, pontual e local. Nesse sentido, parece ser o momento da leitura-navegação o constituidor da autonomização de leitor que faz dele um “interator”. Todavia, Juanals (2007) alerta para o fato de que essa autonomia pode ser relativa, afinal, ao utilizar as performances materiais e programadas da informação, combinadas à tecnologia hipertextual e às suas capacidades de filtragem e seleção da informação, o leitor, na verdade, dialoga consigo mesmo (com seu próprio perfilamento prévio feito pelos algoritmos), e não com as instâncias de produção de maneira ideal.

Se existe uma forma de intercâmbio entre duas instâncias, há principalmente uma conexão entre o leitor-usuário e os softwares que, por intermédio das interfaces informáticas, auxiliam o leitor em suas atividades de pesquisa e de consulta de dados. Essa forma de “inteligência programada” amplia a inteligência humana, tomando o encargo do armazenamento (memória externa) e realizando uma parte das tarefas de pesquisa, de filtragem ou de classificação de dados, mas deixa ao pesquisador questões subjetivas e contextuais relativas às suas finalidades, às suas escolhas e à construção de sentido. (JUANALS, 2007, p. 78).

Ao observar esses três movimentos de uso do *Google*, inferimos que a leitura pelas conexões oferecidas pela plataforma não contribui, necessariamente, para uma autonomização do leitor, justamente porque são movimentos guiados e vigiados. É verdade também que o produtor de texto não dispõe mais da autoridade suprema sobre o plano da estruturação do discurso, já que é impossível antecipar todas as combinações de escolhas possíveis do leitor, conforme veremos no Capítulo 5. Portanto, vale trazer o questionamento de Juanals (2007) sobre escrita e leitura em ambiente digital, em que ela sugere certo continuísmo em relação a outros suportes, como o impresso: “[...] essa forma de ‘liberdade vigiada’ promove a evolução de uma abordagem centralizadora *versus* uma abordagem descentralizadora?” (JUANALS, 2007, p. 74).

4.6.4 Interação pós-SERP

Esse último movimento de uso do *Google* que vamos descrever é o que mais interessa aos produtores de texto. A partir do momento que um leitor clica em um *link* para deixar a SERP e seguir rumo à interação com o conteúdo eleito, é como se o processo de mediação da plataforma de busca se encerrasse. Mas, na visão do produtor de texto, é nesse ato de clicar que a interação começa. A chegada do leitor – com origem na plataforma de busca, ou seja, com origem no tráfego orgânico – é rastreada pelas ferramentas de *analytics*, em especial, o *Google Analytics*, como demonstramos em pesquisa anterior (CRUZ, 2014). Essa ferramenta coleta e compila dados em relatórios, ajudando no reconhecimento do público-leitor (com dados geográficos) na análise de comportamento, interesse e engajamento (com a lista de palavras-chave buscadas para chegar ao *site*), na compreensão de taxas conversão (com o número de leitores que retornam ao *site* ou quantos cliques em conteúdos foram convertidos em vendas de produtos e serviços).

Nesse ponto, é possível retomar Anne Helmond (2015), quando analisa a programabilidade das plataformas e as compreende a partir dos fluxos de dados. Para a autora, um dos alicerces de funcionamento das plataformas é configurar canais que possibilitam fluxo de dados, inclusive, para terceiros. Ferramentas de *analytics* – não só do *Google*, mas *Facebook Analytics*, *Instagram Analytics* e *Twitter Analytics* – parecem fechar um ciclo da plataformização quando envolvem os produtores de texto no mito relevância, por meio de métricas. Rieder e Rohler (2019) alertam para o gosto exacerbado pelos modos de análise quantitativos ou “científicos” que podem indicar, na verdade, uma virada computacional

indesejável por nós sociedade, especialmente, para os setores que lidam com a produção cultural (imaterial).

Amplos e, muitas vezes, incomuns conjuntos de dados, técnicas de visualização avançadas e processamento impreciso conduziram, por muito, alguns daqueles que tinham em mente números, cálculos e computadores, a uma distância segura para prepararem as novas possibilidades computacionais. Assim, a nossa questão central era: se esses novos métodos são mais do que apenas outro conjunto de ferramentas em nosso arsenal, como lidamos com as transformações fundamentais que desafiam práticas e paradigmas epistemológicos estabelecidos? (RIEDER; ROHLER, 2019, p. 20).

Na prática, esses instrumentos de medição oferecidos pelo próprio *Google* e por outras empresas, para que produtores conheçam os movimentos de leitura, são responsáveis por validar numericamente as ações de escrita, informando em tempo real quais texto performam melhor. Isso auxilia na reconfiguração de práticas, nas decisões editoriais, determina estratégias de conteúdo e, muitas vezes, de negócios. Significa que, atualmente, empresas cujo serviço fim é a produção de texto em ambiente digital, a exemplo do jornalismo e do *marketing*, parametrizam caminhos a seguir de curto, médio e longo prazo por meio de dados coletados nas ferramentas de *analytics*. No caso do jornalismo, Livia Vieira (2018) afirma que a notícia *online* é moldada por um conjunto de procedimentos técnicos que interfere em decisões editoriais, e essa dinâmica se dá por meio do monitoramento das audiências ativas, que têm seus comportamentos de navegação transformados em dados, processados de acordo com indicadores – *pageviews*, visitantes únicos, engajamento, recirculação, fidelidade, tempo de permanência. Nesse sentido, “[...] toda audiência ativa é também uma audiência observada”. (VIEIRA, 2018, p. 224).

Isso abriu possibilidades até para profissionais multidisciplinares atuarem como analistas desses dados, de modo a interpretá-los em âmbito corporativo e ajudar a traçar planos. A seguir, relacionamos algumas métricas mais comuns¹³⁹, mostrando como elas entrelaçam a interação pós-SERP e com as práticas adotadas por produtores.

¹³⁹ Muitas dessas métricas foram alteradas em 1º julho de 2023, após a desativação da versão *Universal Analytics* e a estreia do G4 – *Google Analytics 4*. O novo sistema permite conferir as jornadas do leitor e não depende exclusivamente dos *cookies*, aqueles arquivos que servem para identificar usuários que visitam um determinado *site*. A mudança é uma resposta ao processo de alteração, ocorrido em 2020, a partir da aprovação da Lei Geral de Proteção de Dados (LGPD), que não permite a coleta compulsória de *cookies* dos usuários. Apesar de não serem os únicos dados, os *cookies* são informações importantes para alimentar sistemas de *analytics*.

Quadro 9 – Movimentos do leitor vistos nas ferramentas de *analytics*

Métrica	O que representa como movimento do leitor?	Por que é importante para produtores de texto?
Usuário/ Novo usuário	Contabilização de pelo menos uma utilização de uma sessão em um <i>site</i> durante um intervalo de datas/ contabilização de utilizadores pela primeira vez de uma sessão em um intervalo de datas.	Ajuda a compreender a quantidade real de leitores que estão visitando os conteúdos. Além disso, ajuda a entender se é um público fiel, que volta recorrentemente, ou se novos leitores chegam com frequência. Exemplo: se tenho um <i>blog</i> sobre música e no Dia Nacional do Rock produzi um texto que trouxe muitos leitores ao <i>site</i> com origem de tráfego orgânico, ou seja, os leitores vieram por meio do <i>Google</i> , há grandes chances de naquele evento aumentarem os números de novos usuários, afinal, as pessoas vieram ao <i>blog</i> por meio de um <i>link</i> , de modo fragmentada e topográfica, somente por causa do tema sazonal. O produtor poderá usar como estratégia investir nesses conteúdos sazonais ou entender que um público de roqueiros pode ser mantido no <i>blog</i> a partir dessa primeira visita, portanto, trabalhar mais conteúdos de <i>rock</i> .
Visualizações de páginas	Número total de visualizações de páginas. São contabilizadas as visualizações repetidas em uma mesma página.	Esta é uma informação importante para monetização dos conteúdos. Se compararmos com o conceito de audiência – que se refere às visualizações de página é que vão servir para legitimar se um conteúdo funcionou ou não em termos de audiência e performance. Portanto, é um dado que guia, principalmente, produtores de texto performáticos que pretendem crescer essa métrica. Esse dado se relaciona muito com o critério da popularidade, base essencial de funcionamento dos motores de busca, porque indica quantitativamente as entradas em uma página. Se são muitas entradas, há chances de que essa página seja enquadrada como popular.

Métrica	O que representa como movimento do leitor?	Por que é importante para produtores de texto?
Taxa de rejeição	A porcentagem de visitas do <i>site</i> a uma única página, ou seja, não interagiu com o <i>site</i> , saiu na mesma página que entrou. Ex.: usuário acessou a sua <i>home</i> , não foi para uma outra página e saiu, isso é uma rejeição.	Indica um movimento de leitura topográfico e fragmentado em que o leitor não navega, ele entra por um <i>link</i> e não explora outros conteúdos de um <i>site</i> . O produtor de texto deve tentar reduzir a taxa de rejeição, oferecendo ao leitor caminhos para ficar no <i>site</i> , navegar e permanecer em estado de leitura naquele espaço. Uma estratégia técnica-textual para isso é a inserção de <i>hiperlinks</i> e chamadas ao leitor para <i>links</i> internos, ou seja, outras páginas do próprio <i>site</i> . Ex.: o <i>Google Analytics</i> considera algumas taxas de rejeição comuns para segmentos específicos. No caso do jornalismo, entre 40% e 60% das visitas terminam em uma única página, o que faz sentido se pensarmos que um leitor pode sair do <i>Google</i> em busca de uma informação única ou pode entrar em um portal de notícias em busca somente de se atualizar sobre um fato.
Tempo médio na página	A média de tempo que leitores passaram vendo uma página.	Ajuda a inferir se o leitor consumiu todo o texto ou não o leu por completo. O trabalho do produtor de texto nesse caso é aumentar o tempo médio e tentar manter o leitor na página, dando a ele materiais multimidiáticos que puxem o interesse e mantenham o processo de leitura, ou criando chamadas no texto que façam parecer um diálogo em andamento. Ex.: a inserção de galeria de fotos, vídeos e áudios são estratégias usadas para aumentar tempo médio, porque o leitor gastará mais tempo consumindo o texto em outros formatos que não só o verbal.
Palavra-Chave	Expressões usadas pelos leitores para chegamem ao <i>site</i> .	O sistema <i>analytics</i> entrega ao produtor de conteúdo como se deu processo de leitura-escritura do seu público, o que ajuda o profissional do texto a entender os termos importantes para o universo do conteúdo que produz. Esse auxilia na técnica-textual de titulação de conteúdos, por exemplo, afinal se sabe quais

Métrica	O que representa como movimento do leitor?	Por que é importante para produtores de texto?
Origem de tráfego/mídia	Retrata onde os leitores estavam antes de verem o conteúdo e por qual meio chegaram ao texto. Esses lugares de origem podem ser: plataformas de busca ou de redes sociais, um <i>e-mail</i> , ou algum outro <i>site</i> .	<p>palavras funcionam como atratoras de leitor para aquele <i>site</i>.</p> <p>Com esse dado, o produtor de texto compreende o processo de circulação do seu conteúdo, inferindo quais são as principais plataformas mediadoras da sua audiência. Ex: se muitos leitores estão chegando a um <i>site</i> após clicarem em um <i>link</i> no <i>Facebook</i>, quer dizer que talvez essa plataforma seja um espaço importante de circulação. A origem de tráfego é o dado que auxilia na decisão de investimento de mídia. Uma empresa de <i>marketing</i>, por exemplo, pode pensar que se a origem de tráfego mais representativa é a busca orgânica, talvez seja melhor continuar investindo em técnicas de SEO do que em conteúdos para redes sociais.</p>

Fonte: Elaborado pela autora, 2023.

Atentar-se a essas métricas é importante para produtores de texto não somente para entender movimentos do leitor e a reação deles com conteúdos, mas para se legitimar diante do *Google*, afinal é sabido que a plataforma de busca conhece esses dados e pode usá-los para indicar popularidade, autoridade ou qualidade. Isso nos remete à discussão do Capítulo 2 desta tese sobre “jornalistas metrificados” e autores de livros espalhados em um espectro de modos produtivos mais ou menos aderentes às lógicas das plataformas de autopublicação. Nosso olhar crítico é para o fato de que, atualmente, essas métricas e números atravessam a relação entre esfera de produção e esfera de leitura, determinando o encontro entre quem escreve e quem lê.

Enfim, enfatizamos o forte monitoramento do *Google* e dos produtores de texto sobre os movimentos dos leitores em todas as instâncias aqui descritas – intenções de busca, leitura-escritura, leitura-navegação, interação pós-SERP – para argumentar a materialidade de plataforma *Google* na triangulação do sistema produção, circulação e leitura. O leitor nesse processo é, sim, dono dos fragmentos que reúne, mediante uma lógica associativa e de mapas cognitivos personalizados e intransferíveis (SANTAELLA, 2004), mas as especificidades de uso das plataformas de busca o colocam diante da engrenagem de leitura topográfica que se

torna literalmente escritura, e vai formando as mensagens à medida que se dão as interações leitor-algoritmos-produtor-algoritmos-leitor.

4.7 DESCOMPACTANDO O CONJUNTO SOCIOTÉCNICO DE ALGORITMOS

Como último passo para a caracterização da plataforma de busca *Google* e compreensão dessa infraestrutura para a plataformização dos textos, trazemos uma breve análise sobre atualizações de algoritmos (*Core Updates*)¹⁴⁰ que mais influenciam na produção textual em ambiente digital. Nosso foco é a implementação do BERT, em 2019, mas faremos uma retomada até esse ponto para contextualizar. O próprio *Google* (2023m) define os *Core Updates* como mudanças que visam melhorar o sistema de avaliação de conteúdo em geral. Defendemos que a manipulação desses algoritmos só tem consequências porque esses códigos têm significados. Ochigame (2021a) nos lembra que a computação também é um processo semiótico. Nesse sentido, a metáfora de “linguagem” para programação de computadores não é injustificável. Para o autor, assim como línguas codificam as suposições culturais de seus falantes, o computador codifica línguas.

Muitos dos valores, metas, objetivos comerciais e simbólicos do *Google* estão expressos no imenso conjunto sociotécnico de algoritmos, que incentiva uma rede de profissionais do texto e de desenvolvimento Web a fim de reconstruir seus modos de trabalho para se adequar ou resistir ao imperativo da plataforma de busca. Os *Core Updates* são também respostas da plataforma aos usos táticos e estratégicos que as pessoas fazem, principalmente os especialistas em SEO e produção de texto para *marketing*, que acabam criando sempre novos métodos de performance na SERP.

Entre 1998 e 2013, apesar das muitas atualizações de algoritmos, o *PageRank* continuava representando uma base muito importante para o processo de seleção do *Google*, especialmente, com o esquema de relevância dos *links*. Na direção de tornar o motor de busca mais qualitativo, sem eliminar a importância dos sinais coletados pelo *PageRank*, em 2003, o *Google* lançou o algoritmo *Florida* para remover *sites* de baixa qualidade, especialmente os *spams* que usavam palavras-chave de modo excessivo e aleatório; em 2011, implantou o *Panda*, que penalizava *sites* com grande quantidade de anúncios.

Em 2012, lançou o algoritmo *Penguin*, conhecido como *Webspam Update* e que também teve como foco conter ações excessivas de otimização como esquemas de *black hat*

¹⁴⁰ Disponível em: <https://status.search.google.com/products/rGHU1u87FJnkP6W2GwMi/history?hl=pt-br>. Acesso em: 8 jun. 2023.

SEO e *spamdexing*, que ações antiéticas, deliberadas e maliciosas de manipulação de mecanismos de buscas com o objetivo de aumentar a relevância de um *site* em resultados. Algumas dessas técnicas consistem em acrescentar dentro de *sites* como fóruns, *blogs*, listas de discussões palavras-chaves e *links* sem relação com a página, visando somente a encontrabilidade por meio do termo. O *Penguin* também foi uma reação ao *Google bomb*, fenômeno *hacker* sem fins comerciais que tentava influenciar a classificação de uma determinada página nos resultados da SERP normalmente com intenções humorísticas ou políticas. O caso mais conhecido foi o de ativistas que otimizaram o termo “fracasso miserável” para vincular à página oficial do presidente George W. Bush. Ainda em 2012, o *Google* lançou o recurso *Knowledge Graph*, que já citamos na seção 4.6.3, e serve para fornecer respostas mais contextualizadas diretamente na SERP, em vez de apenas listar *links*. Essas mudanças culminaram em uma revisão completa da lógica algorítmica da plataforma de busca em 2013, que foi o *Core Update Hummingbird*, implantado em 26 de setembro daquele ano.

O *Hummingbird* serve para que o mecanismo de pesquisa entregue respostas para além da palavra-chave literal digitada na barra de busca. De acordo com a principal patente¹⁴¹, ele é um sistema de identificação e avaliação de sinônimos e termos não adjacentes usados para a revisão de consultas no *Google*. É capaz de determinar valor de confiança para pares de termos adjacentes, estabelecendo uma medida de coerência entre eles, capacidade importante para um direcionamento de busca menos literal e mais semântica. Com isso, a plataforma passa a analisar, durante o processo de seleção, o contexto em que os termos estão inseridos, o que foi considerado uma expansão de consulta (ampliação de abordagem) por entender melhor as pesquisas de linguagem natural mais longas (ou termos cauda longa, que detalharemos no Capítulo 5). A mudança afetou cerca de 90% dos resultados de busca, de acordo com *sites* especializados. Vejamos um exemplo: para a pergunta “Qual é o melhor lugar para comer pão de queijo em BH?”, o *Google* poderia entender que o leitor provavelmente ficaria mais satisfeito com a entrega de “restaurante” como sinônimo de “lugar”. Assim, os algoritmos avaliam a coerência desse sinônimo-substituto em potencial para a entrega na SERP. Muitos especialistas em otimização analisaram essas mudanças como sinais de que o *Google* caminhava para um processo de entrega mais conversacional e humana, que pudesse, inclusive, compreender as intenções de pesquisa e a busca por voz que começava a crescer naquela época.

¹⁴¹ *Synonym identification based on co-occurring terms*. Disponível em: <https://patents.google.com/patent/US8538984B1/en?q=US+Patent+8%2c538%2c984>. Acesso em: 8 jun. 2023.

Em 26 de outubro de 2015, foi lançado o *Rankbrain*, que é uma evolução do *Hummingbird*. O sistema é popularmente conhecido como o “cérebro humano do *Google*”, justamente por ser um processamento de dados dos resultados de busca, baseado em inteligência artificial e aprendizado de máquina. Os algoritmos do *Google* passaram a interpretar ainda melhor as consultas e, sem a intervenção humana, eles aprendem automaticamente sobre a intenção de busca e entrega resultados mais assertivos. São três as principais patentes¹⁴² que descrevem o funcionamento e entre as principais funções estão definição de que termos de consulta sequenciais representam um conceito, recuperação de dados anteriormente armazenados para associar aos resultados de pesquisa e delimitação de pontuação combinando as métricas e os pesos para classificar os resultados. Por causa desse último funcionamento específico, o *Rankbrain* tornou-se o sinal mais importante que contribui para seleção e ranqueamento, ao lado da relevância de *links* (definida pelo *PageRank*).

O foco principal do *RankBrain* está em consultas inéditas e em fornecer as informações mais relevantes usando dados históricos de pesquisa (mesmo dados com apenas alguns meses), em vez de confiar apenas nos dados da página de destino. Um exemplo disso é a aplicação a notícias locais ou mundiais. Embora digitar “pandemia” possa ter resultado anteriormente em definições ou estudos de caso anteriores, no início de 2020, o *RankBrain* aprendeu a identificar a intenção dos pesquisadores de entender mais sobre o novo coronavírus. O *RankBrain* também influenciou como os profissionais de texto tentam classificar para determinadas consultas. Em vez de tentar usar estrategicamente o sistema por meio de sinais específicos (por exemplo, preenchimento de palavras-chave), esses produtores devem se concentrar na criação e entrega de conteúdo de alta qualidade de uma perspectiva humana. Portanto, é possível perceber um enfoque do *Google* em aperfeiçoar a perspectiva textual e contextual da busca, emulando cada vez mais os modos de pensar e escrever humanos. Essa movimentação se completa em 2019, com o *Core Update BERT*.

4.7.1 BERT

Apesar da importância dos *Core Updates* anteriores, o *Bidirectional Encoder Representations from Transformers* (BERT) é nosso ponto de atenção, porque aconteceu no

¹⁴² *Ranking search results based on entity metrics*. Disponível em: <https://patents.google.com/patent/WO2014089776A1/en>; *Using concepts as contexts for query term substitutions*. Disponível em: <https://patents.google.com/patent/US9104750>; *Question answering using entity references in unstructured data*. Disponível em: <https://patents.google.com/patent/US9477759B2/en>. Acesso em: 8 jun. 2023.

ano de início da pesquisa de doutorado, portanto, toda a análise documental e simulações em espaço de escrita e edição, ou seja, a coleta de dados é feita após essa atualização. O BERT é o carro-chefe do *Google* e institucionaliza ainda mais a plataformização do texto, especialmente, quando entendemos esses processos considerando uma grande rede de atores – humanos e não humanos. Ademais, o BERT acontece em concomitância com a implementação de diretrizes EEAT e contextualmente, em nossa interpretação, configuram uma resposta do *Google* a momentos políticos e sociais marcantes. A plataforma de busca enfrentou críticas sobre o quanto seu modelo de funcionamento permitiu e foi preponderante na proliferação de notícias falsas e desinformação nos meses que antecederam e imediatamente após a eleição nacional dos EUA em 2016.

Metaxa e Torres-Eheverry (2017), em busca de compreender como os processos democráticos e políticos nos Estados Unidos foram afetados pelo *Google*, construíram três linhas principais de argumentos. Todos já foram analisados aqui na tese, mas vale retomar: 1) a confiança tácita das pessoas na entrega e ordenação das plataformas de busca; 2) os vieses algorítmicos atuando de forma determinante no espriamento de conteúdos; e 3) a possibilidade de o *Google* influenciar intencionalmente seus resultados de pesquisa ou sugestões de pesquisa do usuário para obter algum resultado político. Nesse último caso, a pesquisa de Metaxa e Torres-Eheverry (2017) não revelou evidências de manipulação do algoritmo para fins políticos, mas coletou casos em que a plataforma demonstrou claros posicionamentos. Em janeiro de 2012, o *Google* apagou seu logotipo na página de busca em protesto contra dois projetos de lei, o *Stop Online Piracy Act* (pare com a pirataria *online*, em tradução livre) e *Protect IP Act* (ato para proteção da propriedade e intelectual), que ampliariam os meios legais para que detentores de direitos autorais possam combater o tráfego *online* de propriedade protegida. Os projetos previam bloqueio de acesso a plataformas e aplicações na internet que fossem consideradas violadoras da propriedade intelectual. Já em 22 de maio de 2015, a página inicial do *Google* na Irlanda disse a seus usuários que eles deveriam *#VoteYes* no referendo irlandês sobre o casamento *gay*. Esses exemplos completam o caso brasileiro em que o *Google* é investigado pela Polícia Federal (PF), em 2023, suspeito de privilegiar o ranqueamento de conteúdos contra o PL 2.630/20, inclusive fazendo circular conteúdos desinformativos.

A conclusão de Metaxa e Torres-Eheverry (2017) é de que a desinformação se espalhou por meio do *Google* em resultados de pesquisa política naquela época, apesar desse espriamento ter acontecido em menor proporção que nas plataformas de redes sociais (como *Facebook* e *Twitter*). Para os pesquisadores, o *Google* investiu mais rápido em medidas para

mitigar o problema das notícias falsas e se posicionou para manter o nível de confiança dos usuários, por isso não teve a reputação tão abalada como outras plataformas *online* no pós-eleição dos EUA.

Afinal, por que foi importante retomar aqui a preponderância do *Google* na proliferação de desinformação e lembrar os posicionamentos da empresa em momentos políticos? Justamente para apoiar a análise feita por *sites* especializados¹⁴³, na época do lançamento do BERT, anunciando este *Core Update* como a maior mudança dos últimos cinco anos e que impactaria uma em cada dez pesquisas. Grande parte desses *sites* especializados citou o *Core Update* BERT como uma reação de posicionamento do *Google* a qualquer investida contra seu modelo de negócio, principalmente, contra qualquer acusação de espalhar desinformação.

Tecnicamente, o BERT é um esquema de aprendizado profundo (*deep learning*) para pré-treinamento de processamento de linguagem natural (NLP). O NLP é um campo interdisciplinar com componentes de linguística computacional, ciência da computação, aprendizado de máquina e inteligência artificial. O NLP está preocupado em analisar, decifrar, entender e dar sentido à linguagem natural do ser humano. O BERT, portanto, é uma solução técnica para algumas das deficiências em sistemas convencionais no campo do NLP.

De acordo com a principal patente do *Core Update*¹⁴⁴, por meio de um modelo de linguagem pré-treinado, os algoritmos serão capazes de compreender, prever precisamente, comparar e associar e responder com elevado nível de confiança a conjuntos de texto. O sistema BERT pode descobrir o contexto completo de uma palavra observando os termos que vêm antes e depois (bidirecional), o que o torna único. É que outros modelos de NLP tradicionais treinam os modelos de linguagem na ordem da sequência de palavras (da direita para a esquerda ou da esquerda para a direita). Os algoritmos do sistema BERT usam o contexto e as relações de todas as palavras em uma frase, em vez de analisar uma por vez. Isso significa uma grande melhoria na interpretação de uma pesquisa e a intenção por trás dela, pois ajuda computadores e máquinas a entenderem a linguagem de modo mais semelhante aos humanos. Para facilitar associações e interpretações, o sistema BERT é construído em representações contextuais de pré-treinamento, os chamados *transformers*, que foram treinados em um *corpus* gigantesco de texto que ultrapassa 33 milhões de itens. Por

¹⁴³ A cada *Core Update* *sites* como *Agência Mestre*, *Conversion*, *Blog Yoast SEO*, *Blog Moz*, *Resultados Digitais* e *Neil Patel* trazem análises técnicas e contextuais das mudanças.

¹⁴⁴ *Pre-trained contextual embedding models for named entity recognition and confidence prediction*. Disponível em: <https://patents.google.com/patent/US11568143B2>. Acesso em: 12 jun. 2023.

fim, o BERT é considerado um *Core Update* complementar ao *Rank Brain*, que foi o primeiro sistema de IA do *Google*.

Vejamos um dos exemplos apresentados pelo próprio *Google* no lançamento do BERT e observemos como o ponto de atenção desse exemplo é uma compreensão mais eficiente da preposição. Se alguém buscar por “viajantes do Brasil para os EUA em 2019 precisam de visto?”. A preposição “para” e sua relação com as outras palavras na consulta são particularmente importantes para entender o significado. É sobre um brasileiro viajando para os Estados Unidos e não o contrário. “Anteriormente, nossos algoritmos não entendiam a importância dessa conexão e retornavam resultados sobre cidadãos americanos viajando para o Brasil” (NAYAK, 2019). Com o BERT, a pesquisa consegue captar essa nuance e saber que a palavra “para” realmente importa.

Enfim, podemos perceber que o BERT atua de modo diferenciado em consultas mais complexas, em especial, para busca com frases-chave de cauda longa. É por isso que ele é um sistema algorítmico de grande influência na produção textual, por trazer novas camadas de inteligência ao buscador que exigirão de quem escreve esforços de correspondência cada vez mais eficientes (no sentido de busca por ranqueamento, visibilidade e circulação). Durante o processo de pesquisa, acompanhamos¹⁴⁵ a percepção de especialistas em SEO e produção de conteúdo em ambiente digital, não como uma metodologia, mas como um processo complementar. Observamos que o *Core Update BERT* é visto como um conjunto de algoritmo que está mudando a forma como buscamos na internet de modo que as palavras-chave terão importância, mas o contexto em que elas aparecem será decisivo. O uso de uma linguagem natural – próxima da oralidade ou dos modos de diálogo que usamos no dia a dia – terá protagonismo. Assim, a produção textual que pretende performar (circular) deverá atender cada vez mais dessa intenção da plataforma de busca em entregar conteúdos conectados aos modos de pensar e ao sistema de expectativas humanos.

Neste Capítulo 4, ao enfatizar as lógicas o funcionamento, governança, construção de “métricas” e conjunto sociotécnico de algoritmos, caracterizamos sistematicamente o *Google* como plataforma cuja a indissociabilidade entre os serviços prestados, estratégias comerciais e datificação são responsáveis por novos modos de escrever, editar e ler textos em ambiente digital. Ou seja, a inserção de critérios algorítmicos e normatizações definidas pelo *Google* na cadeia de produção textual são pilares da platformização do texto.

¹⁴⁵ Por meio de *blogs* e fóruns em que opinam especialistas da *Conversion*, *SemRush*, *Hubspot*, *Agência Mestre*, *Yoast SEO*, *Moz*, *Resultados Digitais* e *Neil Patel*.

Aos meus colegas jornalistas,

Depois de 7 anos longe da rotina das redações de jornal, mas ainda imersa no universo vivenciado por colegas contemporâneos, gente que permaneceu e permanece, a impressão é de que alguns profissionais não se surpreendem mais com o fato de um certo “imperativo algorítmico” atravessar as práticas de escrita. Incorporaram esses afazeres de forma automatizada, o que é preocupante.

No entanto, há outros jornalistas que se colocam no oposto distante, negando-se a compreender e embrenhar no jogo das plataformas online e apresentando uma resistência nada subversiva ou uma rebeldia por ignorância, daquelas que chamamos “sem causa”.

Convidei um colega para participar de uma aula minha, em um curso de pós-graduação, a fim de partilharmos vivências sobre escrita e produção textual em ambiente digital. Ele, um jornalista esportivo e experiente, apresentou alguns relatos de rotinas de um modo muitíssimo normalizado: “somos orientados a fazer parágrafos menores”, “a gente tem que fazer títulos de até 75 caracteres, porque o sistema nem aceita mais que isso”, “precisamos incluir palavras-chave no título, de preferência o início dele”.

Questionei sobre o imperativo que ali se desenhava e de onde vinham tantas ordens em relação à produção textual. Fiquei pensando silenciosamente: os editores, as empresas, os próprios repórteres definiram que todos devem escrever a partir de tendências de otimização e performance de conteúdo que obedecem à lógica de visibilidade do Google?

Não sei dizer quais e quantos colegas são suficientemente empoderados em relação ao conhecimento que se constrói acerca das plataformas e das influências algorítmicas, mas me preocupou o modo como aquele jornalista expressou a naturalização da reconfiguração de suas práticas.

Em contrapartida, certa vez, preparei uma breve oficina de SEO para colegas jornalistas com os quais eu trabalhava. Aconteceu em 2017, momento em que boa parte da grande mídia começava a considerar a otimização como fundamental à sobrevivência, pois perceberam que a origem de tráfego dos buscadores representava muito para as coberturas jornalísticas. Lá pelas tantas na oficina, depois de discutir funcionamentos e modos de fazer, um colega se manifestou: “Eu me recuso a escrever para o Google! Quer dizer agora que o Google vai definir como eu faço meus títulos?”.

Bem... O Google tem, sim, re(definido) modos de escrita e edição, o que não significa dizer que precisamos escrever PARA a plataforma de busca, talvez o melhor seja escrever COM a plataforma de busca. Devemos, sim, questionar o determinismo de uma escrita a partir de parâmetros de ranqueamento. É bem isso que estou fazendo aqui na tese, acreditando que um dos caminhos para a rebeldia seja compreender as entranhas, cercar a caixa-preta (mesmo que não consigamos abri-la). Jogar esse jogo pode ser o único jeito de enxergar brechas. Vale a pena questionar, porém é fundamental conhecer os modos de operação e abraçar causas.

Luana Cruz

5 PLATAFORMIZAÇÃO DO TEXTO: PRÁTICAS DE ESCRITA E EDIÇÃO INCORPORADAS

Neste capítulo, vamos adentrar o universo da parametrização contingenciada da produção textual em ambiente digital, analisando as implicações. O funcionamento e as diretrizes do *Google*, o mito da relevância e o funcionamento do sistema algorítmico BERT atravessam a rotina dos produtores e editores de texto justamente no momento em que esses profissionais incorporam práticas. No relato a seguir, a jornalista Natália Viana, da *Agência Pública*, ironiza a reinvenção de uma escrita em ambiente digital sob a batuta das lógicas algorítmicas, especialmente, com indicações de ferramentas e *plugins* que emulam o funcionamento do *Google* para “ensinar” o produtor de texto a escrever de forma mais otimizada. Neste capítulo, verificamos como essa reconfiguração acontece.

Existem várias ferramentas que analisam se o seu conteúdo está “apropriado” para aparecer bem na busca do Google. Eles chegam a dizer que os parágrafos estão muito longos, que você deveria colocar intertítulos aqui ou ali, que as frases tinham que ser mais curtas e diretas. Inventam um jeito de escrever que é o dos robôs e não o nosso. (VIANA, 2023).

Gillespie (2018) fomenta esse debate da incorporação de práticas. Para o autor, apesar de ser difícil documentar como usuários podem mudar suas visões de mundo para acomodar as lógicas subjacentes e os pressupostos implícitos dos algoritmos que usam regularmente, há uma razão para que as lógicas de funcionamento desses algoritmos não apenas moldem as práticas, mas os levem a internalizar normas e prioridades. Essa razão é o modo como as plataformas *online* habitam seus usuários por meio de configurações cuidadosamente projetadas.

O *Google*, por exemplo, projeta-se para oferecer um serviço relevante, à medida que se torna mais comum e confiável, fazendo com que suas lógicas sejam autoafirmativas. As entregas da plataforma de busca são asseguradas, em grande parte, pela ratificação pública, conforme já discutimos ao explicitar o critério da popularidade. Para Gillespie (2018), o *Google* e seu algoritmo ajudam a afirmar e a normalizar essa lógica de conhecimento como “correta”, tão correta quanto seus resultados parecem ser. Defendemos que uma das principais consequências disso é que produtores de texto contingenciam sua produção para que seus textos apareçam, sejam ratificados e possam competir por atenção. É um jeito de garantir *feedback* positivo, pois saber que um texto foi escolhido para o *ranking* e,

ainda por cima, clicado pelo leitor, fecha um ciclo de sucesso, mesmo que seja um sucesso dotado de miopias.

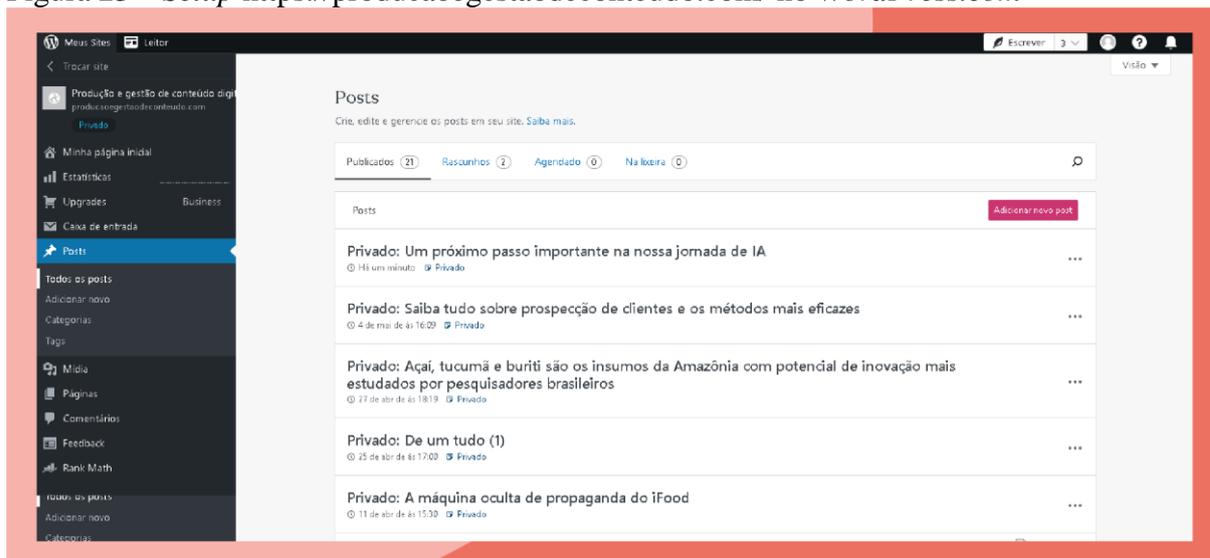
5.1 ESPAÇO SIMULADO DE ESCRITA E EDIÇÃO

Na etapa 2 de nossa pesquisa, criamos um espaço para verificar a incorporação de práticas de escrita e edição, submetendo cinco textos já publicados na *Web* a dois *plugins* do *WordPress.com*. Nossa intenção foi testar e parametrizar essas produções textuais, editando-as para que pudéssemos – simuladamente – fazer esses conteúdos serem mais elegíveis aos algoritmos da plataforma de busca *Google*. O método consistiu em inserir o texto escolhido no rascunho de um *post* de *WordPress.com*, dentro do site <https://producaoegestaodeconteudo.com/>, criado exclusivamente para atividades de ensino, aprendizagem e pesquisa. Ao submeter os textos às análises oferecidas pelos *plugins*, conseguimos detalhar as fases de edição desses textos para verificar o quanto é possível alterar uma produção para que ela seja mais ranqueável, já que os *plugins* têm a função de automatizar esse serviço e propiciar uma escrita mais encontrável.

Muniz Jr. (2020) define o ato de editar como interferir em textos, sons, imagens estáticas e imagens em movimento. Diz também que editar é trabalhar para que uma materialidade discursiva ganhe adequada circulação. Interferimos o menos possível nos textos, editando só o necessário para alcançar os melhores padrões indicados pelos *plugins*, um movimento de testagem que nos ajudou a verificar reconfigurações das práticas de escrita. Trabalhamos com os textos na função “privado” do *WordPress*, em que os conteúdos não ficam visíveis para o público¹⁴⁶, afinal, a intenção nunca foi publicá-los, mas usá-los como *inputs* nos *plugins*, inspirando-nos em metodologias de engenharia reversa.

¹⁴⁶ Os *posts* privados são indexáveis pelos mecanismos de busca. Isso significa que alguém poderia chegar ao *link* (URL) do nosso espaço simulado de textos, mas assim que clicasse não conseguiria visualizar a página.

Figura 23 – Setup <https://producaoegestaodeconteudo.com/> no *WordPress.com*



Fonte: Elaborado pela autora, 2023.

O *WordPress* é um CMS projetado há 20 anos para atender, em configuração mínima, a leigos que não sabem criar uma página na *Web*. O *WordPress* é usado por 63,2% de todos os *sites* cujo sistema de gerenciamento de conteúdo conhecemos. Isso representa 43,1% de todos os sites da *Web*. Ele está quase isolado à frente dos concorrentes, pois, em segundo lugar, vem o *Shopify*, com 5,6% e, em terceiro, o *Wix*, com 3,6%. No *WordPress.com*, o *site* fica integralmente hospedado, com funcionalidades autogerenciadas, opções de *plugins* e temas personalizados. Nesse caso, não há necessidade de contratar um provedor de hospedagem, nem baixar *software*, diferentemente do *WordPress.org*, que exige hospedagem de terceiros e *download* do *software*. Para o espaço simulado de escrita e edição, usamos a versão hospedada, e não a versão aberta do *software*, que pode ser customizada de modos infinitos.

O CMS funciona em interdependência com quase 60 mil *plugins* que podem ser acoplados, adicionando funcionalidades ao *WordPress.com*. Os dois *plugins* que selecionamos para usar nesta pesquisa funcionam como ferramentas (re)definidoras de práticas produtivas com a intenção de auxiliar produtores e editores de texto na criação e publicação de conteúdos mais elegíveis ao *Google*. Olhando mais criticamente, posicionam-se como “bússolas editoriais” e, por vezes, fazem “papel de linguistas”, indicando alterações textuais mais apropriadas para o ranqueamento.

Tanto o *Yoast SEO*, quanto o *Rank Math* trabalham com uma lógica metrificável ou gamificada (Zicherman; Cunningham, 2011) para escrita e edição de texto, pois vão indicando código de cores (verde/bom, laranja/ok e vermelho/ruim), notas, índices e

percentuais que os produtores/editores precisam alcançar em cada parâmetro. Isso envolve os criadores numa lógica de jogo (nesse caso, nada lúdico e nem divertido), na tentativa de motivá-los a escrever atendendo a requisitos de ranqueamento. No processo de simulação, todos os textos foram, primeiramente, submetidos ao *Yoast SEO* e a versão já editada passou pelo crivo do *Rank Math*, para que pudéssemos observar as sugestões de edição de forma complementar.

5.1.1 *Yoast SEO*

O *Yoast SEO*, lançado em 2010, é o *plugin* mais popular do *WordPress* e conta com cerca de 5 milhões de instalações ativas¹⁴⁷. Portanto, é conhecido como o *plugin* de SEO do *WordPress* “de fato”, como se representasse uma ferramenta original do CMS. Os recursos do *plugin* se dividem em dois pilares: análise de SEO On Page (palavra-chave, título, hipertextualidade, imagem) e análise de escaneabilidade e legibilidade (uso de palavras de transição e da voz passiva, a extensão de frases e parágrafos, a hierarquização do texto com divisão de intertítulos).

Os *feedbacks* dados pelo *Yoast SEO* dizem respeito a uma escrita mais otimizada para plataformas de busca, o que pode soar estranho, porque a maneira que cada produtor escreve é muito pessoal. O que o *plugin* faz é verificar, por meio de algoritmos, o conteúdo em diferentes fatores comprovadamente relacionados ao desempenho em mecanismos de pesquisa. Vale ressaltar que o *Google* e outros buscadores decidirão sobre a seleção e o ranqueamento, mas, ao otimizar os textos, é possível indicar aos motores de busca como um conteúdo pode aparecer. Ademais, as técnicas de SEO não são testes preparados por examinadores que definem todas as respostas certas, mas, sim, estratégias para enviar sinais para os mecanismos de pesquisa e ajudá-los a entender melhor um texto. Embora o algoritmo do *Google* permaneça secreto, mais de uma década de experimentos e movimentações da indústria de SEO resultou em uma boa ideia sobre os fatores importantes no ranqueamento.

A configuração de parâmetros SEO, por meio do *Yoast*, pode ser feita durante a produção e edição de texto dentro da plataforma *WordPress.com*. A ferramenta vai pontuando as ações do editor – no sentido de atender as demandas de visibilidade dos motores de busca – e apresentando o resultado da análise. A ideia é guiar o editor para que ele não precise trabalhar intuitivamente na organização e hierarquização textual, na seleção de palavras-

¹⁴⁷ Dados do próprio *plug-in*. Disponível em: <https://yoast.com>. Acesso em: 29 maio 2023.

chaves, na divisão de parágrafos, na construção de *links* internos e externos, na escaneabilidade, entre outros processos de edição. Todos esses itens são importantes marcadores para algoritmos de motores de busca durante o rastreamento, a indexação e o ranqueamento de conteúdos no ambiente digital.

Figura 24 – *Yoast SEO*



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

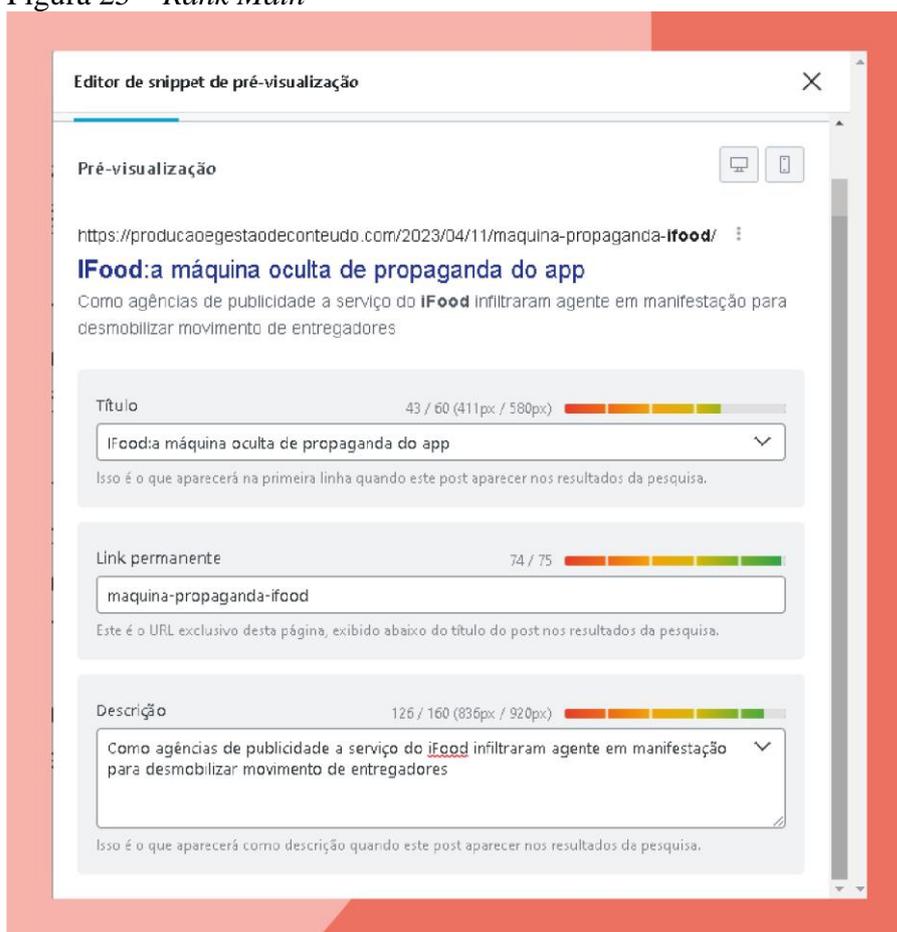
Com o texto inserido no rascunho do *WordPress.com*, o *plugin* aponta os bons resultados alcançados e os pontos a melhorar no texto sob a perspectiva da elegibilidade algorítmica. Parte desses parâmetros se baseia na palavra-chave (ou frase-chave) escolhida pelo autor do texto como a operadora do processo de atração – tanto de leitores, quanto de relevância pelos algoritmos. Por exemplo, ao solicitar a inserção de *links* internos, a ideia é criar autorreferências do *site* que indicam autoridade na produção de informação, fator considerado pelos motores de busca no ranqueamento. Ao sugerir o uso da frase-chave na introdução, a intenção é auxiliar na identificação temática feita pelos algoritmos nos textos, seguindo uma ordem de leitura que, geralmente, privilegia título, subtítulo, primeiro parágrafo e intertítulos. Enfim, o *plugin* se coloca como guia e poderá sugerir alterações de tamanho do texto, escolhas semânticas (como mudar a frase-chave), uso de conectivos (conjunções e pronomes), entre outras práticas.

5.1.2 *Rank Math*

O *Rank Math*, lançado em 2018, apesar de menos popular que o *Yoast SEO*, conquistou rapidamente o mundo dos *plugins* do *WordPress.com*. Atualmente, tem mais de 1

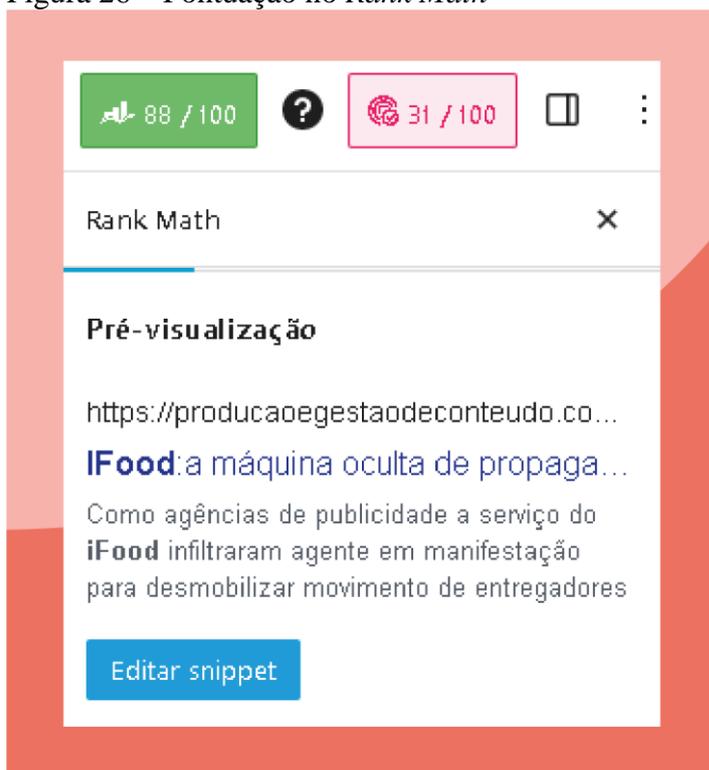
milhão de usuários ativos e se autoproclamou como o “canivete suíço” do *WordPress SEO*, por oferecer múltiplas funcionalidades. Os recursos do *plugin* se dividem nos mesmos pilares do *Yoast SEO*: análise de *SEO On Page* e análise de escaneabilidade e legibilidade, no entanto, com apresentações visuais diferentes. Os dois *plugins* usam IA para apoiar a edição de texto, oferecendo sugestões de quantas palavras, cabeçalhos, *links* e mídia o produtor pode usar com base no que está classificando atualmente para cada palavra-chave. Mas, vale dizer, que o *Rank Math* dá mais centralidade ao uso da IA como recurso.

Figura 25 – *Rank Math*



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

O processo de parametrização se dá de modo similar ao *Yoast SEO*, durante a escrita e edição de textos em rascunho no *WordPress.com*. A principal diferença está no modo de pontuar as melhorias sugeridas no texto, afinal, o *Rank Math* não trabalha como sistema de cores do *Yoast SEO* (verde/bom, laranja/ok e vermelho/ruim), mas determina notas, como é possível ver na Figura 26.

Figura 26 – Pontuação no *Rank Math*

Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

5.1.3 Seleção do *corpus*

No decorrer da tese, defendemos que a lógica algorítmica que determina visibilidade de conteúdos em plataformas de busca atravessa as rotinas de produção e edição de várias áreas da escrita, quando produtores e editores têm a intenção de performar. Por isso, definimos como *corpus* de pesquisa cinco textos que representam diferentes modos de produção, objetivos, públicos, gêneros textuais. Essa diversidade nos ajudou a verificar se a incorporação de práticas com vistas à elegibilidade algorítmica pode ocorrer em muitas rotinas produtivas. Certamente, em suas publicações originais, o nível de intenção de performar desses textos é diferente, porque há autores/sites/veículos mais ou menos preocupados em ter visibilidade em ambiente digital. O jornalismo, por exemplo, se preocupa ostensivamente com a visibilidade porque isso sustenta seu modelo de negócio. O que não necessariamente é o caso da literatura. Veremos que o texto jornalístico pode ser concebido e editado para o ranqueamento e uma crônica literária pode, simplesmente, ser publicada sem qualquer parametrização para elegibilidade algorítmica. Se pensarmos na perspectiva de gêneros textuais, faz pouco sentido que uma crônica seja alterada em suas essências criativas para atender a parâmetros de ranqueamento em plataforma de busca, mas essas questões vão emergir nas análises que faremos.

Quadro 10 – *Corpus etapa 2*

Número	Títulos
Texto 1	A máquina oculta de propaganda do iFood ¹⁴⁸
Texto 2	De um tudo (1) ¹⁴⁹
Texto 3	Açaí, tucumã e buriti são os insumos da Amazônia com potencial de inovação mais estudados por pesquisadores brasileiros ¹⁵⁰
Texto 4	Saiba tudo sobre prospecção de clientes e os métodos mais eficazes ¹⁵¹
Texto 5	Um próximo passo importante na nossa IA ¹⁵²

Fonte: Elaborado pela autora, 2023.

Os textos selecionados já estão publicados em seus *sites* originais, sendo quatro deles produzidos dentro do CMS *WordPress*, o que nos motivou a incluí-los no *corpus*. Tentamos fazer nosso espaço simulado parecido com o espaço de escrita em que os conteúdos foram originalmente criados. Para descobrir se um *site* é feito no *WordPress*, basta analisar o código-fonte das páginas em busca de identificações como “/wp-content/”, ação que empreendemos em todos os textos candidatos a entrar no *corpus*. Nessa mesma análise de código-fonte, é possível descobrir se os produtores publicaram com auxílio dos *plugins Yoast SEO* e *Rank Math* e alguns dos textos selecionados estavam, sim, parametrizados com a ajuda dessas ferramentas.

Figura 27 – Código-Fonte indicando uso do *WordPress.com*



```

<!DOCTYPE html><html lang="pt-BR" class="no-js">
  <head>
    <meta charset="UTF-8">
    <!--
      <meta http-equiv="content-type" content="text/html; charset=utf-8" />
    -->
    <meta http-equiv="x-ua-compatible" content="ie=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <link rel="profile" href="http://mpg.org/xfn/11">
    <link rel="shortcut icon" type="image/png" href="https://apublica.org/wp-content/uploads/2018/02/favicon3.png"/>
  
```

Fonte: Disponível em: <https://apublica.org/>. Acesso em: 29 maio 2023.

O texto 1 – *A máquina oculta de propaganda do iFood* – é um conteúdo jornalístico escrito por Clarissa Levy e publicado pela *Agência Pública*. É narrada a estratégia

¹⁴⁸ Disponível em: <https://apublica.org/2022/04/a-maquina-oculta-de-propaganda-do-ifood/#Agente>. Acesso em: 29 maio 2023.

¹⁴⁹ Disponível em: <https://rascunho.com.br/liberado/de-um-tudo-1/>. Acesso em: 29 maio 2023.

¹⁵⁰ Disponível em: <https://abori.com.br/amazonia/acai-tucuma-e-buriti-sao-os-insumos-da-amazonia-com-potencial-de-inovacao-mais-estudados-por-pesquisadores-brasileiros/>. Acesso em: 29 maio 2023.

¹⁵¹ Disponível em: <https://www.conversion.com.br/blog/prospeccao-de-clientes/>. Acesso em: 29 maio 2023.

¹⁵² Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/um-proximo-passo-importante-na-nossa-jornada-de-ia/>. Acesso em: 29 maio 2023.

de *marketing* da plataforma de *delivery* em contra-ataque à mobilização de motociclistas que se uniram em prol de condições menos precarizadas de trabalho. O texto é denso e com muitas possibilidades de edição, o que deu potencial para nossas análises. A temática é adjacente às discussões sobre plataformização da nossa pesquisa, o que nos motivou a juntar esse texto ao *corpus*. Este é um dos textos já publicados com o apoio do *plugin Yoast SEO*, usado pela *Agência Pública* no *WordPress* e foi interessante submetê-lo novamente ao crivo do *plugin*.

Figura 28 – Código-Fonte indicando uso do *Yoast SEO*

```
<!-- This site is optimized with the Yoast SEO plugin v20.4 - https://yoast.com/wordpress/plugins/seo/ -->
<title>A máquina oculta de propaganda do iFood - Agência Pública</title>
<meta name="description" content="Agências de publicidade a serviço do iFood criaram perfis falsos em redes sociais e infiltraram
<link rel="canonical" href="https://apublica.org/2022/04/a-maquina-oculta-de-propaganda-do-ifood/" />
<meta property="og:locale" content="pt_BR" />
<meta property="og:type" content="article" />
<meta property="og:title" content="A máquina oculta de propaganda do iFood - Agência Pública" />
<meta property="og:description" content="Agências de publicidade a serviço do iFood criaram perfis falsos em redes sociais e infil
<meta property="og:url" content="https://apublica.org/2022/04/a-maquina-oculta-de-propaganda-do-ifood/" />
<meta property="og:site_name" content="Agência Pública" />
<meta property="article:publisher" content="https://www.facebook.com/agenciapublica" />
<meta property="article:published_time" content="2022-04-04T09:00:00+00:00" />
<meta property="article:modified_time" content="2023-03-28T19:42:57+00:00" />
<meta property="og:image" content="https://apublica.org/wp-content/uploads/2022/03/capa-facebook-a-maquina-oculta-de-propaganda-dc
<meta property="og:image:width" content="1200" />
<meta property="og:image:height" content="630" />
<meta property="og:image:type" content="image/png" />
<meta name="author" content="Clarissa Levy" />
<meta name="twitter:card" content="summary_large_image" />
<meta name="twitter:image" content="https://i0.wp.com/apublica.org/wp-content/uploads/2022/03/capa-twitter-a-maquina-oculta-de-pro
<meta name="twitter:creator" content="@clalevy_" />
<meta name="twitter:site" content="@agenciapublica" />
<meta name="twitter:label1" content="Escrito por" />
<meta name="twitter:data1" content="Clarissa Levy" />
<meta name="twitter:label2" content="Est. tempo de leitura" />
<meta name="twitter:data2" content="39 minutos" />
<script type="application/ld+json" class="yoast-schema-graph">{"@context":"https://schema.org","@graph":[{"@type":"Article","@id":
<!-- / Yoast SEO plugin. -->
```

Fonte: Disponível em: <https://apublica.org/>. Acesso em: 29 maio 2023.

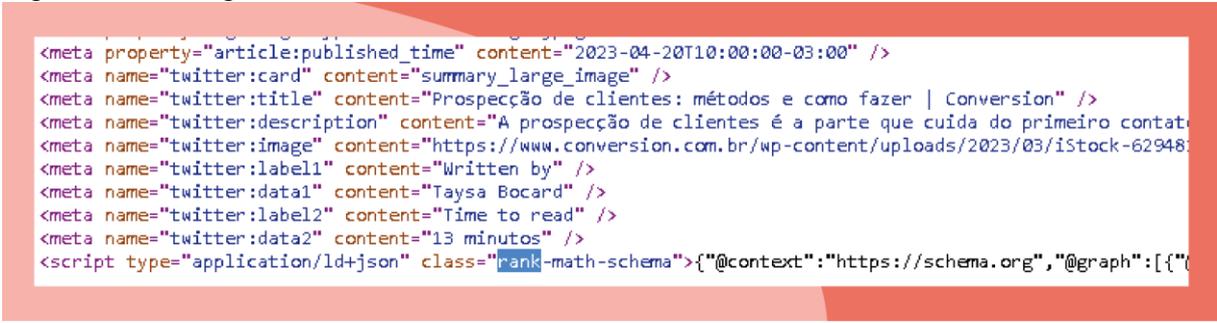
O texto 2 – *De um tudo (1)* – é uma crônica de autoria da Ana Elisa Ribeiro e publicada na versão *online* do jornal *Rascunho*, especializado em literatura brasileira e estrangeira. A coluna explora, com uma escrita refinada e tom irônico, o potencial do *YouTube* como plataforma de busca e de vídeo. Portanto, é também uma temática que dialoga com os assuntos da nossa pesquisa, o que nos seduziu a inseri-lo no *corpus*. Ademais, pressupondo objetivos de alcance e de públicos bem diferente da *Agência Pública*, por ser o *Rascunho* um jornal de nicho, consideramos importante trazer essa diversidade. Também nos instigou a investigar um texto curto, com apenas uma imagem inserida, mais literário e sem finalidade, necessariamente, informativa.

O texto 3 – *Açaí, tucumã e buriti são os insumos da Amazônia com potencial de inovação mais estudados por pesquisadores brasileiros* – é um conteúdo de divulgação científica. Ele está publicado na *Agência Bori*, um *hub* para jornalistas encontrarem conteúdos sobre pesquisas científicas inéditas e que auxilia na cobertura dos temas de ciência. A matéria

tem caráter de *release*, portanto, muito diferente dos textos 1 e 2, o que nos motivou pela diversidade. Este é um dos textos já publicados com o apoio do *plugin Yoast SEO*, usado pela *Agência Bori* no *WordPress*, dessa forma, propusemos uma nova otimização.

O texto 4 – *Saiba tudo sobre prospecção de clientes e os métodos mais eficazes* – é um exemplo de *Marketing* de conteúdo. Ele foi escrito por Taysa Bocard e publicado no *blog* da *Conversion*, uma agência especializada em SEO e *marketing* digital. O conteúdo é sobre prospectar clientes em ambiente digital usando técnicas e metodologias de *marketing* digital. Portanto, instigou-nos a incluir no *corpus* por ser um texto pensado para dar visibilidade e tornar a empresa uma referência no tema de SEO, além de perpassar temas centrais da nossa pesquisa. Diferentes dos outros *sites* em que selecionamos textos, o *blog* da *Conversion* usa o *plugin Rank Math* como ferramenta para otimização de textos.

Figura 29 – Código fonte indicando uso do *Rank Math*



```

<meta property="article:published_time" content="2023-04-20T10:00:00-03:00" />
<meta name="twitter:card" content="summary_large_image" />
<meta name="twitter:title" content="Prospecção de clientes: métodos e como fazer | Conversion" />
<meta name="twitter:description" content="A prospecção de clientes é a parte que cuida do primeiro contato" />
<meta name="twitter:image" content="https://www.conversion.com.br/wp-content/uploads/2023/03/istock-62948" />
<meta name="twitter:label1" content="Written by" />
<meta name="twitter:data1" content="Taysa Bocard" />
<meta name="twitter:label2" content="Time to read" />
<meta name="twitter:data2" content="13 minutos" />
<script type="application/ld+json" class="rank-math-schema">{"@context":"https://schema.org", "@graph": [{"t

```

Fonte: Disponível em: <https://www.conversion.com.br/>. Acesso em: 29 maio 2023.

O texto 5 – *Um próximo passo importante na nossa IA* – é de Sundar Pichai, CEO da *Google/Alphabet Inc.*, e está publicado no *blog* do *Google Brasil*. Há pelo menos três motivos para incluí-lo no *corpus*: 1) ser um texto de *blog* institucional de uma grande empresa de tecnologia, que dá a ele a função de conteúdo da marca e informação primária para assuntos mais recentes da área (no caso, IA); 2) ser um texto do *Google*, que é a plataforma em foco em nosso estudo; e 3) tratar de temáticas adjacentes à nossa pesquisa, afinal, é um conteúdo para apresentar o serviço experimental de IA de conversação do *Google* com tecnologia LaMDA (modelo de linguagem neural conversacional). Este é o único texto não publicado a partir do *WordPress.com*, o que nos faz crer que o *Google* tem um CMS próprio.

5.1.4 Categorias de análise

As categorias emergiram da observação do *corpus* estudado e da articulação de toda teoria que fundamenta nossa pesquisa. As diretrizes do *Google*, analisadas na etapa

documental, também nos orientaram a propor as categorias de análise. Estas, por sua vez, foram organizadas durante os testes de inserção dos textos nos *plugins Yoast SEO* e *Rank Math*. Ao todo, são cinco categorias e, dentro de uma delas, oito subcategorias, como apontamos no Quadro 10. Todos os parâmetros indicados pelos *plugins* são baseados na leitura do código HTML de páginas de conteúdo, da mesma forma que fazem os algoritmos do *Google*. Portanto, alguns dos elementos que analisaremos têm uma nomenclatura para produtores e editores de texto que trabalham com CMS – na camada de edição de texto –, mas levam nomes diferentes no código HTML – que é a camada constitutiva da página.

Quadro 9 – Categorias de análise

Nomenclatura para produtores de texto		Elementos e atributos HTML
Categoria 1	Palavra-Chave	“keywords”
Categoria 2	Título, metadescrição e URL	<title>, “description”
Categoria 3	Hipertextualidade (<i>links</i>)	<link>
Categoria 4	Legibilidade e escaneabilidade	Tamanho do texto; Tamanho do parágrafo; Frases consecutivas; Palavras de transição; Tamanho das sentenças; Voz Passiva; Hierarquia do texto – distribuição de intertítulos ou cabeçalhos (<i>headings</i>)
Categoria 5	Imagem	

Fonte: Elaborado pela autora, 2023.

Muitas dessas categorias fizeram grande sentido ao relacionarmos com diretrizes e métricas de relevância descritas na seção 4.4. Por exemplo, a diretriz “Boa experiência de uso” aparece na categoria “Legibilidade e escaneabilidade”; e a diretriz “Links rastreáveis, títulos descritivos e úteis” é correlata da categoria “Título, metadescrição e URL”. Sendo assim, conseguimos, ainda no processo de categorização, perceber conexões entre as regras de

funcionamento do *Google* com os modos como a escrita se desencadeia no CMS (no caso, *WordPress.com*).

5.2 RECONFIGURAÇÃO DE PRÁTICAS: ANÁLISES E DISCUSSÕES

Ao iniciar a simulação com cada texto do *corpus*, observamos que as recomendações de reescrita, adaptações e revisões foram completamente distintas e encaminharam as decisões de edição de modo muito particular. Cada alteração sugerida pelos *plugins Yoast SEO* e *Rank Math* se converte em uma alteração de código HTML para otimizar um texto, sinal de que os processos de escrita e edição à luz de lógicas algorítmicas são emergentes e recorrentes nos modos de construir texto na atualidade. Quando especialistas em SEO e produção de conteúdo indicam que o *Core Update BERT* muda a forma como se busca na internet, estão também dizendo sobre as formas de escrita que acabam por se reconfigurar às demandas de elegibilidade para o *Google*.

Sendo assim, a edição de textos que subjaz às mediações algorítmicas dessa plataforma de busca é permeável por mudanças na lógica de rastreamento, indexação e, principalmente, ranqueamento. Referindo-se à produção de livros, Medeiros (2009) afirma que a edição surge de modo contingente, plural, aberto e situado historicamente, e, se estamos vivenciando um momento único – de cultura digital pós-plataformas –, vamos nos inspirar nessa ideia do autor para verificar que rupturas e recorrências nas práticas de produção de texto são inevitáveis. Leal (2018) corrobora essa ideia ao refletir que o texto não é simplesmente um produto, resultado final de uma prática sociodiscursiva historicamente situada, mas algo que emerge em seu desenrolar, como veremos na análise de cada categoria a seguir.

5.2.1 Palavra-Chave

Nesta categoria, analisamos a percepção dos *plugins Yoast SEO* e *Rank Math* sobre a inserção e uso estratégico de palavras-chave nos cinco textos selecionados. Essa é a principal parametrização apontada pelos *plugins Yoast SEO* e *Rank Math*, que a consideram fundamental para que as plataformas de busca reconheçam do que se trata um texto. O *Google* é capaz de reconhecer as palavras juntas ou separadas mesmo que digitemos termos longos na barra de pesquisa. Os *plugins* tentam, ao analisar os textos, imitar o *Google*, indicando ao produtor de texto a pertinência do termo escolhido para ser o foco do conteúdo.

Portanto, a palavra-chave ou frase-chave em foco é o termo de pesquisa para o qual um produtor de texto deseja que uma página ou uma postagem sejam mais classificadas. No código HTML, algoritmos vão enxergar o atributo “keywords” para descobrir qual é a palavra-chave em foco na página. Se o produtor ou editor de texto definir o termo-foco, os *plugins Yoast SEO* e *Rank Math* avaliarão o conteúdo da página e fornecerão *feedback* sobre como melhorar o texto para aumentar suas chances de classificação para esse termo de pesquisa.

Baseado na noção de cauda longa (ANDERSON, 2004), criou-se um espectro de palavra-chave *head tail*, referente às mais genéricas e com alto volume de buscas e palavra-chave *long tail*, que se refere às mais específicas, com menor volume de busca, mas que se aproximam mais com a forma como os leitores fariam a pesquisa no buscador. Por exemplo, uma palavra-chave *head tail* poderia ser “tênis nike” e a versão *long tail*, “tênis nike michael jordan”. No primeiro caso, por ser uma combinação de termos mais comum, muitas pessoas podem fazer essa busca, mas haverá muitos *sites* (de *e-commerce* ou não) otimizando textos e postagens para essa combinação de termos, portanto, haverá maior concorrência na classificação do *Google*. No segundo caso, por ser uma combinação mais específica, há menor concorrência, porém há menos pessoas buscando. Portanto, existe maior chance de o *Google* selecionar e classificar textos e postagens que atendam as necessidades específicas de leitores.

Durante as simulações com o *corpus*, verificamos a importância dada à palavra-chave, porque, nos cinco textos, o ato de escolher frases-chave, inserindo-as em campo adequado, alterava completamente o *status* de análise da ranqueabilidade do conteúdo. Pelo *Rank Math*, no texto 4, começamos com nota 21/100¹⁵³ e somente a inserção da palavra-chave “prospecção de clientes” (*head tail*) nos fez saltar para 75/100, algo muito representativo, que indica grande peso desse parâmetro. Pelo *Yoast SEO*, no texto 1, somente a inserção da palavra-chave “ifood” (*head tail*) alterou a condição de vermelho/ruim para laranja/ok. No texto 3, começamos a edição inserindo a frase-chave “açai tucumã castanha” (*long tail*) e, como não ocorreu mudança de *status*, alteramos para apenas “açai” (*head tail*), que melhorou a condição de vermelho/ruim para laranja/ok. Nesse caso, percebemos que o termo “açai” sozinho funcionou bem, por ser mais popular e ter maior potencial como termo de pesquisa em plataformas de busca. Mesmo sendo os outros dois insumos – tucumã e castanha – importantes para o conteúdo principal, eles ficaram taticamente de fora da otimização.

¹⁵³ Lemos 21 pontos em 100, onde 100 é o índice máximo de otimização do texto para que seja mais ranqueável.

Figura 30 – Análise de frase-chave no texto 1 pelo *Yoast SEO*



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

Ainda sobre usos táticos de palavra-chave, para escolher o termo foco para o texto 1, fizemos o teste usando “ifood” (*head tail*), que, apesar de menos específica, representa um nome próprio ou marca e pode obter um volume de buscas representativo. A versão sem edições da reportagem tinha, segundo contabilização do *Yoast SEO*, 43 repetições de “ifood”, o que configura a diretriz de espalhar naturalmente os termos no decorrer do texto, atribuindo “densidade de termo-chave” (o quanto a palavra-chave aparece). Tentamos outras versões mais específicas, como “ifood propaganda”, “ifood movimento de entregadores”, “publicidade ifood”, “ifood marketing”. Consultamos a ferramenta *Google Trends* para ter ideias de combinações como “ifood entregador”, que tem bom volume de busca, mas contribuiu pouco com a otimização da reportagem, porque não ajudou com a densidade de termo-chave, afinal essa exata combinação não se repete no decorrer do texto.

Sobre densidade de palavras-chave, observamos diferenças nas análises dos dois *plugins*, inclusive, porque o *Rank Math* metrifica esse parâmetro matematicamente. Por exemplo, o texto 2, analisado pelo *Yoast SEO* com a inserção do termo-foco “youtube”, foi considerado adequado, até mesmo na densidade (apareceu sete vezes no conteúdo). Pelo *Rank Math*, essa quantidade não foi considerada suficiente: “A densidade da palavra-chave é 0,89%, pois o termo aparece 7 vezes. Ideal que fique entre 1 e 1,5”, apontou a análise. Aliás, observamos esse aviso do *Rank Math* em pelo menos outros dois textos, mostrando que esse *plugin* contabiliza densidade a partir de alguma fórmula que exija maior quantidade de palavras-chave distribuídas no texto. Para o texto 4, ao inserir a palavra-chave “prospecção de

clientes”, a análise do *Rank Match* foi de densidade 0,47, quando o ideal é ficar entre 1 e 1,5. Para esse mesmo texto 4, o *Yoast SEO* avaliou como verde/bom a densidade de termo-chave.

Essa aparente falta de consenso entre as quantificações apontadas pelos *plugins Yoast SEO* e *Rank Math* é mais uma prova de que não se conhece exatamente as métricas ideais de presença de palavra-chave na otimização de um texto para que ele seja mais ranqueável. Essa afirmação nos remete a vários momentos do Capítulo 4 em que apontamos subjetividades, intangibilidades e interpretabilidades nas diretrizes do *Google*, por isso defendemos uma construção mitológica da relevância. Se os *plugins* emulam o *Google* e não há conhecimento exato dos critérios, não existe, também, exatidão na emulação. O que podemos observar são os entrelaçamentos entre o que conhecemos de critérios do *Google* com as práticas de escrita e edição que passam a fazer parte da rotina dos produtores.

Verificamos que a centralidade da palavra-chave no processo de produção e circulação de textos em ambiente digital parece ter um peso maior para alguns gêneros textuais do que em outros. Ao analisar os textos 1, 2, 3, e 5, não há indícios de que tenham sido criados a partir de termos-chave, pelo contrário, são fruto criativo de temas interessantes para quem os publica ou importantes como função social, conforme a reportagem jornalística da *Agência Pública* (texto 1). Isso não significa, por exemplo, que *sites* como *Agência Pública* e *Rascunho* não façam estratégias de pautar ou *brifar*¹⁵⁴ produtores de texto a partir de “trends” de busca. Todavia, nos textos que compunham o *corpus*, não foi possível ver indícios. Já no texto 4, que tem objetivo de *marketing*, notamos pistas de que ele tenha nascido a partir da palavra-chave “prospecção de cliente”. O padrão de URLs do *site* indica que há uma lista de temas, capitaneados por palavras-chave para a produção de conteúdo. A URL do texto 4 era <https://www.conversion.com.br/blog/prospeccao-de-clientes/>, em que o termo-chave compõe o endereço, assim como em outros exemplos do mesmo *site*: <https://www.conversion.com.br/blog/inboundmarketing/>; <https://www.conversion.com.br/blog/marketing-de-conteudo/> e <https://www.conversion.com.br/landing-pages/>. Outra pista é a integração desse conteúdo sobre prospecção de clientes com outros textos *linkados* dentro do *post*, configurando-se como um episódio de uma série de conteúdos construídos para tornar a *Conversion* referência no mercado. Conhecendo a dinâmica de agências e *marketing*, essa é uma prática comum: o mapeamento de palavras-chave para montar estratégias e agendas de conteúdos a serem produzidos nos *blogs*. Ao apontar essa especificidade produtiva, não fazemos juízo de valor, mas observamos o quanto a lógica de ranqueamento permeou o

¹⁵⁴ O termo *brifar* vem de *briefing*, que são instruções objetivas dadas a um produtor de textos sobre a tarefa a ser executada. É uma expressão muito usada no mercado de produção de conteúdo para marcas.

processo de produção textual, a ponto de ser premissa para uma criação. Nesse sentido, também resumimos, no Quadro 11, a relação entre mediações algorítmicas e reconfiguração das práticas de escrita para a categoria 1 – palavra-chave:

Quadro 102 – Entrelaçamento entre plataforma de busca e escrita na *Web* (categoria 1 – palavra-chave)

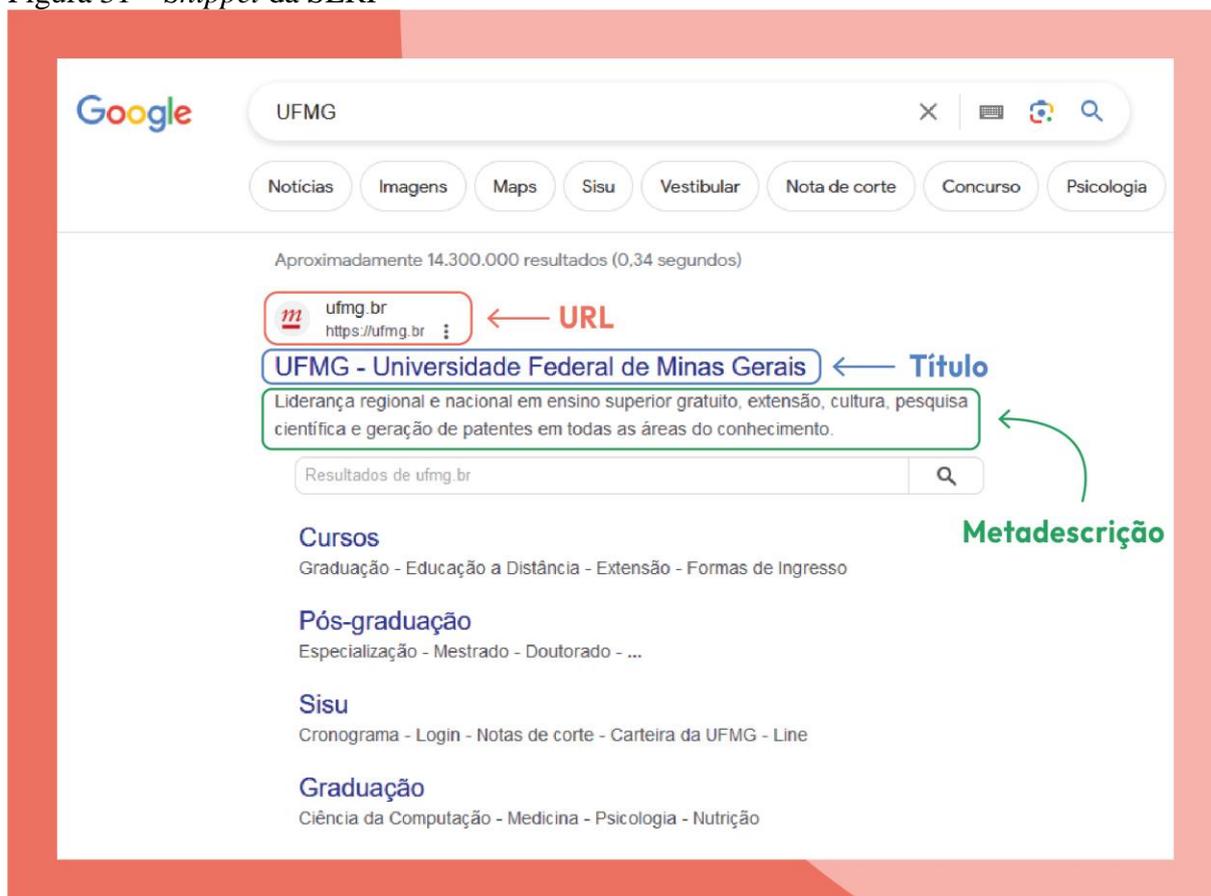
Mediações algorítmicas e/ou diretrizes do Google	Práticas incorporadas na escrita e edição
A frase-chave – ou atributo “keyword” – será usada pelos <i>webcrawlers</i> para percorrer páginas buscando correspondências textuais.	Escolher e usar estrategicamente a palavra-chave, além de inseri-la naturalmente a palavra-chave no decorrer do texto, atribuindo densidade.
Algoritmos agrupam as páginas encontradas que têm conteúdo semelhante.	Usar palavras-chave em títulos, subtítulos, primeiro parágrafo, intertítulos (<i>headings</i>) de conteúdos e páginas.
Indicações de ferramentas de <i>Analytics</i> sobre o processo de leitura-escritura mostrando quais palavras-chave foram usadas para chegar a um texto.	Submeter texto a sistemas de análise de palavras-chave, a exemplo do <i>Google Trends</i> ou <i>Keyword Planner</i> ; desenvolver estratégias e planejamento de pautas e <i>briefing</i> no entorno das palavras-chave.
Algoritmos capazes de compreender, prever precisamente, comparar e associar e responder com elevado nível de confiança a conjuntos de texto como BERT, que atua de modo diferenciado em consultas mais complexas, em especial, para busca com frases-chave de cauda longa.	Planejar textos na lógica da engenharia reversa. É produzir textos para os quais já sabemos existirem demandas de busca e fazer títulos, subtítulos exatamente iguais aos termos usados frequentemente para barra do <i>Google</i> .

Fonte: Elaborado pela autora, 2023.

5.2.2 Título, meta descrição e URL

Nesta categoria, analisamos a percepção dos *plugins Yoast SEO* e *Rank Math* sobre os outros elementos, além da palavra-chave, que vão fazer uma página ou conteúdo ser reconhecido pelo *Google*. Título, metadescrição e URL são fundamentais na construção de métricas de relevância da plataforma, porque são os primeiros aspectos visualizados pelo leitor no *snippet*¹⁵⁵ da SERP, como a Figura 31 exemplifica. Além disso, no HTML, o marcador de título <*title*> compõe um grupo de elementos que guardam informações essenciais da página, conhecidos como metadados. Daí a importância desse elemento durante a leitura algorítmica.

¹⁵⁵ No contexto de SEO, *snippet* é o bloco em destaque que inclui um resumo da resposta, extraído de uma página da *Web*, além de um *link*, um título da página e o URL.

Figura 31 – *Snippet* da SERP

Fonte: Google.com, 2023.

O título SEO *<title>* é diferente do título do conteúdo, que os leitores visualizam ao acessar a página com um texto. Ele é feito para os “olhos dos algoritmos” de plataformas de busca, afinal ele funciona como marcador no HTML de uma página, e tem como objetivo que as pessoas cliquem e visitem o *site*. Por isso, é considerado um elemento essencial na escrita e edição, com vistas à visibilidade, porque deve ajudar uma página a classificar uma palavra-chave e tem influência direta no ranqueamento. Assim, reforça-se a importância de a palavra-chave estar presente no título SEO e atrair cliques que, por sua vez, podem levar o *Google* a compreender a relevância do *site*, posicionando-o bem, ou entender que ele deve ser posicionado mais abaixo nos resultados de pesquisa. O *Google* notará se as pessoas clicarem muito em um título, o que alimenta a cultura caça-cliques (CHRISTOFOLETTI; VIEIRA, 2015) e os títulos testáveis (BUENO; REINO, 2019), muito comuns no jornalismo e no *marketing* digital.

Os *plugins Yoast SEO* e *Rank Math* verificam aspectos cruciais dos títulos, como largura e a presença da palavra-chave, para indicá-lo como apropriado ou não, segundo as parametrizações de performance em mecanismos de busca. O *Rank Math*, especificamente,

faz uma análise de legibilidade para o título e sugere frequentemente a inserção de números no título. De acordo com o tutorial¹⁵⁶ do *plugin*, os títulos com números soam mais clicáveis. O exemplo dado no tutorial é: “Como limpar sua garagem?” e “7 maneiras fáceis de limpar sua garagem”, sendo esta segunda opção apontada como a mais clicável. Na análise do *corpus*, ficou completamente inviável editar os títulos inserindo números, porque não faria qualquer sentido para os textos escolhidos. Acreditamos que essa tática faria sentido no planejamento de uma pauta ou *briefing*, antes do processo de escrita. Da forma como simulamos, com textos já produzidos, seria muito difícil acatar a sugestão do *Rank Math*, haja vista nossa ideia de interferir pouco nos textos. Na seção 5.3, olharemos mais criticamente para essa sugestão como uma forma padronizadora exacerbada, que coloca os textos em *templates* para performance no *Google*.

A metadescrição `<description>` é também uma marcação no HTML de uma página, que é optativa em uma postagem, mas que nenhum produtor em busca de visibilidade deixará de criar. Isso porque esses caracteres descritivos dão ao leitor uma noção mais expandida sobre o conteúdo ou sobre a página. O próprio nome “meta” dá essa noção de descrição em nível superior, que reflete o conteúdo e serve como guarda-chuva de sentido. A metadescrição fica posicionada abaixo do título no *snippet* da SERP e, assim como o título, existe para convencer o leitor a clicar e gerar tráfego para os *sites*. O *Google* usa as taxas de cliques de maneira estatística para determinar se certos *sites* são “boas entregas” para os leitores, o que reforça a importância de títulos e metadescrições apropriadas. Os *plugins* ajudam os produtores a parametrizar as metadescrições e, geralmente, sugerem que tenham até 155 caracteres (o espaço dado pelo *Google* no *snippet* da SERP); voz ativa, verbos de comando; chamadas para ação; palavra-chave (por mais que isso pareça repetitivo no título); unicidade e originalidade.

Os URLs são os endereços das páginas e itens também usados por *crawlers* de plataformas de busca para selecionar *sites*. Por isso, a parametrização que os *plugins* vão sugerir está baseada na lógica algorítmica dos buscadores. Embora URLs curtos sejam preferidos, não há diretrizes para dizer quais URLs são considerados “curtos”, portanto, não é fácil quantificar o comprimento, mas esses endereços devem ter precisão e objetividade, além de trazer a palavra-chave foco. Criadores do *Rank Math*, por exemplo, definiram, após testes, que um URL apropriado é menor ou igual a 75 caracteres, sendo assim, o *plugin* contabiliza para o produtor de textos esse limite numérico.

¹⁵⁶ Disponível em: <https://yoast.com/tag/content-seo/>. Acesso em: 02 ago. 2023.

Assim como na análise da Categoria 1 – palavras-chave –, ao olharmos para a categoria 2, verificamos a importância de título, metadescrição e URL para a performance de textos nas plataformas de busca. Por exemplo, pelo *Rank Math*, no texto 2, iniciamos a simulação com nota 14/100 e a alteração de palavra-chave, título, metadescrição e URL elevaram a nota para 86/100, um salto muito representativo de nota, indicando grande peso desses parâmetros.

Nas escolhas de título, podemos verificar, na prática, como produtores e editores de texto trabalham para atender a duas instâncias de relevância, como explicitamos na seção 4.4: relevância para públicos, com títulos do conteúdo que atendam a um sistema de expectativas de leitores e façam sentido na linha editorial da organização publicadora; e relevância para algoritmos, com títulos SEO que atendam critérios de elegibilidade das plataformas de busca. Vejamos os exemplos:

Quadro 113 – Titulação: relevância para leitores *versus* relevância para algoritmos

	Título do conteúdo	Título SEO
Texto 1	A máquina oculta de propaganda do <i>iFood</i>	<i>iFood</i> : a máquina oculta de propaganda do app
Texto 2	De um tudo (1)	Qualquer coisa está no <i>YouTube</i> e isso pode nos salvar a vida
Texto 3	Açaí, tucumã e buriti são os insumos da Amazônia com potencial de inovação mais estudados por pesquisadores brasileiros	Açaí, tucumã e buriti são os insumos mais estudados por cientistas

Fonte: Elaborado pela autora, 2023.

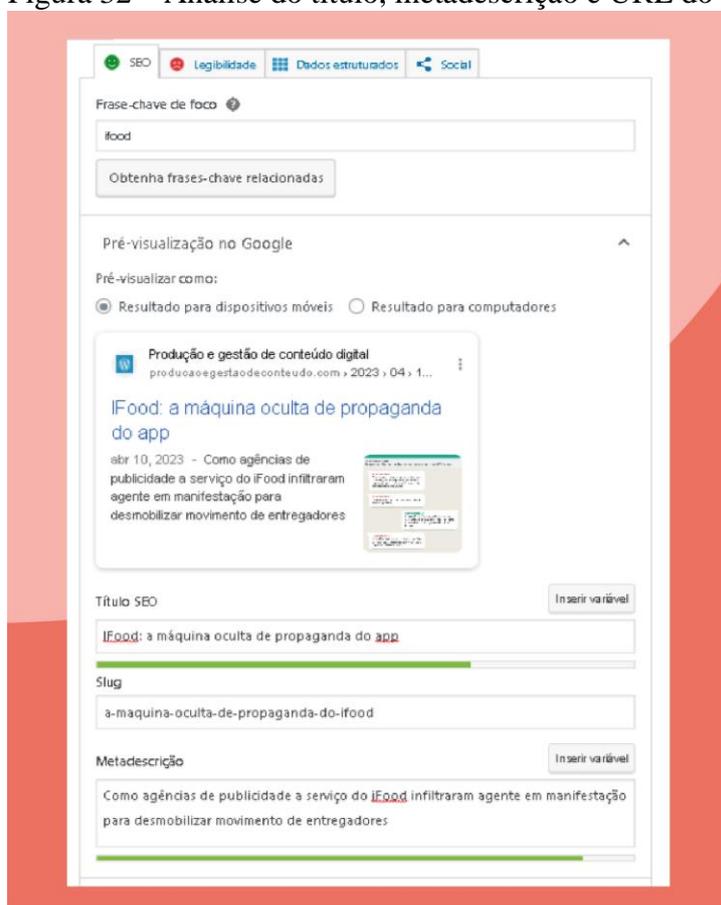
Por que ocorre essa diferença? É justamente porque os *plugins Yoast SEO* e *Rank Math* vão indicar a parametrização do título SEO, aquele encontrado pelos buscadores e que, portanto, precisa atender os parâmetros de ranqueabilidade. Pelo *Yoast SEO*, no texto 1, tentamos usar o título original da reportagem como título SEO, mas o *plugin* indicou que a palavra-chave, “ifood”, não estava no início da sentença, um aspecto importante para o rastreamento dos motores de busca.

No texto 2, o *Yoast SEO* considerou apropriada a escolha da palavra-chave “YouTube”, no entanto, a inserção do termo-foco não alterou muito o *status* da avaliação, quando percebemos que seria preciso mudar o título do conteúdo. Entre todos os conteúdos do *corpus*, o texto 2 era aquele com título menos otimizado, aliás, priorizando um título não literal justamente pela característica do jornal *Rascunho* e do gênero proposto pela autora Ana Elisa Ribeiro. A escolha de um título não literal, metafórico ou poético nos parece bem

adequado para o meio e públicos presumidos, no entanto, não atende aos critérios algorítmicos. Os títulos dessa natureza, geralmente, não trazem a palavra-chave e não são apropriados para a lógica do ranqueamento, por isso a diferença entre os títulos no texto 2 é tão marcante.

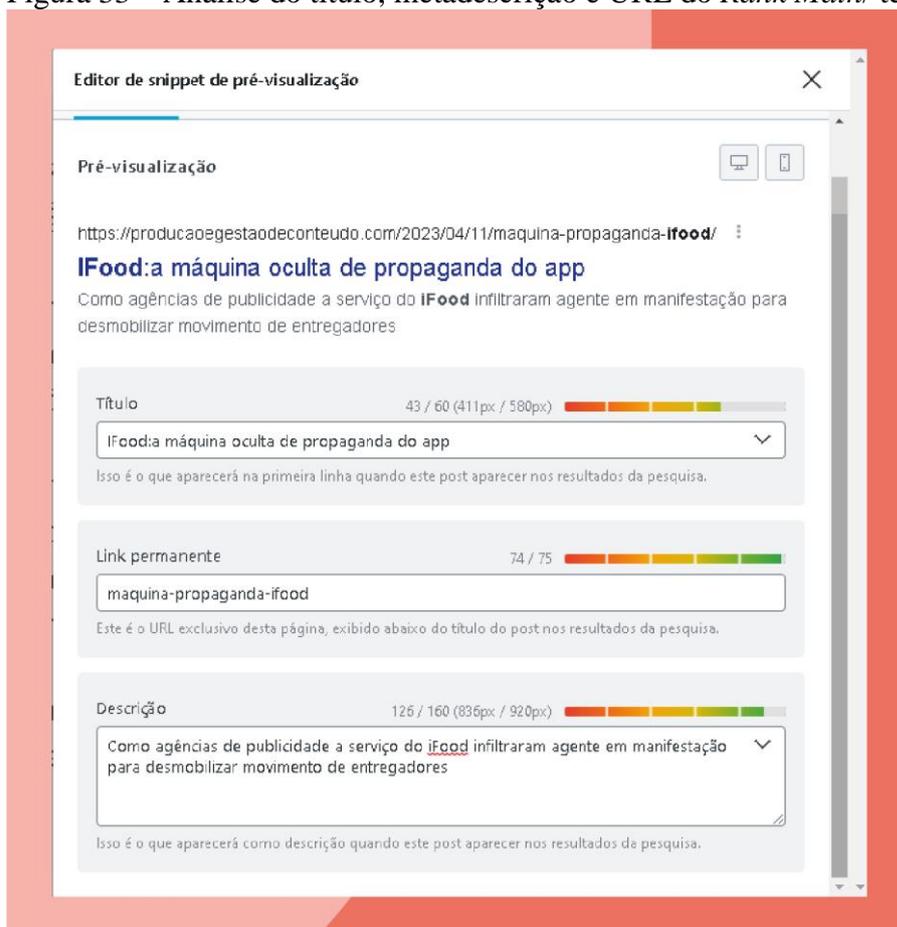
Analisando o conjunto – título, metadescrição e URL – para os textos 1, 2, e 3, o que há de mais importante no apontamento dos *plugins* são os tamanhos das sentenças escolhidas. No texto 1, a primeira metadescrição que criamos excedeu 155 caracteres e isso não garantiria a visibilidade dela na SERP, portanto, o *Yoast SEO* sugeriu a redução. Já o *Rank Math* é mais criterioso com o tamanho do URL. O primeiro URL que criamos para o texto apresentava 89 caracteres, mas foi sugerida a redução para 75, como vemos ao comparar a Figura 32 com a Figura 33. A diferença está apenas no endereço: “a-maquina-culta-de-propaganda-do-ifood” (*slug*), no *Yoast*, e “maquina-propaganda-ifood” (*link* permanente), no *Rank Math*.

Figura 32 – Análise do título, metadescrição e URL do *Yoast SEO*/ texto 1



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

Figura 33 – Análise do título, metadescrição e URL do *Rank Math*/ texto 1



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

No texto 2, o *Rank Math* também foi criterioso com o tamanho do URL, que, inicialmente, apresentava 110 caracteres, mas é sugerida a redução para 75. Sendo assim, começamos com “qualquer-coisa-esta-no-youtube-e-isso-pode-nos-salvar” (*slug*), no *Yoast*, e encerramos com “tudo-esta-no-youtube” (*link* permanente), no *Rank Math*. No texto 3, já sabendo que o tamanho do URL seria um ponto de atenção, configuramos no *Yoast SEO* o *slug* “acai-tucuma-buriti”, que tem apenas 68 caracteres e foi aprovado como *link* permanente no *Rank Math*.

Para a categoria “Título, metadescrição e URL”, nos textos 4 e 5, fizemos um movimento de simulação um pouco diferente, pressupondo os usos estratégicos que a empresa de *marketing Conversion* e a *Google/Alphabet Inc.* poderiam fazer em seus próprios *blogs*. Nesses dois casos, antes de atender a demandas de edição do *Yoast SEO* e do *Rank Math*, fizemos uma busca simples no *Google* com trechos dos textos para que pudessemos chegar exatamente neles. Assim, vimos pela SERP como os *snippets* dos conteúdos apareceram e apenas reproduzimos aquela mesma combinação de título, metadescrição e URL em nosso espaço simulado, para ver como os *plugins* avaliaram. Ou seja, fizemos um caminho inverso

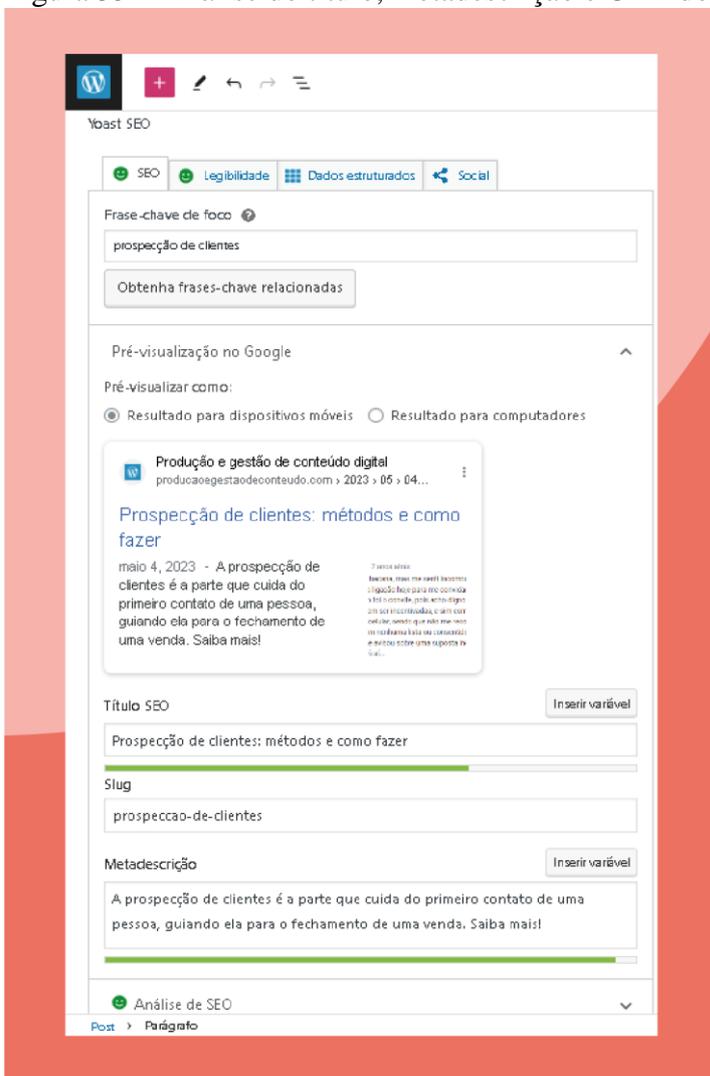
para adivinhar qual foi o uso estratégico dos elementos (título, metadescrição e URL) adotado pelos autores. Como vemos nas Figuras 34, 35, 36 e 37 a parametrização foi aprovada de primeira pelo *Yoast SEO*.

Figura 34 – *Snippet* da SERP/ texto 4



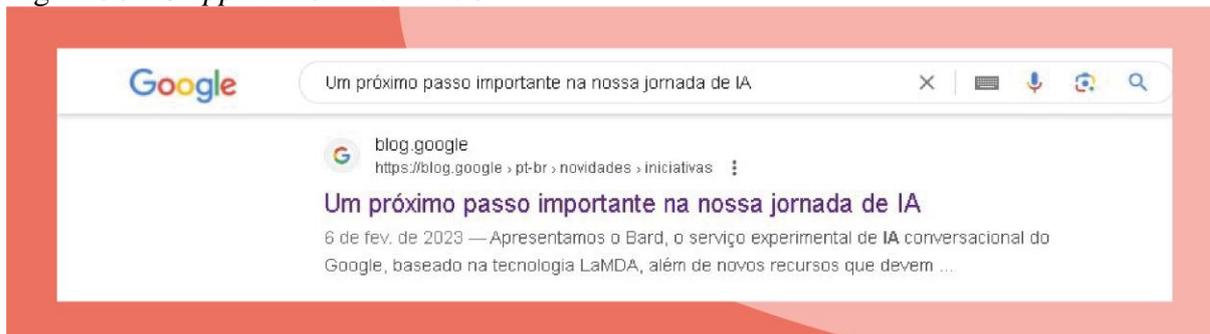
Fonte: Google, 2023.

Figura 35 – Análise do título, metadescrição e URL do *Yoast SEO*/ texto 4



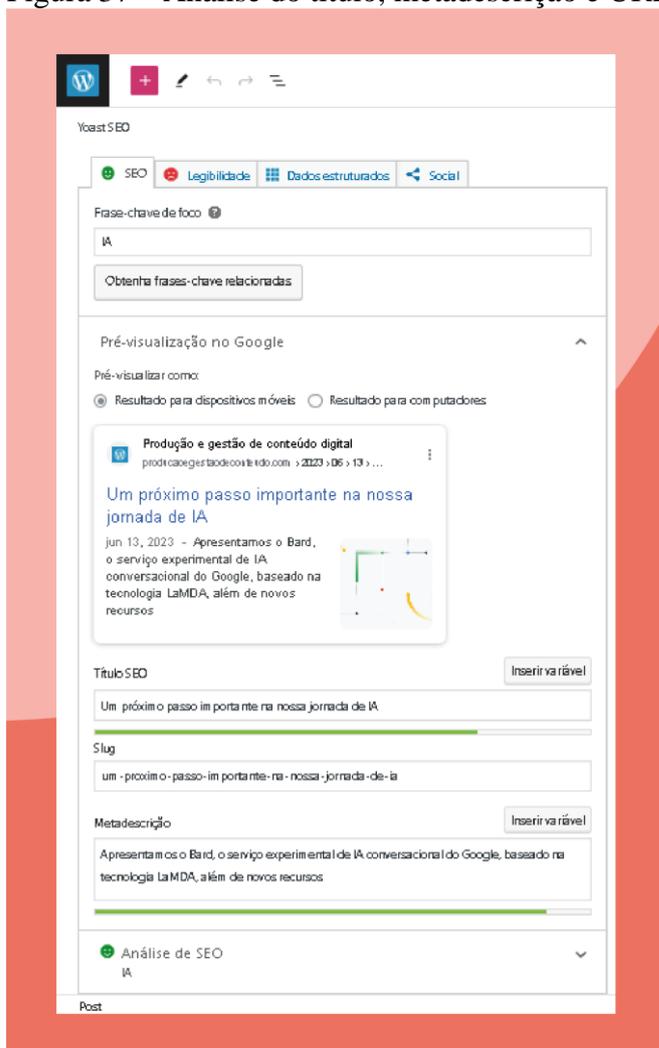
Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

Figura 36 – *Snippet* da SERP/ texto 5



Fonte: Google, 2023.

Figura 37 – Análise do título, metadescrição e URL do *Yoast SEO*/ texto 5



Fonte: Disponível em: <https://producaoegestaoeconteudo.com/>. Acesso em: 29 maio 2023.

Enfim, nessa categoria, verificamos que produtores e editores de texto incorporam parametrizações e preocupações como a quantidade de caracteres ou modelos de títulos, a fim de tornar suas páginas e conteúdos mais visíveis. A noção de visibilidade e performance

desses conteúdos está muito atrelada à eficiência do tripé título, metadescrição e URL, que, muitas vezes, funciona como porta de entrada. Assim, fica verificada a relação entre circulação de conteúdos e a necessidade de escolhas textuais que atendam a critérios das plataformas de busca. Para evidenciar ainda mais, resumimos no Quadro 13 a relação entre mediações algorítmicas e reconfiguração das práticas de escrita para a categoria 2 – título, metadescrição e URL:

Quadro 124 – Entrelaçamento entre plataforma de busca e escrita na *Web* (categoria 2 – título, metadescrição e URL)

Mediações algorítmicas e/ou diretrizes do <i>Google</i>	Práticas incorporadas na escrita e edição
A indexação do <i>Google</i> inclui processamento e a análise do conteúdo textual e das principais <i>tags</i> e atributos do conteúdo, como elementos <i><title></i> (títulos).	Submeter títulos a sistemas de análise de palavras-chave, a exemplo do <i>Google Trends</i> e desenvolver estratégias e planejamento de pautas e <i>briefing</i> com títulos descritivos e literais.
Algoritmos “leem” prioritariamente URLs, títulos, metadescrições, <i>headings</i> e trechos negritados.	Planejar textos na lógica da engenharia reversa. É produzir textos para os quais já sabemos existirem demandas de busca e fazer títulos, subtítulos exatamente iguais aos termos usados frequentemente para barra do <i>Google</i> .
O <i>Google</i> incentiva produtores de texto a criarem títulos descritivos tal como as pessoas escrevem na barra de busca, ou seja, o criador do texto precisa aprender a fazer escolhas textuais para atender às tendências da plataforma de busca	Contabilizar caracteres de títulos, metadescrições e URLs para serem visíveis na SERP.
Para o <i>Google</i> , conteúdo de qualidade é aquele que baseado em outras fontes, é reescrito adicionando valor e originalidade; não usa de termos exagerados ou chocantes no título principal ou da página.	Usar títulos diferentes que atendam uma linha editorial (título do conteúdo) e atendam lógicas algorítmicas (título SEO).

Fonte: Elaborado pela autora, 2023.

5.2.3 Hipertextualidade (*links*)

Nesta categoria, analisamos a percepção que os *plugins Yoast SEO* e *Rank Math* têm da hiperconectividade de um texto, elemento alcançado por meio do processo de construção de *links* (*link building*) ou construção da hipertextualidade. Em outros momentos da tese, já descrevemos a importância do *link* para o rastreamento de páginas e conteúdos. Também explicitamos que os *links* são a essência da *Web*, por isso o primeiro algoritmo do *Google* (*Pagerank*) era completamente baseado na autoridade de *links*; reforçamos que esses elementos continuam sendo muito importantes, mesmo depois de tantas atualizações de algoritmos. Vale ressaltar que os algoritmos enxergam na página HTML o elemento *<link>*, que especifica as relações entre o documento atual e um recurso externo, demonstrando estrutura relacional para navegação. O elemento *<link>* faz parte do grupo de metadados de

uma página, o que eleva a importância dele para os algoritmos. Como, afinal, os *plugins* vão ajudar o produtor de texto a compreender o peso dos *links* na determinação de visibilidade do conteúdo?

A inserção de *links* internos (que referenciam outras páginas do próprio *site*) e externos (que dão votos de confiança a *sites* terceiros) ajuda o *Google* a compreender a estrutura de uma página, conteúdo ou *site*. É como se o mecanismo de busca imitasse o processo de leitura hipertextual que o ambiente digital, inerentemente, permite ao leitor. Se leitores usam *links* para encontrar conteúdo e *sites*, o mecanismo de busca também usa. O *Google* rastreia *sites* seguindo *links*, internos e externos, usando o *Googlebot*¹⁵⁷. Esse robô acessa a página inicial do *site*, renderiza a página e segue o primeiro *link*. Seguindo os *links*, o *Google* pode descobrir a relação entre as várias páginas, postagens e outros conteúdos.

Os *links* internos são aqueles em que referenciamos páginas iniciais, menus, *feeds* de postagem, ou aqueles que chamamos *links* contextuais, que estão sobre algum trecho de texto ou termo-chave e dão aos leitores caminhos para expandir a leitura em conteúdos relacionados. Já os *links* externos são importantes para que os fatores EEAT sejam percebidos pelo *Google*, porque mostram em qual rede de autoridade um *site* está inserido ou como ele se conecta a outras organizações que são referência em assuntos específicos. *Links* que vêm de outros *sites* chamamos de “*backlinks*” e os *links* que enviamos para outro *site* são nomeados “*links* de saída”. Quanto mais *links* significativos uma página recebe, mais importante ela parece para os mecanismos de pesquisa. O *Google* tem um lance estatístico para valorar *links* de uma página, mas não vamos aqui detalhar essa fórmula. Importante é que os *plugins* *Yoast SEO* e *Rank Math* vão contabilizar o número de *links* internos e externos, além de medir o valor desses *links* para o ranqueamento. Assim, vão auxiliar o produtor ou editor de texto a (re)configurar essas “linkagens”.

Para todos os textos do *corpus*, criamos uma solução de simulação de *links* interno, considerando que os *plugins* apontaram a necessidade de inclusão desse parâmetro. Quando inserimos os textos no espaço simulado, eles perderam a linkagem interna porque, afinal, saiam do *site* original da publicação. A nossa solução foi inserir como *link* interno a página principal <https://producaoegestaodeconteudo.com/> para que os *plugins* detectassem a existência de uma linkagem autorreferente. Por exemplo, no texto 1, toda vez que aparecia “Agência Pública”, o *link* era enviado para a página principal do nosso *site* de simulação,

¹⁵⁷ Como exemplificamos no capítulo 4, o *Googlebot* é um *webcrawlers* que percorre páginas buscando correspondências textuais. Ele acessa automaticamente páginas da Web disponíveis ao público e segue os *links* desses sites.

como se indicassem retorno para a página principal. No texto 3, fizemos o mesmo para as referências à “Agência Bori” e, no texto 5, usamos a mesma tática quando aparecia a empresa “Google” no conteúdo.

Para o *link building* do texto 4, usamos uma tática baseada na estratégia presumida da empresa *Conversion*. Como sinalizamos na análise da categoria 1, o conteúdo do *blog* da *Conversion* dá muitos indícios de fazer parte de um planejamento maior de textos sobre temáticas de especialidade da agência. Sendo assim, no conteúdo original publicado no *blog* da empresa, o processo de linkagem é muito eficiente, pois a autorreferência é orquestrada para que os termos apareçam e sirvam de texto-âncora ou chamadas para outros *posts* do *blog*. Na página sobre prospecção de clientes, que escolhemos para o *corpus*, aparecem no decorrer do texto termos como “inbound marketing”, “SEO” e “marketing de conteúdo”, que são colocados de modo estratégico para servir de base para linkagem, ou seja, cumprindo a função de agentes do *link building* e ensinando aos algoritmos que, ao seguir o *link* sobre um conceito fundamental (como *marketing*), encontrará uma página da própria *Conversion*, afinal, ela se coloca como autoridade sobre o tema. A empresa cria um ciclo de autoridade, uma dinâmica de autorreferência baseada no *link building*. Aproveitamos essa trilha montada pela produtora de texto da *Conversion* e usamos os mesmos textos-âncora para linkagem, simulando uma página interna ao usar o *link* <https://producaoegestaodeconteudo.com/>.

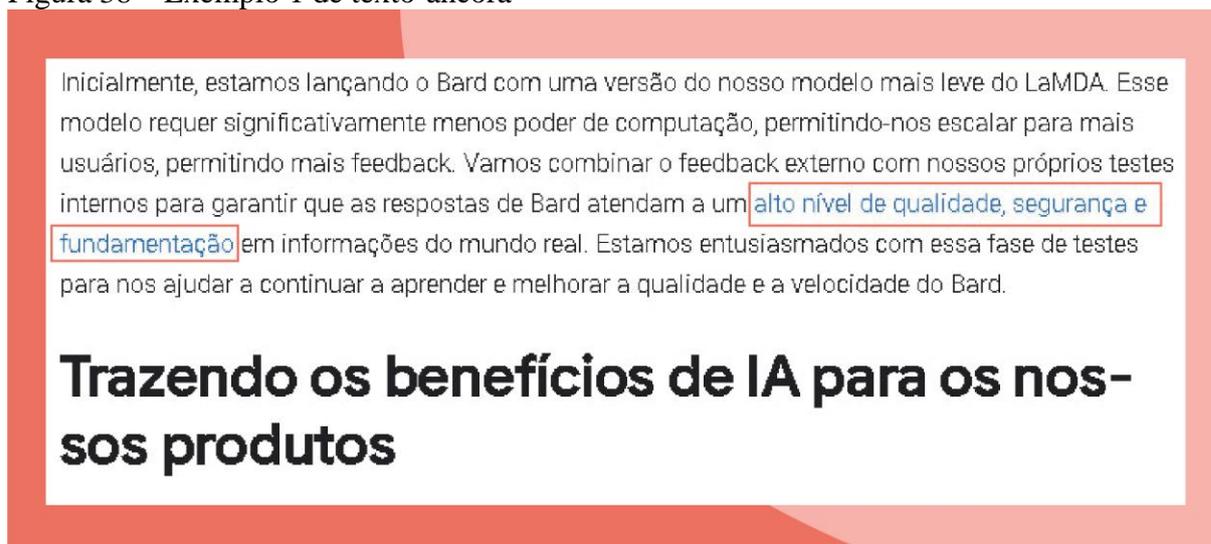
Ainda sobre o texto 4, saltou aos olhos a estratégia de relacionamento externa que envia *links* para grandes empresas de *marketing*, como *Hubspot*, *Hotmart*, *Rock Content* e *Buzz Lead*, que, na teoria, são concorrentes da *Conversion*. No entanto, a produtora de texto entende que é melhor referenciar *links* para concorrentes e mostrar para o *Google* que faz parte de uma rede de autoridade do que não se relacionar com essas outras empresas e parecer isolado. Enviar um *link* para o *site* concorrente é como dar um voto de confiança para ele, talvez esperando receber de volta e, assim, construir relacionamento digital para ser visto por algoritmos de buscadores. Essa estratégia de linkagem externa é o ponto mais controverso entre especialistas em SEO, porque, para algumas organizações, não é interessante ou faz pouco sentido editorialmente referenciar um *link* do *site* concorrente. Todavia, o que pode fazer pouco sentido na curadoria editorial, é muito apropriado na curadoria algorítmica (e vice-versa). Empresas de jornalismo e assessorias de imprensa, por exemplo, são muito resistentes em usar essa estratégia de *link building*. Apesar de enxergarem valor no reconhecimento de autoridade pelo *Google*, não querem abrir mão de que leitores tracem a rota da informação sem visitar o veículo concorrente. No texto 1, percebemos que a *Agência*

Pública é uma das que adota o *link building* com referência a outros veículos de comunicação, como *Folha de S.Paulo* e *Yahoo Notícias*. Enfim, essas referências entre empresas de um setor ou mercado acabam por treinar os algoritmos das plataformas de busca para reconhecer essas redes como *hubs* de autoridade.

Se trouxermos os conceitos fundamentais da hipertextualidade, já citados no capítulo 2 desta pesquisa, veremos que uma construção hipertextual está relacionada a uma oferta de possibilidades de rotas para o leitor de um texto. Verificamos que, no caso da hipertextualidade, em que há necessidade de ranquear em plataformas de busca e fazer circular os conteúdos, o *hiperlink* ganha *status* de “moeda de troca”. Ele é mais que um fator textual para ampliar sentidos ao leitor, tornando-se um elemento de mostraçãõ aos algoritmos com implicações importantes na performance dos conteúdos, portanto, com grande peso para os objetivos financeiros de produtores e editores de texto (especialmente aos que servem a empresas e organizações).

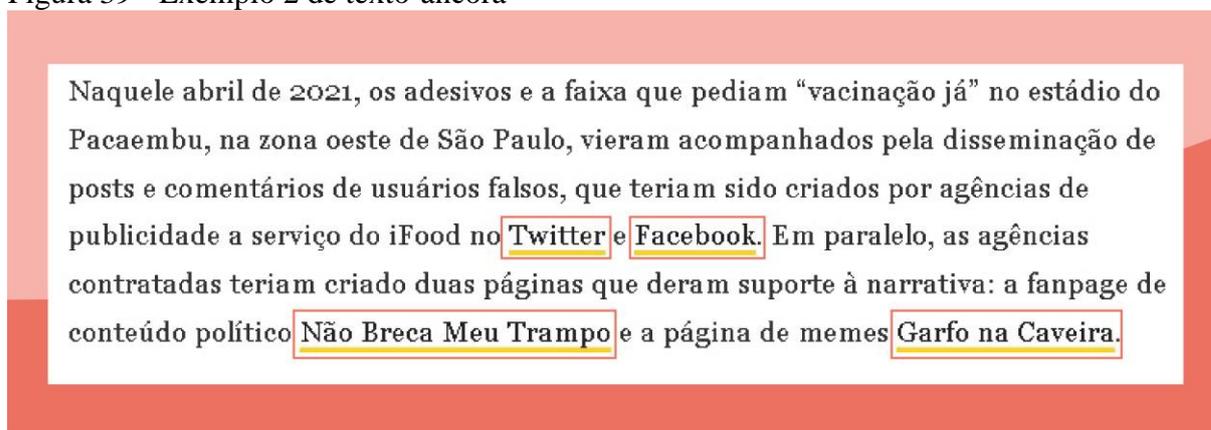
Já no texto 2 foi necessário um trabalho mais específico de linkagem, considerando que o conteúdo original não tinha *hiperlinks* e, claro, o *Yoast SEO* e *Rank Math* sinalizaram isso como um ponto negativo. O primeiro processo de edição foi inserir *links* externos, como, por exemplo, *hiperlink* no termo *YouTube*, enviando para o *site* da plataforma; e *hiperlink* no livro de Maty Ellen, indicando para o *site Estante Virtual*, onde é possível comprá-lo. Esse trabalho foi suficiente para ser “aprovado” em *links* externos, mas o *Yoast SEO* alertou sobre um problema: “Links com frase-chave: você está vinculando para outra página com as palavras que você quer que essa página classifique. Não faça isso!”. Quando inserimos o *hiperlink* no termo *YouTube*, que já estava configurada como a palavra-chave da postagem, sinalizamos para os algoritmos que um *site* externo era uma boa referência para informações sobre “YouTube” e isso é uma autossabotagem na estratégia de ranqueamento, porque precisamos fazer o algoritmo entender que nosso *site* é a referência para classificar o termo “YouTube”. Esse alerta mostrou a importância da escolha de texto-âncora – aquele texto visível e clicável de um *link* – no processo de escrita e edição. A seguir, estão dois exemplos de texto-âncora, cada um visualmente destacado de um jeito diferente. Geralmente, os textos-âncora aparecem sublinhados ou de cor diferente na página.

Figura 38 – Exemplo 1 de texto-âncora



Fonte: *Blog do Google*, 2023.

Figura 39 - Exemplo 2 de texto-âncora



Fonte: *Agência Pública*, 2023

De acordo com o tutorial do *Yoast SEO*, um bom texto-âncora diz ao leitor o que esperar e aumenta a chance de alguém clicar no *link*. Além disso, ajuda os mecanismos de pesquisa, ao fornecer contexto a eles, deixando a estrutura do texto clara para os buscadores. Quando um produtor cria um *link* para outro conteúdo, o *Google* usa o texto do *link* como um indicador do tópico da página vinculada. Se o texto do *link* corresponder à frase-chave (como erroneamente fizemos com *YouTube*), o *Google* não será capaz de dizer qual artigo é o mais relevante para esse tópico e essa confusão complicará o ranqueamento. Portanto, entendemos que o texto-âncora é um fator de classificação para o *Google*, pois esclarece ao algoritmo o contexto e o valor de um *link*. Diante disso, concluímos que usar a palavra-chave como texto âncora para *links* externos não é uma boa opção, portanto, retiramos o *hiperlink* do termo *YouTube* no texto 2.

Verificamos nesta categoria que produtores e editores de texto incorporam usos táticos e estratégicos de *links* em suas produções textuais com objetivo de promover a hipertextualidade para o leitor, mas, especialmente, performar em plataformas de busca. No espaço de simulação foi necessário reconfigurar a estrutura dos textos (re)fazendo linkagens para atender as indicações dos *plugins* sobre a ranqueabilidade dos conteúdos, portanto, evidenciamos o entrelaçamento entre escrita na *Web* e lógicas algorítmicas das plataformas de busca. A seguir, resumimos essas relações:

Quadro 135 – Entrelaçamento entre plataforma de busca e escrita na *Web* (categoria 3 – hipertextualidade)

Mediações algorítmicas e/ou diretrizes do Google	Práticas incorporadas na escrita e edição
Algoritmos perfilam <i>sites</i> pela quantidade de <i>links</i> recebido pela análise de quais <i>links</i> eles estabelecem (“qualidade” de <i>links</i>).	Inserir <i>links</i> internos (referenciam outras páginas do próprio <i>site</i>) e externos (dão votos de confiança a <i>sites</i> terceiros) que ajuda o <i>Google</i> a compreender a estrutura de uma página, conteúdo ou <i>site</i> .
Escanear e fazer <i>links</i> é um dos pilares do <i>Google Search</i> , responsável por rastrear, indexar e classificar conteúdos, considerando que o buscador não hospeda essas informações, mas as ordena (VAIDHYANATHAN, 2001)	Fazer <i>lobby</i> por <i>links</i> , um trabalho de <i>Digital PRs</i> ou Relações Públicas Digitais, porque os <i>links</i> são como “moeda de troca” para demonstração de autoridade.
<i>Page Rank</i> funcionava a partir de sistemas de “avaliação de importância” dos <i>links</i> espalhados na <i>Web</i> e que podem ser rastreados.	Criar dinâmicas para que <i>sites</i> e conteúdos recebam boas avaliações de clientes ou leitores, menções na <i>Wikipédia</i> – considerada pelo <i>Google</i> como <i>site</i> confiável – e menções em fóruns especializados. Tudo isso em busca do reconhecimento de autoridade.
<i>Webcrawlers</i> percorrerem páginas buscando correspondências textuais. Esses rastreadores acessam automaticamente páginas da <i>Web</i> disponíveis ao público e seguem os <i>links</i> desses <i>sites</i> , da mesma forma que um leitor faria se estivesse navegando no mesmo conteúdo e fazendo um processo de leitura hipertextual.	Inserir <i>hyperlinks</i> e chamadas ao leitor para <i>links</i> internos, ou seja, outras páginas do próprio <i>site</i> com objetivo de reduzir a taxa de rejeição e aumentar tempo de permanência em páginas.
Algoritmos contabilizam quantas vezes os <i>links</i> são citados em <i>sites</i> e fóruns de grande autoridade em algum tema.	
Diretriz do <i>Google</i> incentiva a presença de <i>links</i> rastreáveis nas páginas de conteúdo.	

Fonte: Elaborado pela autora, 2023.

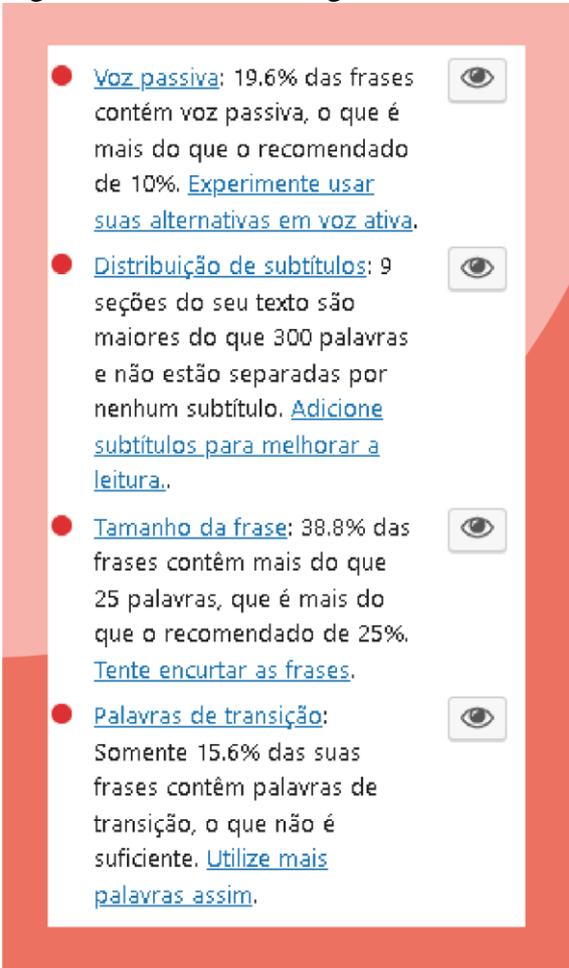
5.2.4 Legibilidade e escaneabilidade

Nesta categoria, analisamos a percepção dos *plugins Yoast SEO* e *Rank Math* sobre o quanto os textos analisados são legíveis aos olhos dos leitores e aos “olhos dos algoritmos”. A legibilidade e a escaneabilidade foram descritas no Capítulo 4 porque eram

diretrizes da “boa experiência de uso” declarada nas documentações do *Google*, mas vamos destrinchá-las nas análises, inclusive, abrindo oito subcategorias muito relacionadas à produção dos textos e que nos ajudarão a olhar para as reconfigurações de práticas de escrita.

Veremos o quanto os *plugins* vão sinalizar questões objetivas do texto como se fizessem, efetivamente, o “papel de linguistas”, auxiliando os produtores de texto a construir conteúdos mais otimizados. A parametrização de legibilidade e escaneabilidade parecem ser as mais trabalhosas para o produtor de texto, por requerer um esforço de edição. As alterações sugeridas pelos *plugins* nessa categoria têm implicações estruturais no texto, porque não são ações – como a criação de um título SEO – que ficam ocultas ao leitor. Essas são alterações visíveis ao leitor e que mudam bastante a primeira versão dos textos. A seguir, é possível ver as análises iniciais para alguns dos textos do *corpus* e verificar o quanto a legibilidade dos conteúdos era inapropriada aos “olhos” do *Yoast SEO*.

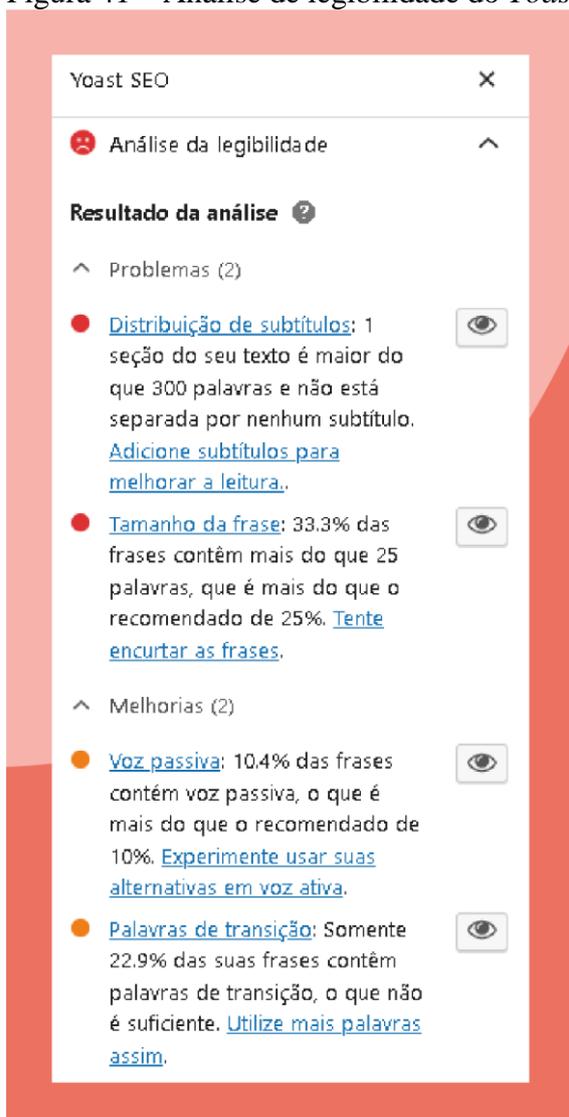
Figura 40 – Análise de legibilidade do *Yoast SEO* para o texto 1



- **Voz passiva:** 19,6% das frases contém voz passiva, o que é mais do que o recomendado de 10%. [Experimente usar suas alternativas em voz ativa.](#)
- **Distribuição de subtítulos:** 9 seções do seu texto são maiores do que 300 palavras e não estão separadas por nenhum subtítulo. [Adicione subtítulos para melhorar a leitura.](#)
- **Tamanho da frase:** 38,8% das frases contém mais do que 25 palavras, que é mais do que o recomendado de 25%. [Tente encurtar as frases.](#)
- **Palavras de transição:** Somente 15,6% das suas frases contém palavras de transição, o que não é suficiente. [Utilize mais palavras assim.](#)

Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

Figura 41 – Análise de legibilidade do *Yoast SEO* para o texto 5



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

Percebe-se que o *plugin* estabelece percentuais ideais para cada parâmetro e sinaliza ao produtor/editor o que falta para alcançar o nível apropriado de otimização. O texto 1, como sabemos, já foi submetido ao *plugin Yoast SEO* na publicação original da *Agência Pública*, mesmo assim, apresenta inconformidades na legibilidade. Isso indica que nem sempre editores e produtores de texto se mobilizam para cumprir todas as parametrizações indicadas, haja vista o texto 5, que é do *blog* do próprio *Google*, e não tem equilíbrio ideal para otimização, segundo as análises do *Yoast SEO*.

5.2.5 Tamanho do texto

Os *plugins Yoast SEO* e *Rank Math* sinalizam que, para serem encontrados, os textos precisam ter mais de 300 palavras; a chance de classificação no *Google* aumenta se um conteúdo tiver alta qualidade de mais de 600 palavras. O *Rank Math*, como vemos na indicação a seguir, aponta para uma métrica mais pontual, inclusive, dando nota zero para textos com menos de 600 palavras:

Tabela 1 – Tamanho do texto

Palavras	Pontuação
Mais de 2500	100%
2000-2500	70%
1500-2000	60%
1000-1500	40%
600-1000	20%
Abaixo de 600	0%

Fonte: Elaborado pela autora com base em *Rank Math*, 2023.

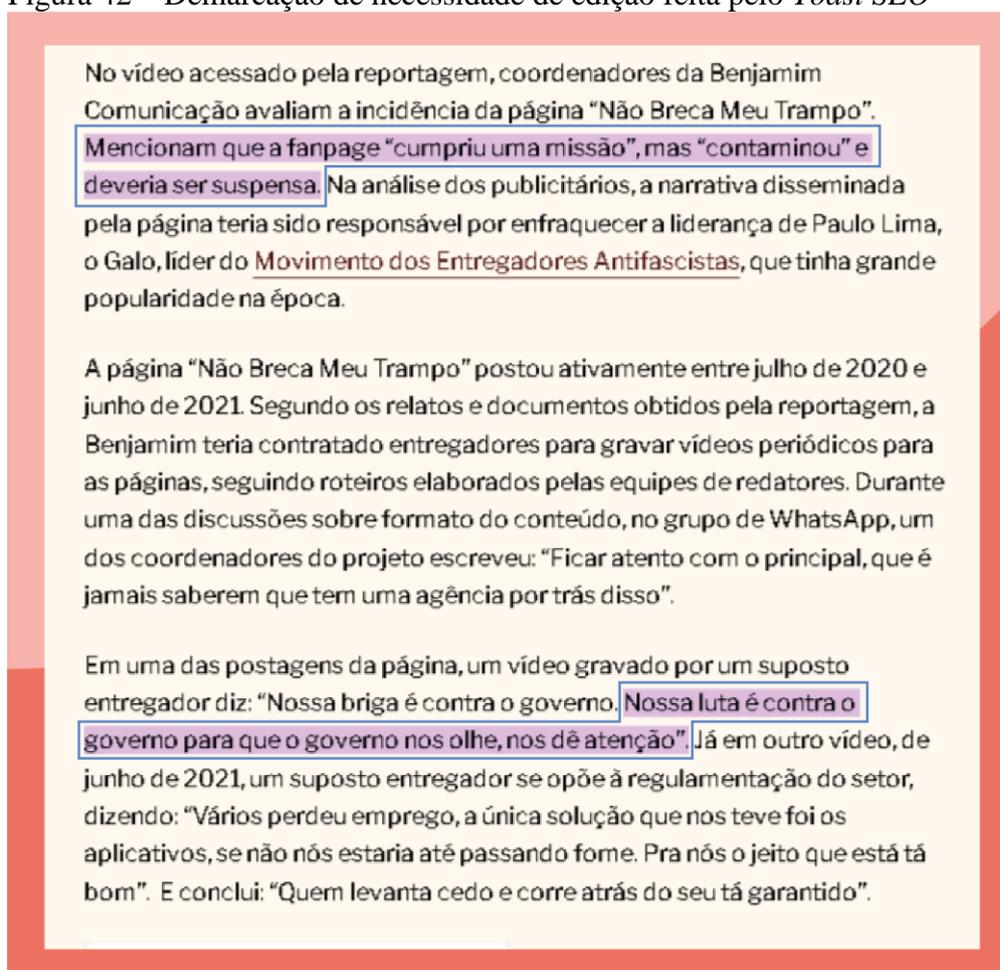
Quanto mais longo, parametrizado e otimizado um texto estiver, mais pistas ele deixará para o *Google* determinar do que se trata aquela produção, considerando que, mais frequentemente, a palavra-chave em foco aparecerá. Ao analisar o *corpus* da pesquisa, o texto 1 tem maior número de palavras (5.759 palavras), seguido do texto 4 (2.451), texto 5 (1.152), texto 2 (881) e, por fim, o texto 3 (522). Esse último, em especial, recebeu avaliação laranja/bom no *Yoast SEO* e pontuação 0% no *Rank Math*, no entanto, como nosso propósito não incluía a reescrita do texto, não inserimos mais palavras. Observando essa parametrização, verificamos que produtores e editores de texto podem ser impelidos até mesmo a aumentar um texto que esteja fora dos padrões da otimização, sendo necessária a refeição, para torná-lo mais elegível aos algoritmos das plataformas de busca. Ou, então, o tamanho do texto pode ser uma diretriz para produtores que condiciona a entrega, a exemplo de como trabalham as agências de *marketing* de conteúdo, que determinam o número de palavras para um redator produzir (CRUZ; RIBEIRO, 2016). Essa determinação visa, muitas vezes, atender as demandas de plataformas de busca.

5.2.5.1 Tamanho do parágrafo

Os *plugins Yoast SEO* e *Rank Math* sinalizam que parágrafos de tamanho adequado desempenham um papel importante para tornar seu texto legível (para o leitor) e escaneável (para algoritmos), pois ajudam a dividir seu texto em pedaços mais palatáveis. Para fazer essa sinalização, o *Yoast SEO* vai contabilizar palavras e sugerir que parágrafos tenham menos que 150 palavras, podendo chegar a 200, mas nunca ultrapassar essa marca. Já o *Rank Math* notifica se um parágrafo estiver com mais de 120 palavras. Vemos aqui mais uma diferença da parametrização dos *plugins*, um sinal de que não há matemática exata para o ranqueamento no *Google*.

Outra diferença entre os *plugins* e que tem implicações no trabalho dos produtores e editores de texto é que o *Yoast SEO* tem uma ferramenta de destacar (*highlight*) nos trechos em que são necessárias adequações, como vemos na Figura 42. No exemplo a seguir, estão demarcados trechos que precisam de palavras de transição no texto 1.

Figura 42 – Demarcação de necessidade de edição feita pelo *Yoast SEO*



Fonte: Disponível em: <https://producaoegestaodeconteudo.com/>. Acesso em: 29 maio 2023.

A funcionalidade de demarcação não existe no *Rank Math*, portanto, quando esse *plugin* apontou parágrafos longos nos textos 1 e 2, tivemos que garimpar esses trechos para reduzir. Sendo assim, fizemos uma detecção visual de quais parágrafos poderiam ultrapassar 120 palavras e copiamos cada um deles em um editor de texto (*Google Docs*) para usar a ferramenta de contar palavras. Tudo isso gasta muito tempo no processo de edição, mesmo que simulado, como fizemos. Para se ter uma ideia, no texto 2, foi preciso repetir a procura por parágrafos longos cinco vezes.

5.2.5.2 Frases consecutivas

Ainda na perspectiva da legibilidade, os *plugins* estão preparados para sinalizar elementos que tornam a leitura repetitiva, cansativa e pouco fluida. Por isso, *Yoast SEO* e *Rank Math* sinalizam a presença de frases que começam com as mesmas palavras ou que reiteram uma ideia sem adicionar sentido ou que poderiam ser reduzidas. Os *plugins* vão contabilizar e indicar, percentualmente ou numericamente, a necessidade de incluir mais diversidade e variedade ao texto. *Yoast SEO* e *Rank Mat* vão sugerir estratégias de escrita, como inserir sinônimos, usar palavras de transição, reescrever aumentando ou reduzindo sentenças de modo mais fluido. Mas, afinal, por que esta é uma estratégia importante para a visibilidade do texto? Por que algoritmos consideram que um texto com “Boa experiência de uso” seja mais legível, portanto, um candidato mais apropriado para o ranqueamento. Na análise do *corpus*, este foi o único entre todos os parâmetros que os cinco textos passaram ilesos, sem apontamentos para adequação.

5.2.5.3 Palavras de transição

As palavras de transição também estão incluídas na noção de “Boa experiência de uso”; portanto, os *plugins* vão contabilizar esses conectivos mostrando que uma quantidade inadequada pode deixar o texto menos interessante e coerente. Portanto, *Yoast SEO* e *Rank Math* sugerem a inserção de conjunções, pronomes demonstrativos e expressões de conexão de sentido que evidenciem ao leitor a relação entre frases, sentenças ou até mesmo parágrafos, permitindo ligação de ideias e previsão do que está por vir no texto. O *Yoast SEO* sinaliza como vermelho/ruim quando menos de 20% das frases contêm palavras de transição;

laranja/ok acima de 21% e verde/bom quando há 30%. Ao analisar essa subcategoria no *corpus*, foi possível verificar que os *plugins* – com base na lógica de motores de busca – valorizam muito a presença de conectivos, afinal, entre todas as parametrizações, essa foi a que exigiu mais esforço de edição para alcançar índices ideais nos cinco textos do *corpus*.

No texto 1, a primeira análise do *Yoast SEO* apontou que apenas 15,6% das frases da reportagem continham palavras de transição, portanto, estávamos com avaliação vermelha/ruim. Iniciamos a inserção dessas palavras mantendo a coerência do texto, mas percebemos que, após três inserções, o percentual indicado pelo *plugin* subiu apenas para 16,8%. Sendo assim, a edição de palavras de transição demandou mais de três horas de dedicação contínua no processo simulado de edição. Usamos até uma lista de conectivos do *Manual de Redação da Folha de S.Paulo* para auxiliar nas ideias e opções de adequações. Porém, o *Yoast SEO* não considerava qualquer conectivo inserido, mesmo se o termo fosse coerente no contexto. O advérbio “inclusive” e a expressão de relação causa e consequência “de tal maneira” não foram aceitos como palavras de transição válidas. Percebemos também, no decorrer das edições, que vários termos com função conectora (de conformidade e de adição) poderiam ser apenas trocados para que o *Yoast SEO* os considerasse adequados. Por exemplo, o *plugin* validou a mudança da expressão “segundo consta em um relatório de junho de 2021” por “conforme consta em um relatório de junho de 2021”, ou ainda, trocar “De acordo com os documentos obtidos” por “Conforme documentos obtidos”.

No texto 2, a primeira análise do *Yoast SEO* apontou que apenas 15% das frases da reportagem continham palavras de transição. Nesse caso, a grande dificuldade da adequação aconteceu porque este não é um texto informativo, como os textos 1, 3 e 4, é um conteúdo com tom de crônica para uma coluna fixa do jornal *Rascunho*. A tessitura proposta pela autora, Ana Elisa Ribeiro, realmente exclui as palavras de transição, pois é para o texto ser dessa maneira. Mesmo assim, encaramos o processo de simulação das palavras de transição no texto 2 com o objetivo, inclusive, de testar as respostas do *plugin* para um texto não informativo. Verificamos que a parametrização para otimizar pode reconfigurar uma prática de escrita a ponto de retirar a personalidade e o estilo do texto, caso produtores e editores optem por priorizar o ranqueamento. Não parece ser esse o objetivo do jornal *Rascunho* e poderia soar estranho uma despersonalização dos textos de colunas, escolhidos por curadoria humana, em detrimento da lógica algorítmica. Vejamos a seguir outras alterações de palavras de transição nos textos 1 e 2:

Quadro 146 – Inserção de palavras de transição

	Antes	Depois
	A petição <i>online</i> havia sido criada sete dias antes pela equipe da agência de publicidade que o contratou.	<u>Ademais</u> , a petição <i>online</i> havia sido criada sete dias antes pela equipe da agência de publicidade que o contratou.
Texto 1	O tema da vacinação prioritária não era inédito para o <i>iFood</i> . No final de janeiro, a empresa anunciou uma doação de R\$ 5 milhões para a construção de novas instalações do Instituto Butantan.	O tema da vacinação prioritária não era inédito para o <i>iFood</i> . <u>Isso porque</u> , no final de janeiro, a empresa anunciou uma doação de R\$ 5 milhões para a construção de novas instalações do Instituto Butantan.
	No <i>Twitter</i> , perfis de usuários falsos que seriam administrados pelas agências de publicidade também postaram sobre a vacinação dos entregadores.	<u>Só para ilustrar</u> , no <i>Twitter</i> , perfis de usuários falsos que seriam administrados pelas agências de publicidade também postaram sobre a vacinação dos entregadores.
	Herdei esse livro. Não sei quando.	Herdei esse livro, <u>todavia</u> , não sei quando.
Texto 2	Acessível: fala as horas de uma em uma. Que beleza. Nem precisava de tanto. Era só ser relógio e estaria bom. Levei.	Acessível: fala as horas de uma em uma. Que beleza. Nem precisava de tanto. Era só ser relógio e estaria bom. <u>Sendo assim</u> , levei.
	Abri a caixa, nada de folheto de instrução. Tudo autoexplicativo.	Abri a caixa, nada de folheto de instrução, <u>afinal</u> , era tudo autoexplicativo.

Fonte: Elaborado pela autora, 2023.

Os textos 4 e 5 alcançaram logo na primeira análise do *plugin* o padrão verbe/bom. Em contrapartida, a análise do *Yoast SEO* para o texto 3 indicou a ausência de palavras de transição em todas as frases, ou seja, foi necessário inserir termos de transição em todo o conteúdo. Nesse caso, analisamos que um texto sem conectivos pode dificultar a legibilidade, considerando que entre todos os textos este tinha menos fluidez e coesão, além de dificultar a escaneabilidade, porque algoritmos não o reconhecerão como um conteúdo passível de ser classificado. Entre todos os textos do *corpus*, o conteúdo da *Agência Bori* era o menos elaborado linguisticamente, constando erros gramaticais e trechos de difícil compreensão.

5.2.5.4 Tamanho das sentenças

Frases longas são consideradas pelos *plugins* como elementos dificultadores de leitura, portanto, o tamanho das sentenças é um parâmetro sinalizado na análise textual. *Yoast SEO* e *Rank Math* vão contabilizar o número de palavras da sentença e, sem seguida, calcular quantas frases têm mais de 20 palavras. Se mais de 25% das frases tiverem mais de 20 palavras, será emitida a sugestão de encurtá-las. Ao analisar essa subcategoria no *corpus*, observamos que o texto 2 apresentou a redação adequada na análise dos *plugins*, e não foi

necessária edição para esse parâmetro. Já o texto 3 tinha 43,5% das frases longas; o texto 1, 37,2%; o texto 5, 33,3%; e o texto 4, 25,8%. Em todos esses casos, iniciamos o processo simulado de edição na tentativa de reduzir, refazer e enxugar essas sentenças. Essa ação foi a que mais alterou estruturas de textos do *corpus*, porque, eventualmente, era necessária uma reescrita. No texto 3, as soluções quase sempre foram de pontuação, como no exemplo a seguir, em que inserimos ponto e abrimos nova sentença para descrever a participação da *Agência Bori* na pesquisa científica citada:

Antes:

É o que mostra a publicação “Bioeconomia amazônica: uma navegação pelas fronteiras científicas e potenciais de inovação”, realizada pela World-Transforming Technologies (WTT) com a participação da Agência Bori, divulgada na sexta-feira (8).

Depois

É o que mostra a publicação “Bioeconomia amazônica: uma navegação pelas fronteiras científicas e potenciais de inovação”, realizada pela World-Transforming Technologies (WTT). A Agência Bori participou da pesquisa, que foi divulgada na sexta-feira (8).

Já no texto 1, retiramos todas as marcações condicionais, especialmente o uso de verbos no futuro do pretérito, como “A presença do funcionário teria como objetivo”, “as agências teriam produzido conteúdo para redes sociais”, pois avaliamos que, numa perspectiva jornalística, faz pouco sentido esse uso, afinal, a repórter está cercada de documentos comprobatórios que sustentam a reportagem e tornam as afirmativas inquestionáveis. Essa decisão que tomamos reduziu muitas sentenças da reportagem, ou seja, foi uma ação editorial que acabou por atender a uma demanda algorítmica de redução de sentenças.

5.2.5.5 Voz passiva

O último aspecto textual que os *plugins* sinalizam para a legibilidade é evitar o uso da voz passiva, indicando que a melhor construção gramatical para os textos alcançarem performance em plataformas de busca é a voz ativa. Portanto, *Yoast SEO* e *Rank Math* vão contabilizar e sinalizar sentenças que desrespeitam a ordem – sujeito, verbo, predicado. A lógica dos *plugins*, imitando a lógica dos buscadores, considera que a voz passiva torna o texto mais distante e menos legível, por exigir mais esforço cognitivo do leitor. Na avaliação da voz passiva, *Yoast SEO* e *Rank Math* vão considerar apropriado o valor de 10% de frases na voz passiva e será aceitável o índice de 10% a 15%, não podendo ultrapassar este

parâmetro. As sugestões de reescrita, geralmente, são para trocar a ordem de sujeito e predicado, além da alteração verbal adequada para a nova combinação.

Ao analisar essa subcategoria no *corpus*, observamos que o texto 2 apresentou, novamente, a redação adequada na análise dos *plugins*, e não foi necessária edição para esse parâmetro. É importante observar como a tessitura dessa coluna no jornal *Rascunho* parece fazer bastante diferença nas análises de legibilidade e escaneabilidade, um indicativo de que os cuidados com aspectos linguísticos são preponderantes. Isso pode contrariar os que preconizam a otimização de textos como técnicas não relacionadas às operações linguísticas, cognitivas e de construção textual. Nossas análises demonstram a valorização de uma produção que envolva organização, compreensão e funcionamento do texto para seus leitores.

A voz passiva estava presente no texto 1 em 18,1% das frases, no texto 3, em 17,4%, no texto 4, em 12,9% e, no texto 5, em 10,4%. Sendo assim, para essa categoria, o processo de edição simulada foi mais crítico para os textos 1 e 3, em que foram necessárias muitas intervenções, como veremos nos exemplos a seguir:

Quadro 157 – Mudança para voz ativa

	Antes	Depois
	Nesse sentido, oito dias após o “Breque dos Apps” foi criada a página <i>Não Breca Meu Trampo</i> no <i>Facebook</i> .	Nesse sentido, oito dias após o “Breque dos Apps”, os estrategistas criaram a página <i>Não Breca Meu Trampo</i> no <i>Facebook</i> .
Texto 1	Quando a “Não Breca Meu Trampo” foi criada, um grupo com o mesmo nome foi aberto por Moriael Paiva, também no <i>Facebook</i> .	Quando surgiu a “Não Breca Meu Trampo”, Moriael Paiva abriu um grupo com o mesmo nome, também no <i>Facebook</i> ..
	Ademais, a empresa foi contratada por João Doria para fazer monitoramento de suas redes, segundo publicação da <i>BBC Brasil</i> .	Ademais, a empresa fez o monitoramento das redes sociais de João Doria, segundo publicação da <i>BBC Brasil</i> .
	Os dados foram obtidos a partir de um mapeamento de pesquisas recentes publicadas na forma de artigos científicos na base internacional de periódicos <i>Web of Science</i> .	Um mapeamento de pesquisas recentes publicadas na forma de artigos científicos na base internacional de periódicos <i>Web of Science</i> deu origem aos dados.
Texto 3	As matérias-primas mais estudadas pelos grupos de pesquisa brasileiros são açaí, tucumã, buriti, piper, aniba, castanha do Brasil, andiroba, cupuaçu, guaraná e bacaba. A aplicação desses produtos	As matérias-primas mais estudadas pelos grupos de pesquisa brasileiros são açaí, tucumã, buriti, piper, aniba, castanha do Brasil, andiroba, cupuaçu, guaraná e bacaba. Diversas atividades industriais como produtos artesanais, fabricação

Antes	Depois
pode ser feita em diversas atividades industriais: produtos artesanais, fabricação de tecidos, artesanato, fios e redes de pesca, compósitos cimentícios para construções sustentáveis, filmes biodegradáveis.	de tecidos, artesanato, fios e redes de pesca, compósitos cimentícios para construções sustentáveis e filmes biodegradáveis usam esses insumos da Amazônia.

Fonte: Elaborado pela autora, 2023.

Observando as alterações de voz passiva *versus* voz ativa, verificamos que um produtor ou editor de texto que tenham como objetivo otimizar um texto, seguindo a parametrização de *plugins* SEO, precisará lançar mão de recursos textuais diversos e equilibrar as mudanças que fazem sentido para o contexto de cada produção. Isso porque fazer as alterações sinalizadas pelo *Yoast SEO* e *Rank Math* não tem relação com melhorar ou piorar o texto, mesmo porque essa avaliação é subjetiva. A análise dos *plugins* é muito mais quantitativa do que qualitativa, ficando à “escolha” do profissional de texto escolher caminhos de adequação. Claro, escolher é relativo aos objetivos do texto, da instituição ou organização que o publica, mas isso nos faz refletir sobre a ambivalência desses produtores/editores e a importância deles como curadores para definir a quais demandas algorítmicas das plataformas de busca um texto se submeterá ou não. É por isso que faz todo sentido os *plugins* não serem *softwares* editores automatizados que interferem no texto, mas, sim, ferramentas que indicam as possibilidades de edição, aguardando uma ação humana.

5.2.5.6 Hierarquia do texto – distribuição de intertítulos ou cabeçalhos (*headings*)

O que tornará um texto mais escaneável para leitores e para algoritmos das plataformas de busca é o modo como ele está hierarquizado, conforme adiantamos ao descrever a escaneabilidade na seção 4.4. O que os *plugins Yoast SEO* e *Rank Math* vão fazer é sinalizar a necessidade de inserção de intertítulos, que são convertidos em cabeçalhos (*headings*) no HTML, e que organizam a hierarquização das informações no texto. Os *plugins* consideram que os cabeçalhos orientam os leitores no decorrer de um artigo e também os algoritmos no decorrer do código HTML. Sendo assim, cabeçalhos devem indicar do que se trata uma seção ou parágrafo para que as pessoas e os robôs saibam o que esperar daquele trecho. É como se o produtor de texto desse opções para leitores e algoritmos decidirem quais seções do texto vão ler. Além disso, os cabeçalhos facilitam a acessibilidade porque ficam demarcados no HTML da página como pontos de atenção ao texto, sendo assim ajudam os

aplicativos leitores de tela a compreender melhor a estrutura de um conteúdo, permitindo que pessoas com deficiência visual possam decidir se querem ler um artigo na íntegra ou navegar por intertítulos específicos.

O *Yoast SEO* e *Rank Math* vão sinalizar para o produtor ou editor de texto como usar de forma eficaz a divisão de cabeçalhos, sugerindo a distribuição pelo conteúdo, além de indicar a necessidade de inserção de palavras-chave nesses intertítulos. Assim, os *plugins* indicam níveis de hierarquia – de cabeçalho 1 a 6 – que serão convertidos no HTML, onde *heading 1 (h1)* é o nível mais alto e *heading 6 (h6)* é nível mais baixo. Ressaltamos que, no HTML, as marcações `<h1>` a `<h6>` fazem parte de um grupo conhecido como elementos de separação, que permitem organizar o conteúdo do documento em partes lógicas.

Conforme a parametrização dos *plugins*, cada bloco de texto demarcado por um intertítulo deve ter no máximo 300 palavras, portanto, essa é a referência numérica que será usada para sugerir nova inserção de cabeçalho. Vejamos um exemplo apresentado no tutorial do *Yoast SEO*¹⁵⁸.

Digamos que temos um post de blog sobre sapatilhas de balé. Escolhemos “sapatilhas de balé” como nossa palavra-chave principal e escrevemos um artigo sobre todos os motivos pelos quais gostamos de sapatilhas de balé. Veja como pode ser a estrutura desta postagem:

H1: As sapatilhas de balé são essenciais

H2: Por que achamos as sapatilhas de balé essenciais?

H3: Eles não são apenas cor de rosa

H3: Você pode usá-las para além de dançar

H3: As sapatilhas de balé podem ser mais baratos do que você pensa

H2: Onde você pode comprar suas sapatilhas de balé?

H3: Os 10 melhores sites de materiais de balé

H3: Nossas lojas de dança locais favoritas

Nesse exemplo, foi montada uma estrutura lógica, usando *tags H2* para planejar sessões e *tags H3* para cobrir tópicos específicos. Ao analisar essa subcategoria no *corpus*, foi possível verificar em que medida os *plugins* – com base na lógica de motores de busca – sugerem a (re)distribuição e estruturação de cabeçalhos. O *Rank Math*, em especial, sugere outro recurso de legibilidade e escaneabilidade, que é a inserção do sumário (*Table of Content*). Ele facilita a navegação pelo conteúdo, permitindo ao leitor uma rota mais hipertextual, não linear e fragmentada dentro do texto, como se ele pudesse ver o sumário de um livro e escolher explorar apenas capítulos específicos.

¹⁵⁸ Disponível em: <https://yoast.com/how-to-use-headings-on-your-site/#:~:text=With%20headings%2C%20you%20should%20always,Google%20understand%20your%20content%2C%20too>. Acesso em: 29 jun. 2023.

Nos textos 3, 4, e 5 não foi necessário editar cabeçalhos. Já no texto 1, foram apontadas alterações em nove seções, que estavam com mais de 300 palavras, sinalizando a necessidade de uma re(hierarquização) geral do texto. Inicialmente, inserimos intertítulos aleatórios, com objetivo de aumentar as repartições, mas percebemos que a mudança não melhorava o *status* de avaliação dos *plugins*. Então, percebemos que seria necessário contar as palavras de cada seção para descobrir parágrafos ou sentenças estratégicas para fazer a divisão. Conhecendo os limites de cada seção, criamos novos cabeçalhos e deslocamos outros, inclusive, mantendo intertítulos originais como H2 e inserindo novos como H3. Essa parametrização alterou bastante a estrutura do texto 1, como vemos no Quadro 17. A principal alteração é o crescimento do número de cabeçalhos, tornando o texto mais repartido, além da inserção estratégica da palavra-chave “ifood” em alguns intertítulos.

Quadro 168 – Distribuição de cabeçalhos texto 1

Cabeçalhos originais publicados na <i>Agência Pública</i>	Cabeçalhos depois da nossa edição
Marketing 4.0: “para que ninguém desconfie” O início: nasce uma página para desmobilizar entregadores “A gente matou o Galo O esquema se amplia: humor e memes O marketing digital chega às ruas: “missão vacina” e agente infiltrado O apoio para todo o esquema: perfis falsos e nanoinfluenciadores Objetivo alcançado: “esvaziar o discurso” Outro lado	Marketing 4.0: “para que ninguém desconfie” Lado B O início: nasce uma página para desmobilizar entregadores Estratégia “Não Breca Meu Trampo” Conteúdo suspeito “A gente matou o Galo Conteúdo frequente e roteirizado Nova parceria O esquema se amplia: humor e memes Piada com a realidade dos entregadores Impulsionamento O marketing digital chega às ruas: “missão vacina” Guerra de narrativas: iFood x entregadores Agente infiltrado Falsa repercussão Ação de “sucesso” O apoio para todo o esquema: perfis falsos e nanoinfluenciadores Os “VIPs” Os falsos “motocas” do iFood Minando mobilizações Objetivo alcançado: “esvaziar o discurso” O (não) diálogo do iFood Reajuste x “trabalho descente” Outro lado O que disse a Benjamin Comunicação O que disse a Social Qi O que disse o iFood

Fonte: Elaborado pela autora, 2023.

O texto 2 foi publicado originalmente sem cabeçalho, apesar de existirem intertítulos, a exemplo do “Procuras úteis”, que vemos na Figura 43. É que eles não estavam demarcados de modo hierárquico, portanto, funcionavam apenas como divisores para o texto, e não como sinalizadores de seções.

Figura 43 – Intertítulos texto 2

A internet tem de um tudo. O YouTube tem mais. Tudo se acha, mesmo sem qualidade, sem condição. De tudo existe um pouco. Tudo muito. Todo assunto, toda instrução, *do it yourself*, com ou sem *ring light*. Em palavra, imagem, som, vídeo e áudio, do jeito que for, de tudo há. Qualquer coisa fica fácil de achar. O filtro, se tiver, é a gente que faz. Ou nem isso. E a prova de que tudo pode estar no YouTube, que desbancou lâmpadas de cristal e enciclopédias de trinta volumes, é que procurei instruções tão díspares quanto *sui generis*, nos últimos tempos, e as soluções estavam sempre lá. Até quando alguém me dizia: ah, mas não deve ter isso no YouTube. E eu dizia: quer ver só?

Procuras úteis

Há alguns anos, quando eu precisava tirar uma mancha da roupa, pelo de casaco, risco de disco ou lustrar um sapato, minha mãe vinha com um livro ou dois e me dizia para ler a página tal. O livro não tinha tudo, mas geralmente dava uma dica ou macete que quebrava bem o galho. Aliás, ainda tenho aqui, em algum lugar da estante, o **Dicas, truques e quebra-galhos**, assinado por Mary Ellen, publicado no Brasil pela Rio Gráfica em 1976. Sim, o manual era quase da minha idade, e minha mãe o brandia sempre que algum desafio esquisito aparecia em casa. No subtítulo da obra lê-se: “maneiras fáceis, rápidas e divertidas de resolver problemas caseiros”. Existia também um volume dedicado ao ambiente da cozinha. (Procurem, que ainda se acha na Estante Virtual).

Fonte: Disponível em: <https://rascunho.com.br/>. Acesso em: 29 maio 2023.

Assim, seguimos a edição transformando os intertítulos em cabeçalhos, mas foi necessário inserir a palavra-chave em um deles. Os intertítulos pareciam formar uma tríade, com certa cadência textual, provavelmente planejada pela autora. No entanto, quando fizemos a edição, esse tom metafórico é atravessado por uma construção mais literal, com a inserção do termo “Youtube”.

Quadro 179 – Distribuição de cabeçalhos texto 2

Cabeçalhos originais publicados no site <i>Rascunho</i>	Cabeçalhos depois da nossa edição
Procuras úteis	Procuras úteis

Procura recente

Procura recente no *YouTube*

A solução, pois

A solução, pois

 Fonte: Elaborado pela autora, 2023.

Enfim, sobre todos os aspectos da legibilidade e escaneabilidade, observamos que os parâmetros se relacionam de modo que, ao inserir uma palavra de transição, era possível que aumentássemos o tamanho da frase e do parágrafo, incorrendo em alerta para dividir uma seção com cabeçalhos. Esse entrecruzamento de diretrizes mostra que nem sempre será possível executar todas ações indicadas pelos *plugins* e que o processo de edição à luz dos parâmetros indicados pelos *plugins* pode ganhar um tom mecânico e pouco reflexivo se aquele texto não foi originalmente pensado para atender à relevância de públicos (linha editorial) e relevância de algoritmos.

Quadro 18 – Entrelaçamento entre plataforma de busca e escrita na *Web* (categoria 4 – legibilidade e escaneabilidade)

Mediações algorítmicas e/ou diretrizes do <i>Google</i>	Práticas incorporadas na escrita e edição
Classificação que valoriza boa experiência de uso ou boa experiência de leitura	Cuidado e atenção especial ao tamanho do texto, dos parágrafos e das frases. Escrita coesa com uso de conectivos e, sem repetições de frases. Escrita que privilegia fluidez com uso da ordem direta nas sentenças, além de uma hierarquização eficiente de informações.
Processo de escaneamento feito por algoritmos em HTMLs de páginas de conteúdo identificando H1, H2, H3.	Divisão do texto em cabeçalhos e inserção de sumário
Processo de coleta de “descoberta de URLs” parece ser dotado de certa objetividade, considerando que a similaridade ou correspondência textual é determinante.	Criação de URLs literais, curtas e descritivas
Organização da SERP por <i>snippets</i> que podem determinar o clique em uma página	Criação de URLs, títulos e meta descrições literais e atrativas

 Fonte: Elaborado pela autora

5.2.6 Imagem

Plataformas de busca não fazem uma leitura visual de imagens adicionadas a textos verbais do mesmo modo que os leitores fazem. Os rastreadores do *Google* leem o HTML das páginas, portanto, quando houver uma imagem, os marcadores de código indicam o elemento ``. Por isso, ao inserir uma imagem no texto, produtores e editores precisam

descrever o que ela representa para a compreensão algorítmica dos buscadores. Essa descrição é nomeada de texto alternativo ou atributo ALT (*alt-text* ou *alt-tag*), que carrega as informações necessárias sobre a imagem, incluindo a palavra-chave daquele conteúdo.

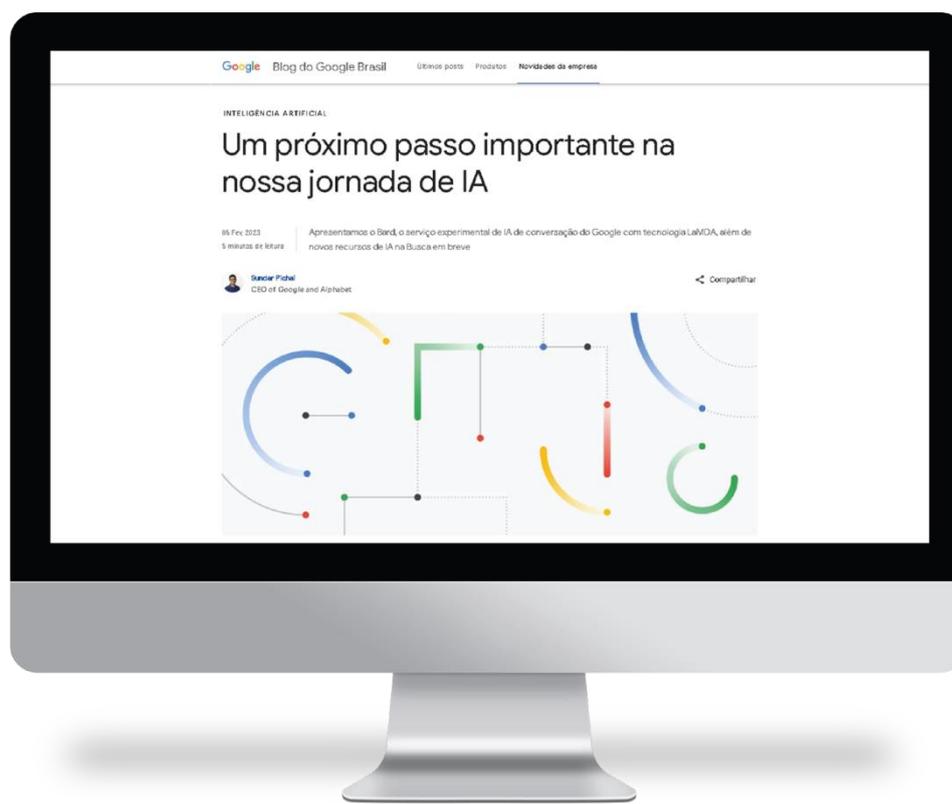
O atributo *alt-tag* é um importante marcador no HTML da página que norteia a leitura de robôs dos mecanismos de busca sobre o propósito das imagens inseridas. Ele não fica visível para leitores no conteúdo, mas, ao ser configurado no CMS, o *alt-tag* aparece nos códigos acessíveis aos algoritmos de plataformas de busca. Nas imagens a seguir, mostramos o atributo *alt-tag* da imagem que está no texto 5, aquele do *blog* do *Google*. A descrição *alt-tag* é exatamente o que vemos na imagem inserida no texto.

Figura 44 – Código-Fonte atributo *alt-tag*/ texto 5

```
<div class="article-image-hero">
  <div class="article-image-hero__container">
    <figure class="article-image--full-aspect article-module">
      <div class="aspect-ratio-image" style="--height-percentage: 41.7%;">
        <div class="aspect-ratio-image__container">
          
        </div>
      </div>
    </div>
  </div>
</div>
```

Fonte: *Blog do Google*, 2023.

Figura 45 – Imagem do texto 5



Fonte: *Blog do Google*, 2023.

Além de ajudar na leitura algrítmica, o atributo ALT auxilia aplicativos leitores de tela que ajudam pessoas com deficiência visual a usar a internet, lendo em voz alta o que está nos *sites*. No exemplo da Figura 44, os aplicativos de acessibilidade leriam a descrição “Ilustração com formas que escrevem a palavra *Google*” com o texto.

O *Yoast SEO* e *Rank Math* vão ajudar esses produtores indicando a necessidade de inserção dessas descrições. Ao analisar essa subcategoria, verificamos que as imagens são elementos fundamentais para tornar um texto encontrável, porque um leitor poderá chegar a uma postagem por meio da imagem. Para isso, é necessário que os algoritmos entendam textualmente o propósito daquela imagem para elegê-la como resposta na SERP. Especialistas em SEO acreditam que o *Google* atribui um valor relativamente alto ao texto alternativo. É por isso que os *plugins* fazem a verificação em duas etapas: a) se há imagens em sua postagem; e b) se essas imagens têm um texto alternativo com a frase-chave em foco.

No processo de análise, consultamos os códigos-fonte de todos os textos do *corpus* para saber se havia atributo ALT para as imagens na publicação original. Para os que já possuíam o descritivo, mantivemos a mesma forma textual no espaço simulado. Para os textos sem atributo ALT, criamos os descritivos de acordo com as recomendações dos *plugins*

Yoast SEO e *Rank Math*. O texto 1, por exemplo, foi publicado originalmente com 31 imagens sem atributo ALT, portanto, ao transportarmos para o espaço simulado, foi necessário criar uma descrição para cada imagem, a fim de alcançar parametrização adequada. No *Yoast SEO*, começamos com a estratégia de inserir um mesmo atributo ALT para todas as imagens, mas o *plugin* apontou que esta não era a melhor escolha, assim, iniciamos a descrição detalhada para cada figura, para melhorar a avaliação no *Yoast*. Algumas das frases usadas foram “reprodução de conversa entre funcionário da agência de *marketing*, contratada pelo *iFood*, em grupo do *WhatsApp*”, “reprodução de página do *Facebook* ‘Não Brega meu Trampo’”, “reprodução de página do *Facebook* ‘Garfo na Caveira’”.

Quadro 191 – Entrelaçamento entre plataforma de busca e escrita na *Web* (categoria/ imagem)

Mediações algorítmicas e/ou diretrizes do <i>Google</i>	Práticas incorporadas na escrita e edição
Leitura textual que os algoritmos fazem, pois enxergam os códigos da página HTML	Inserção de atributo ALT, marcador descritivo de imagens para a leitura algorítmica

Fonte: Elaborado pela autora, 2023.

Nesta seção 5.2, observamos por meio da testagem com o *corpus*, aspectos da linguagem, da construção e edição textual em ambiente digital na atualidade. Se nossas simulações fossem processos reais de produção e edição, certamente, os cinco textos seriam publicados em versões bem diferentes daquelas originais. O texto 1 seria mais enxuto e compartimentado. Essa divisão tem como ponto negativo a quebra de fluidez no texto jornalístico, mas como ponto positivo a organização da massa textual que é enorme naquela reportagem. O texto 2 seria mais literal e óbvio, desviando, talvez, o objetivo da crônica como gênero literário. O texto 3 ganharia coesão e coerência, ficando mais compreensível e potente para a função *release* de divulgação científica. O texto 4 continuaria cumprindo um papel quase burocrático de informar sobre “prospecção do cliente”, sem muita originalidade e atendendo às normatizações para performar. Como apontamos anteriormente, há indícios de que o texto 4 já tenha sido criado dentro de um planejamento editorial voltado para visibilidade em plataformas de busca. Por fim, o texto 5 seria ainda mais objetivo e, ironicamente, mesmo sendo do próprio *Google*, ganharia poder de elegibilidade para SERP.

Essas observações nos fazem crer que a dependência de uma plataforma *online*, como *Google*, para geração de tráfego a *sites* e visibilidade de conteúdos acaba por moldar a “produção, circulação e leitura” a partir dos critérios privados e empresariais dessa plataforma. Sendo assim, percebemos que o fenômeno da “plataformização do texto” é composto pelo atravessamento de critérios algorítmicos e normatizações definidas pelas

plataformas de busca na escrita e edição em ambiente digital. Em consequência, há ainda a plataformização do fluxo editorial do texto provocada pela inserção das camadas técnicas – CMSs, *plugins*, motores de busca – que se tornam quase indispensáveis na cadeia de produção textual (aquela com enfoque em visibilidade).

A dinâmica da plataformização do texto é o que faz emergir o texto plataformizado, incrementado em performance e desempenho e amarrado à retórica do mito da relevância construída pelo *Google*. O que não podemos deixar de observar é que nenhum dos textos publicados levou os parâmetros de ranqueamento ao limite final, portanto, há sinais de que esse compromisso de esforço máximo talvez não esteja no foco de produtores e editores. Isso pode ser um sinal, também, de falta de tempo e recursos humanos para despender no processo de parametrização. Um terceiro ponto é que o movimento de limitar a parametrização pode, por si só, sinalizar algo de subversivo à plataformização, mas voltaremos a discutir subversão mais à frente.

5.3 IMPLICAÇÕES DA RECONFIGURAÇÃO DAS PRÁTICAS DE ESCRITA E EDIÇÃO

A metodologia de espaço simulado de escrita e edição que utilizamos na etapa 2 da pesquisa não foi apenas uma operação instrumental com *plugins*, mas um procedimento que tornou visível a realidade existente na produção de texto em ambiente digital. Além disso, esse procedimento fez emergir leituras e visões de mundo que entrelaçam datificação, mediações algorítmicas, políticas de governança da plataforma de busca *Google* à produção textual na *Web*. Uma das leituras mais óbvias (mas não menos instigante) é de que a natureza contingente do texto plataformizado surge a partir das parametrizações que visam ao ranqueamento. Desse modo, os algoritmos do *Google*, os *plugins* que emulam esses algoritmos e os produtores/editores que avaliam as diretrizes de alteração no texto interferem no fluxo editorial como agentes da textualidade. A inserção de emuladores na cadeia de produção textual é também um pilar da plataformização do texto e que nos faz perceber o quanto o *Google* influencia e quer moldar a escrita em ambiente digital. Percebemos que a tentativa de moldar não é uma “camisa de força” porque nem todos os textos selecionados para o *corpus* estavam parametrizados ao limite, ou seja, nem todo produtor/editor está preocupado com o ranqueamento, mas muitos estão. A seguir, vamos pontuar e analisar algumas implicações dessas camadas de interferência, especialmente, para produtores que escolhem parametrizar em nome da performance.

5.3.1 Incorporação generalizada de técnicas de SEO

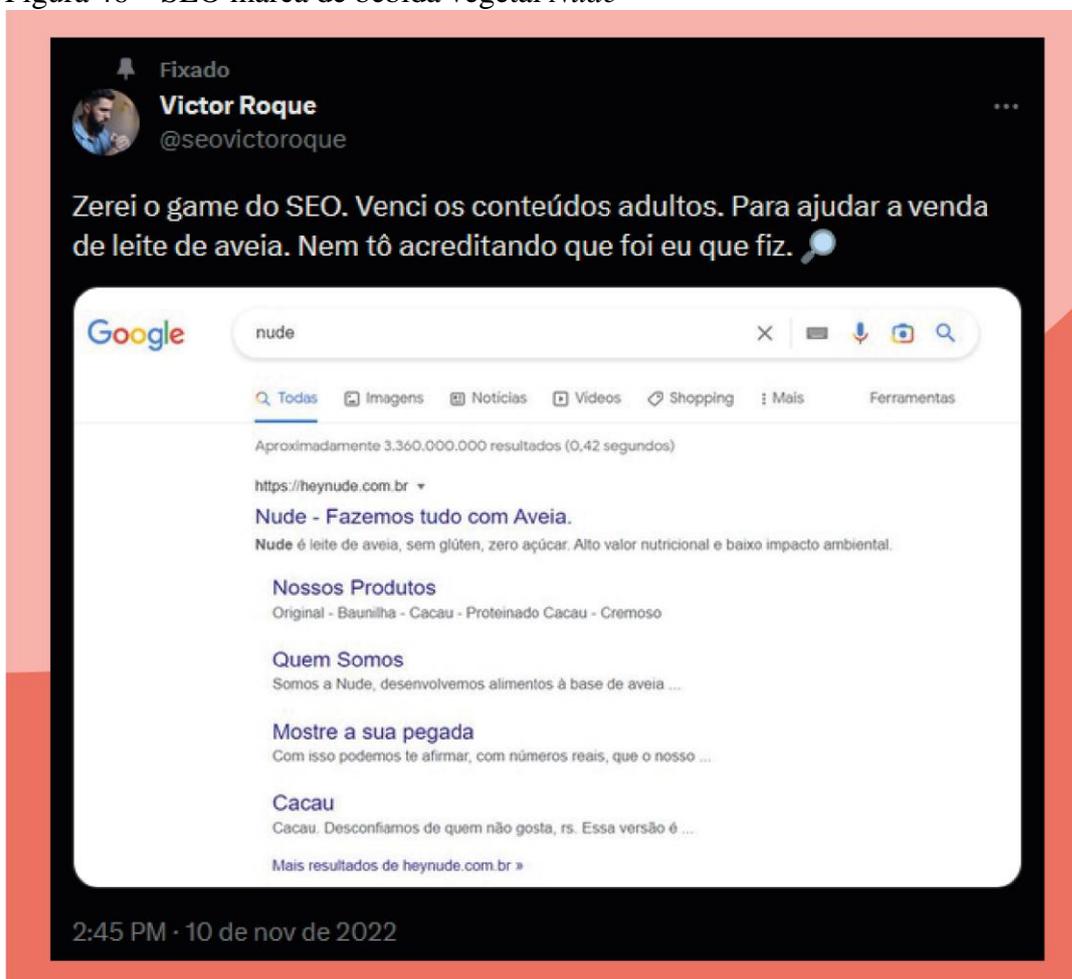
Não é possível nos mantermos ingênuos em relação ao grande negócio que se formou na luta pelos primeiros lugares do *ranking* do *Google*. Martha Gabriel (2010) nomeia essa corrida de “ditadura do top 10”, que motiva a incorporação generalizada de técnicas de SEO, como pudemos perceber nos testes com o *corpus* em que algumas sugestões dos *plugins* soaram exacerbadas. Por exemplo, o *Rank Math* sugerir frequentemente a inserção de números no título das publicações, visando a encontrabilidade delas. Isso poderia ter levado nossas testagens a criações forçadas como: “Conheça as 5 estratégias da máquina oculta da propaganda do iFood”; “Qualquer coisa está no *YouTube*: entenda 5 motivos disso nos salvar a vida; e “Saiba tudo sobre prospecção de clientes e aprenda os 7 métodos mais eficazes”.

A decisão de incorporar sugestões de *plugins* e gabaritar técnicas de SEO nos faz pensar na ambivalência de produtores e editores de texto: usando o planejamento de termos e as técnicas de SEO parecem buscar um caminho tático e, por vezes, estratégico (porque são ações institucionalizadas) que vence a caixa-preta do *Google* e consegue ranquear. Em contrapartida, parecem domesticados à lógica algorítmica da plataforma de busca, porque seguem exatamente o que o *Google* traça como diretriz ideal para uma boa classificação na SERP. Vale lembrar que termo “caixa-preta”, segundo d’Andréa (2020), serve ainda para se referir ao conjunto de associações que, em um dado momento, não se caracteriza por disputas ou controvérsias entre os atores, mas, sim, por um certo consenso entre eles. O exemplo anterior, de a inserção de números no título das publicações, nos parece um uso exagerado, que condiciona os títulos a um template, mas observamos a circulação de muitos textos – de vários gêneros - com essa estratégia em ambiente digital.

Na Figura 46, mostramos um caso curioso em que o estrategista de SEO da marca de bebida vegetal *Nude* comemora o primeiro lugar no *ranking* do *Google*, otimizando o *site* da empresa para o termo “nude”. Essa palavra-chave certamente tem grande volume de busca para o significado de “pessoas despidas” ou de “cor, tom, coloração”, no entanto, a marca de leite de aveia consegue se posicionar, vencendo a corrida pela ratificação do termo. É como se a empresa de bebida vegetal tivesse ganhado a competição por atenção daqueles que buscam por “nude”, independente do sentido e do idioma, porque o termo tem quase os mesmos sentidos em português e inglês. Há pelo menos duas perspectivas de observarmos esse caso, ambas conectam Comunicação e Linguística: 1) como uma marca nova no mercado – fundada em 2020 –, a *Nude* pode ter surgido com um *branding* pensado para explorar volume de busca nas plataformas *online* de forma casada à identidade, integridade e imagem dos produtos; 2) a

marca propositalmente acrescenta polissemia ao termo “nude” – antes conhecido por apenas dois significados – com intuito de ser lembrada justamente por esses múltiplos sentidos.

Figura 46 – SEO marca de bebida vegetal *Nude*



Fonte: Twitter, 10 nov. 2022.

Esse trabalho para acrescentar uma polissemia tão específica e “vencer” no ranqueamento do *Google* certamente envolve não só profissionais do texto. Não conhecemos a metodologia da *Nude*, mas o fato de o estrategista comemorar esse posicionamento (vencer “o game”) prova que a incorporação de práticas de escrita e edição pode soar ao mesmo tempo tática, estratégia e domesticação. Vai depender do ponto de vista em que olhamos. Metaxa e Torres-Eheverry (2017), por exemplo, consideram as práticas de SEO como *spamming*. Os autores colocam os estrategistas em SEO no mesmo patamar de *spammers* que atacaram o modo de funcionamento do *Altavista* e *Lycos*¹⁵⁹, criando centenas de páginas com

¹⁵⁹ *Altavista* foi um buscador, criado em 1995, mas que foi descontinuado em 2013 por não conseguir se sustentar no mercado de motores de busca. O *Lycos* é um buscador, criado em 1994. Ele ainda existe, mas integra a fatia pouco representativa, com menos de 0,50%, do mercado de motores de busca.

a repetição exacerbada de palavras-chave de alto volume de busca para ranquear. Os autores também igualam os estrategistas em SEO aos *spammers* criadores das fazendas de *links*, que são grupos de páginas de *spam* vinculadas umas às outras para parecerem populares ou dotadas de autoridade e, portanto, subir nas classificações do *Google*.

Enfim, quanto mais profissionais do texto e estrategistas de SEO entendem a lógica de funcionamento da plataforma de busca e replicam ações bem-sucedidas para melhorar performance de páginas ou *sites*, conseqüentemente, de sua lucratividade (por meio de ganhos de audiência ou conversões de venda), mais esses profissionais tendem a produzir mercadoria cultural contingente. Isso ilustra o que denominamos plataformação do texto já que esses profissionais tornam seus trabalhos alvo da engrenagem capitalista das plataformas *online*. Vale lembrar que “[...] a precarização do trabalho parece andar junto com os usos táticos motivados por performance” (CRUZ; VECCHIO, 2022, p. 9). Em alguma medida, a busca pelo ranqueamento, a parametrização textual, o uso ostensivo de *plugins* e técnicas de otimização, acabam por precarizar a o trabalho de profissionais do texto. Fica, então, reforçada a dinâmica capitalista que controla recursos, produtos, trabalho, subjetividades, conhecimento e comunicação. O contingenciamento pode refletir até em inversão de valores da posição de trabalhadores para empreendedores de si, colocando essas pessoas na posição de responsáveis pelo desempenho de suas produções, o que pode ser muito opressor.

5.3.2 Homogeneização e surgimento de gêneros textuais

Entender o processo de produção de textos atravessado pelas lógicas algorítmicas, muda nosso olhar para os tipos de textos que circulam na atualidade e nos faz perceber que otimizar conteúdos para ranquear no *Google* implica esquemas estéticos que criam certa homogeneização textual. Essa estética é mais um dos exemplos da dinâmica envolvida na plataformação do texto institucionalizado pelo *Google*. Se os textos fora do *template* para performance – que inclui parametrização de palavras-chave, hierarquização de cabeçalhos, títulos descritivos, entre outros – não serão ranqueados, logo, produtores e editores contingenciam suas produções para formatos que desempenham melhor. Ou seja, passam a produzir formatos mais populares, como já argumentamos ao citar o conceito de “*templatability*” (LEAVER; HIGHFIELD; ABIDIN, 2020).

Na simulação com o *corpus* percebemos várias sugestões do *Rank Math* e do *Yoast SEO* que, quando aceitas, encaminham a edição dos textos para estéticas pasteurizadas. No caso do texto 2, que era uma crônica, a inserção de palavras de transição

retirou a personalidade do texto, colocando-o em um lugar mais óbvio e comum. Em todos os textos do *corpus*, a sugestão apresentada pelos *plugins* de reduzir o tamanho das sentenças também resultou em um enxugamento forçado da escrita, a fim de fazer caber em uma fôrma. Por isso, afirmamos que, no caso específico de textos para *Web* que dependem da seleção do *Google*, as dinâmicas oferecidas pelos CMS e os *plugins* contribuem para homogeneização. Por exemplo, consideremos os 43% de todos os *sites* da *Web* que usam o *WordPress.com*. Esse número tão representativo de *sites* está suscetível ao uso dos *plugins* disponíveis no sistema de gerenciamento. Ao escolher esses *softwares*, o trabalho de produtores de texto e desenvolvedores *Web* é suportado e limitado pelo que o CMS pode fazer (mesmo nos casos personalizáveis, em que o céu não é o limite). A escolha pelo *WordPress.com* passa pelo fato de que ele oferece opções de estruturar páginas e *sites* para performance no *Google*, como *layouts* amigáveis aos motores de busca, condições de construção de páginas com rápido carregamento, possibilidades de estruturação de URLs e facilidade na inserção de *links*, que são elementos importantes para otimização de conteúdo. Ou seja, são muitos profissionais submetendo suas produções a apenas um CMS e com objetivo de performar em apenas uma plataforma de busca, dependência que ilustra o fenômeno da plataformização do texto. Bogost e Montfort (2009) afirmam que a influência dessa submissão às vezes é óbvia: uma plataforma monocromática não pode exibir cores, um console de videogame sem teclado não pode aceitar entrada digitada. Portanto, a diversidade de entrega pode ser reduzida de várias maneiras.

Como argumentamos anteriormente, *affordances* – que concentram as relações e experiências dos usuários com os artefatos tecnológicos projetados/ofertados pelos sistemas – determinam as possibilidades de ação das pessoas. Isso nos faz retomar a Figura 1, em que apresentamos hipoteticamente as camadas do processo de edição de texto no contexto da plataformização, e perceber que a relação complexa entre essas camadas reduz a diversidade textual. Em nossa análise do *corpus*, vimos *WordPress.com* interagindo de modo complexo com seus *plugins* que, por sua vez, agem emulando algoritmos do *Google* para, enfim, conseguir tornar páginas e *sites* elegíveis.

As consequências dessa homogeneização estão associadas ao reforço de determinados perfis textuais, certas configurações de identidade das produções, como fruto da insistência na produção a partir das lógicas de mediação do *Google*. Magda Soares (2002), em uma retomada histórica sobre a escrita e a leitura, afirma que a tecnologia impressa, inaugurada por Gutenberg, “enformou” a escrita, retirando a liberdade do manuscrito. Segundo a autora, aquela ruptura tornou o texto mais estável, monumental e controlado. Já o

ambiente digital, ao contrário do impresso, desestabilizou esse texto, porque tal como os profissionais interferiam na produção, também os leitores podem acrescentar, alterar, definir seus próprios caminhos de leitura. Além disso, o texto perdeu o caráter monumental porque, como consequência de sua não estabilidade, tornou-se fugaz e impermanente. Por fim, Magda Soares (2002) argumenta que texto em ambiente digital se tornou pouco controlado porque é grande o *continuum* de produção e quase ausente o controle da qualidade e conveniência do que é produzido e difundido. Concordamos com a autora sobre a fixidez dar lugar à fluidez, mas acrescentamos que o fato de o texto ser menos estável não o impossibilita de ser “enformado”, afinal, a dinâmica da visibilidade em plataformas de busca está criando formas/fórmulas/formatos. Continuar o raciocínio proposto por Magda Soares (2002) talvez seja dizer que, no decorrer da história, o texto foi “enformado” e “desenformado” de modo quase cíclico, de acordo com demandas de circulação em seu tempo. E, claro, a referência à lógica cíclica aqui também nos remete ao diagrama sobre “movimentos da cultura escrita”, de Ribeiro (2018), ilustrado na Figura 2.

Estamos falando muito, na atualidade, sobre a inserção de IA na escrita, em usos educacionais e profissionais. Refletindo sobre homogeneização da produção textual, Ribeiro (2023) argumenta que fazer “textos formulaicos” torna nossos processos de escrita mais suscetíveis ao atravessamento pela IA. Quando redatores, estudantes, professores, escritores, jornalistas não dominam o texto e as técnicas de escrita, transferindo isso para *softwares*, *plugins* e aplicativos de IA com objetivo de agilizar processos, perde-se uma oportunidade de aprendizado. Quanto menos criativo é um texto e quanto mais “formulaico” (RIBEIRO, 2023), melhor é para máquinas, algoritmos, plataformas e IA. Isso nos faz refletir se os redatores estão fazendo, de modo totalmente consciente, essas escolhas de acelerar, facilitar, otimizar a escrita. O *ChatGPT* ganhou protagonismos com a ideia fatalista de que as IAs vão substituir quem trabalha com escrita e edição, mas estamos há anos moldando e homogeneizando nossos textos para ranquear no *Google*. Ninguém olhou criticamente para isso como um movimento em que nós mesmos permitimos a ascensão de textos “formulaicos”, pasteurizados e homogêneos. E que isso nos coloca em situação complicadíssima agora que as IAs conseguem fazer o trabalho pasteurizado por nós.

A raiz da diversidade reduzida de textos e conteúdos em circulação não está na capacidade dos usuários de controlar seus ambientes de informação – indicando preferências ou usando as plataformas instintivamente –, mas, em grande medida, na sua incapacidade de fazê-lo. Isso corrobora o que Martha Gabriel (2010) chamou de uma inevitável era da busca que acabou dominada pela extrema personalização de informações oferecida pelas

plataformas *online* e que está fora do controle dos usuários. A personalização, da qual não se foge no uso de motores de busca, dá origem a um mundo hiperfiltrado em contraste à rede diversa e hipedescentralizada que poderíamos idealizar. Estranho concluir que haja hiperfiltragem no momento em que temos a impressão de acesso a um bombardeio de informações. É que elas realmente são de uma grandeza imensurável, o que nos faz crer na necessidade de uma curadoria automatizada e capaz de processar em grande escala, como a curadoria algorítmica.

Outro lado da homogeneização é o surgimento de gêneros textuais que hibridizam gêneros já existentes, mas acoplando a perspectiva da performance em plataformas de busca. O *Google* é capaz de projetar estratégias de textualização que geram novos gêneros. Para relativizar essa afirmação, sem incorrer em determinismo tecnológico, recorremos a Marcuschi (2002). Para o autor, não são as tecnologias por si que originam gêneros, e, sim, a intensidade dos usos dessas tecnologias e suas interferências nas atividades comunicativas. Em outro texto, Marcuschi (2003) afirma a existência de forte relação entre funções e formatos de um gênero textual ao seu suporte, o que serve muito para pensar o texto plataformizado. Alinhando-se à perspectiva de Marcuschi (2002), Juliana Assis (2002) considera que textos agrupados sob um mesmo gênero comungam uma base de princípios que orientam sua materialização. Sendo assim, compartilham um mesmo núcleo de funções e propósitos comunicativos e ainda têm sua materialização marcada pelas condições de funcionamento da tecnologia, meio de que dependem. Esse aspecto do funcionamento corrobora nossas argumentações anteriores sobre a interação complexa de *affordances* de CMS e plataformas de busca, como parte do fenômeno causador da plataformização do texto.

O gênero textual *marketing* de conteúdo, que descrevemos em estudo anterior (CRUZ; RIBEIRO, 2016), por exemplo, surfou na necessidade de criação de um modelo menos interruptivo de publicidade, por conta das mudanças de comportamento de consumidores. Os textos feitos para *marketing* inflaram a *Web* a partir de processos de redação que geram conteúdos padronizados e, em tese, eficazes para a conversão da leitura em alguma forma de consumo. Esse tipo de texto mistura o caráter informativo inspirado no jornalismo com o tom persuasivo da publicidade, mas servindo objetivamente a uma estratégia de venda. Ademais, têm formas/fórmulas/formatos adequados ao ranqueamento, porque são produzidos com parametrizações muito alinhadas às técnicas de SEO e ao que é popular na internet.

A era dos textos “enformados” de *marketing* de conteúdo persiste, mas outras produções – embaladas de mais qualidade textual – vão ganhando espaço. Muitas marcas,

empresas e organizações, atualmente, buscando a construção de autoridade digital (EEAT) e compreendendo o protagonismo do conteúdo, passaram a disputar espaços de produção de texto. Algumas foram mais ousadas e criaram seus próprios espaços de produção, movimento conhecido como *brandpublishing*. Por exemplo, a *MRV&Co* criou o *Habitability*¹⁶⁰, um portal que “busca promover o conhecimento e a reflexão sobre o futuro do habitar” (MRV&C, 2023), por meio da publicação de notícias, entrevistas e artigos de opinião de especialistas sobre o tema. Um olhar rápido pelo portal permite ver textos cuidadosamente parametrizados, intitutados e escaneáveis para o ranqueamento em plataformas de busca. O mesmo movimento fez o banco *Inter* ao criar, em 2021, o *Inset*, seu próprio portal de notícias com foco em temas de economia e mercado financeiro. Mas o portal durou menos de dois anos. Como terceiro exemplo, está a *Magazine Luiza* (MGLU3), que comprou, em 2021, o *Jovem Nerd*¹⁶¹, a maior plataforma multimídia voltada para o público nerd e *geek* do Brasil. O grupo empresarial se tornou proprietário de mídias ocupando espaços de produção de conteúdo especializado. Esse movimento reforça uma estratégia de conteúdo digital que visa não só aumentar o relacionamento dos consumidores com a marca, mas se destacar como autoridade em alguns setores. A *Magazine Luiza* ser dona de uma plataforma multimídia que tem público cativo, bom ranqueamento no *Google* e estratégia consolidada de produção frequente de textos, contribui muito para o *brandpublishing*. No *Jovem Nerd*, há influenciadores capazes de ajudar consumidores na tomada de decisão de compras de produtos do universo *Magazine Luiza*.

E o que tudo isso tem a ver com homogeneização? É que temos cada vez mais *hubs* de produção de texto que visam à performance desses conteúdos com fins lucrativos. Logo, o texto ganha protagonismo como estratégia comercial, sendo elevado a ferramenta para conseguir vender. Essa dinâmica alimenta as formas/fórmulas/formatos criadas para ter visibilidade em ambiente digital.

Outro exemplo da homogeneização textual é o sucesso do formato *listicles*, que são textos planejados para viralizar na *Web*. A ideia é que sejam propagáveis (JENKINS, 2014)¹⁶², principalmente em redes sociais, com o objetivo de alcançar audiência. São produções que reúnem informação e entretenimento, por meio de dois gêneros textuais já conhecidos: lista + artigo (*list+article= listicle*) (SARDÁ et al., 2014). Os temas que pautam

¹⁶⁰ Disponível em: <https://habitability.com.br/>. Acesso em: 04 jul. 2023.

¹⁶¹ Disponível em: <https://jovemnerd.com.br/>. Acesso em: 04 jul. 2023.

¹⁶² “A propagabilidade se refere ao potencial técnico e cultural de públicos compartilharem conteúdos por motivos próprios, às vezes com permissão dos detentores dos direitos autorais, às vezes contra o desejo deles”. (JENKINS, 2014, p. 26).

a escrita são geralmente tendências de busca no período da produção, os chamados *trends*. Um dos principais produtores desse modelo é o *BuzzFeed*¹⁶³, que se autodefine como líder de mídia digital independente, levando notícias e entretenimento para centenas de milhões de pessoas em todo o mundo. Os *listicles* são propagáveis justamente porque abraçam as técnicas de SEO, tendo na própria concepção do gênero a padronização hermética de parágrafos curtos (legíveis e escaneáveis). Sardá et al. (2014) analisaram a apropriação dos *listicles* pelo jornalismo, fenômeno que denominaram “buzzfeedização do jornalismo”, para exemplificar como o formato é usado como estratégia coincidente entre a escrita que atende a relevância algorítmica e a escrita que preza pela facilitação da leitura humana.

Poderíamos pensar que o surgimento de gêneros textuais – como citamos o *marketing* de conteúdo e o *listicle* – contribui, na verdade, para heterogeneização, mas aqui argumentamos a homogeneização, porque os tipos que surgiram inundaram a *Web* em grande escala, fazendo parecer que só poderíamos produzir, circular e ler nesses moldes. Entendemos que gêneros textuais evoluem e nascem com as mudanças na cultura, inclusive as de origem tecnológica, mas o que nos incomoda é o fato de as plataformas de busca – com seu processo de ordenação – fazerem circular determinados gêneros textuais em detrimento de outros. O *Google* – como agente principal da plataformização do texto – premia (ranqueia) os *sites* e conteúdos que estão dentro de um padrão e, de modo paradoxal, valoriza qualidade e frequência de publicação. Isso acaba por configurar um ciclo infinito de contingenciamento da produção textual que tem como ponto de partida, quase sempre, os critérios básicos de plataformas de busca (popularidade e similaridade). Demonstramos este contingenciamento, de modo resumido, no Quadro 21 a seguir:

Quadro 202 – Critérios de plataformas de busca x lógicas de produção textual

Critérios de plataformas de busca	Lógicas de produção no contexto da plataformização do texto
Popularidade (que calcula ou contabiliza desempenho/performance)	performance, otimização (SEO), métricas de taxa de clique e engajamento, “jornalista metrificado” (KALSING, 2021), esforços de edição para alcançar o maior melhor ranqueamento, redator/autor “empreendedor de si” para superar metas, texto como “mercadoria cultura contingente” às demandas definidas por cálculos algorítmicos, transcender questões éticas para privilegiar desempenho e audiência.
Similaridade (que reúne semelhanças, correspondências,	repetição (produzir texto que o leitor já viu, já clicou ou já curtiu), texto que reiteram normas e atendem a

¹⁶³ Disponível em: <https://www.buzzfeed.com/about>. Acesso em: 8 abr. 2022.

Critérios de plataformas de busca	Lógicas de produção no contexto da plataformização do texto
correlações estatísticas entre as ações e características dos usuários, com o objetivo de exibir itens que correspondam aos gostos, opiniões e preferências)	templates, “ <i>templatability</i> ” (LEAVER; HIGHFIELD; ABIDIN, 2020), parametrização textual para atender correspondências encontradas pelos algoritmos do <i>Google</i> , operações de perfilamento (CASTRO, 2019) em que quanto mais eficiente se revela mais é reproduzido e maior é sua eficácia na construção de perfis do leitor; insistência em produzir gêneros textuais específicos e similares, que performam melhor.

Fonte: Elaborado pela autora, 2023.

Enfim, a incorporação de práticas de escrita e edição para atender à popularidade e à similaridade propicia a circulação de textos semelhantes na sua aparência e que buscam equivalência na autoridade, a fim de serem vistos como relevantes pelo *Google*. Mais uma vez, parece que nos frustramos sobre construir aquela internet diversa, coletiva, colaborativa, participativa, porque percebemos os critérios privados e empresariais moldando a “produção, circulação e leitura” de textos em ambiente digital, o que é parte do fenômeno de plataformização do texto.

Todavia, os produtores e editores trabalham de modo ambivalente – às vezes, para se adequar às parametrizações textuais impostas pela caixa-preta das plataformas de busca e, em outros momentos, para subvertê-las. Esse trabalho se torna, por vezes, precarizado, especialmente, quando olhamos para as práticas relatadas nesta pesquisa e que se relacionam ao jornalismo (os jornalistas metrificados e a lógica de trabalho implicado por robôs). Essa precarização certamente se relaciona à grande concentração de poder em poucas plataformas *online*, o que pretendemos deixar claro em nossa argumentação sobre o *Google*. Esse poder, no entanto, é exercido de modo contingente, deixando certa permeabilidade às demandas dos públicos, dos produtores e das pressões regulatórias. Daí a importância de buscarmos a compreensão da ambivalência na produção textual e na possibilidade de diversos usos que ativistas, especialistas, instituições, *bots* e outros atores fazem dos recursos técnicos e das regras vinculados a cada plataforma *online*. É sobre isso que dissertaremos na próxima seção.

5.4 PRÁTICAS SUBVERSIVAS: INDÍCIOS E CAMINHOS PARA RESISTÊNCIAS À CURADORIA ALGORÍTMICA

Quando idealizamos a construção desta pesquisa, a seção reservada para a descrição de práticas subversivas talvez tivesse mais protagonismo. No entanto, percebemos

que era preciso consolidar a verificação da reconfiguração de práticas de escrita e edição antes de ir em busca dos movimentos de contrafluxo, que, aliás, são pungentes. Citar esses movimentos que rompem é uma forma de encerrar um ciclo narrativo desta pesquisa, que foge do fatalismo, mostrando que há caminhos em construção. Hissa (2012) diz que, quando se exclui a utopia, empobrecemo-nos imediatamente, sendo ela, a utopia, uma referência movente que nos faz caminhar e, quem sabe um dia, seja uma conquista, e não uma referência. Muito antes das iniciativas que vamos exemplificar aqui, no decorrer da pesquisa, falamos de movimentos *spamming*, *Google Bomb*, das derrotas de plataformas na Justiça encabeçadas por categorias profissionais (caso *Googlebooks*) e das iniciativas de regulação. Defendemos que o “quase monopólio” capitalista do Google e a centralização causada por outras plataformas devem ser combatidos pelas instâncias de tomada de decisão, mas também acreditamos em outros movimentos que semeiam a subversão. Como leitores podemos desempenhar um papel importante de resistência, experimentando a internet de pontos de vista diferentes para refletir como escolhas algorítmicas e regimes geopolíticos moldam nossa vida cotidiana.

No mercado de buscadores horizontais – que indexam as informações da *Web* de forma universal e dos quais fazem parte *Google* e *Bing* –, é possível usar alternativas que garantem mais privacidade, rompem com modelos de negócio exclusivamente mercantilistas ou dão opções de entrega mais personalizáveis. Por exemplo, na mesma linha do *DuckDuckGo*, que citamos no Capítulo 4, existe o *Mojeek*¹⁶⁴. A empresa afirma ser o primeiro mecanismo de pesquisa sem rastreamento de dados, o que garante privacidade. O *Mojeek* tem seu próprio índice, sem precisar depender do *Bing* e do *Google* para obter resultados. Já o *Ecosia*¹⁶⁵, redireciona 80% de sua receita excedente a organizações sem fins lucrativos focadas em reflorestamento. Segundo a empresa, ao realizar 45 buscas na plataforma, um usuário “planta” uma muda. A base de busca do *Ecosia* é fornecida pelo *Bing* (*Microsoft*), ou seja, não há contrafluxo na lógica de curadoria e seleção, mas optar pelo *Ecosia* é escolher por um negócio de impacto social positivo. Outro exemplo é o *Searx*¹⁶⁶, que dá a possibilidade de personalização dos mecanismos de busca de onde os resultados serão extraídos e ainda dá chance de categorizar resultados de pesquisa. Nesse caso, só não é possível escolher como origem de índice o *Google*, que bloqueou as solicitações de pesquisa do *Searx*.

¹⁶⁴ Disponível em: <https://www.mojeek.com/>. Acesso em: 02 ago. 2023.

¹⁶⁵ Disponível em: <https://www.ecosia.org/>. Acesso em: 02 ago. 2023.

¹⁶⁶ Disponível em: <https://searx.thegpm.org/>. Acesso em: 02 ago. 2023.

Fugindo da tentação de sucumbir aos buscadores horizontais – que se posicionam como oráculos digitais –, há um movimento de busca que pode ser muito potente: o uso de buscadores verticais, que têm foco especializado e indexam informações referentes apenas a uma área específica. Não que esta seja uma escolha exatamente subversiva, porque muitos desses buscadores pertencem ao grupo *Big Five*, mas é um modo de compartimentar a demanda e acessar resultados mais específicos. Nesse caso, há dezenas de exemplos em categorias de buscador de voos e viagens, artigos acadêmicos (inclusive por área do conhecimento), imagens, vídeos, músicas e áudios, código de programação, entre outros. O uso de outras plataformas *online* como buscadores causa um efeito inusitado, por exemplo, quando constatamos que a segunda barra de busca com maior volume de pesquisa é a do *YouTube*¹⁶⁷ e a terceira, a da *Amazon*. Entre os jovens de 18 a 24 anos, a barra de busca do *TikTok* já ultrapassou o *Google* em volume de pesquisas¹⁶⁸.

Outras possibilidades são os buscadores semânticos, como *Wolfram Alpha*¹⁶⁹, que respondem às perguntas dos usuários em linguagem natural e sem se basear em palavras-chave, o que pode parecer nada diferente do que o *Google* faz. No entanto, a distinção está no fato de que os buscadores semânticos não estão preocupados em rastrear *links* (como os buscadores horizontais), e sim em procurar informações, sendo muito útil, portanto, a profissionais que precisam lidar com dados e informações que podem ser quantificadas. Se cada um de nós buscarmos por nossos nomes e sobrenomes no *Google*, ele entregará *links* ratificados que contêm a similaridade com o nome. Provavelmente aparecerão perfis de plataformas de rede sociais, textos de nossas autorias em *sites*, lista de livros que escrevemos, verbetes da *Wikipédia*, notícias, nossos currículos ou classificações em concursos já prestados, entre outros *links* que tentarão dizer algo sobre quem aquele nome representa. Se buscarmos por nossos nomes e sobrenomes no *Wolfram Alpha*, ele entenderá como uma recorrência de “nome dado a mulheres” ou “nome dado para homens” e listará quantas pessoas registradas com esse nome, o histórico de nascimentos, a etimologia, a idade mais comum de pessoas com esse registro, entre outros dados.

Além dessas perspectivas de usos alternativos de buscadores, podemos citar iniciativas que escancaram a engrenagem das plataformas *online*, levando a conhecimento de públicos mais amplos. Há o emblemático caso da *Folha de S.Paulo*, que envolve as

¹⁶⁷ Disponível em: <https://www.thinkwithgoogle.com/intl/pt-br/estrategias-de-marketing/video/impacto-economico-cultural-social-youtube-brasil/>. Acesso em: 10 jul. 2023.

¹⁶⁸ Disponível em: <https://veja.abril.com.br/tecnologia/tiktok-substitui-o-google-como-ferramenta-de-busca-entre-os-mais-jovens>. Acesso em: 10 jul. 2023.

¹⁶⁹ Disponível em: <https://www.wolframalpha.com/>. Acesso em: 10 jul. 2023.

plataformas *Facebook* e *Google*. O caso mostra indícios de resistência à lógica de plataformas *online*, mas, ao olhar com mais profundidade, as ações revelam tentativas de sobrevivência no cenário de implosão de seu modelo de negócios.

Em 8 de janeiro de 2018¹⁷⁰, a *Folha* anunciou a suspensão de publicações no *Facebook*, alegando que a plataforma havia se tornado um lugar desvantajoso para circular notícias. O veículo de imprensa afirmou buscar “[...] os melhores caminhos para fazer com que o conteúdo do jornal chegue aos seus leitores” e disse que, após as mudanças algorítmicas – que realmente aconteceram e passaram a priorizar a interação pessoal no *feed* de notícias e reduziram o alcance de *posts* jornalísticos –, ficaram evidentes as “[...] as desvantagens em utilizar o Facebook como um caminho”.

A *Folha de S.Paulo* ainda usou como argumento a acusação de que a nova regra do *Facebook* poderia “[...] diminuir a visibilidade do jornalismo profissional nas páginas de seus usuários”, portanto, “[...] favorecendo a criação de bolhas de opiniões e convicções, e a propagação das ‘fake news’”. Portanto, em um primeiro olhar, a *Folha de S.Paulo* mostrou resistir à lógica de seleção e filtragem do *Facebook*. Naquela época, o veículo foi acusado por muitos de desistir daquele ambiente em um momento importante. Era período pré-eleitoral e a desinformação inundou o Brasil, quando muitos entendiam que a presença da *Folha* e outros jornais presentes na plataforma de rede social poderiam ser um recurso de combate às notícias falsas.

Todavia, quando olhamos para o movimento da *Folha* com mais criticidade, percebemos que a decisão da empresa também se relaciona com a encruzilhada do jornalismo como produto cultural e comercial no contexto da plataformização. O veículo de imprensa sinalizou que “[...] a importância do Facebook como canal de distribuição já vinha diminuindo significativamente”, antes mesmo da mudança algorítmica. Nesse ponto, a *Folha de S.Paulo* argumentou numericamente a queda de origem do tráfego de jornais naquele ano (2018). Em janeiro, o volume total de interações (compartilhamentos, comentários e curtidas) obtido pelas 10 maiores páginas de jornais brasileiros no *Facebook* caiu 32% na comparação com o mesmo mês do ano anterior (2017), segundo dados compilados pela *Folha* em pesquisas da *Parse.ly*, empresa de pesquisa e análise de audiência digital. Com a queda do alcance das páginas, o *Facebook* já vinha perdendo espaço como fonte de acessos a *sites* de

¹⁷⁰ Folha deixa de publicar conteúdo no Facebook. Disponível em: <https://www1.folha.uol.com.br/poder/2018/02/folha-deixa-de-publicar-conteudo-no-facebook.shtm>. Acesso em: 07 jul. 2023.

jornalismo. A participação da rede social nos acessos externos caiu de 39% em janeiro de 2020 para 24% em dezembro de 2021.

De acordo com a própria *Folha de S.Paulo*, a origem de tráfego vinha sendo ocupada principalmente pelos mecanismos de busca, como o *Google*, que no mesmo período avançou de 34% para 45%. E aqui está o ponto de atenção. No momento em que o *Facebook* deixa de ser uma importante origem de tráfego (fonte de audiência, que gira a engrenagem jornalística), a *Folha* decide suspender as publicações. Não coincidentemente, é nesse mesmo momento que a empresa investe grande fôlego na contratação de equipes de SEO para profissionalizar suas estratégias de ranqueamento.

O desfecho desse caso é que em 5 de julho de 2021¹⁷¹, a *Folha de S.Paulo* voltou atrás e retomou suas publicações no *Facebook*. O argumento central foi, novamente, a questão das notícias falsas, pois o veículo de imprensa reconhece a mudança de atitude da plataforma e os novos esforços para combater a desinformação, enumerando uma série de ações: derrubada de páginas que disseminaram *fake news* e discursos de ódio, criação de comissão especializada em direito digital, suspensão de contas falsas, entre outras. Além disso, a *Folha de S.Paulo* enfatizou o investimento do *Facebook* entre 2018 e 2021 na indústria de notícias. Segundo o veículo de imprensa, a empresa de tecnologia destinou R\$ 3 bilhões em treinamentos e incentivos a projetos, pretendendo ainda alocar R\$ 5 bilhões até 2024. A *Folha de S.Paulo* foi contemplada com investimentos do *Facebook*.

Esta ruptura entre *Folha* e *Facebook* indica a resistência estratégica do veículo de comunicação. Dar um basta à curadoria algorítmica foi necessário como posicionamento institucional e aposta e manutenção de credibilidade perante seus públicos. Ficou evidente no discurso da empresa, sua vontade de descortinar mecanismos do *Facebook*, principalmente aqueles que se opõem a ideia de redes sociais como ambiente de convívio e intercâmbio, e que reforçam a lógica de “[...] condomínio fechado das convicções autorreferentes” (FOLHA, 2021) criada pelas plataformas objetivando reduzir o poder dessas corporações. Em contrapartida, a ruptura revela as motivações de fator econômico da *Folha*, afinal, a plataforma *online* que passou a influenciar mais significativamente o modelo de negócio do veículo de comunicação foi o *Google*, ao qual a *Folha* sucumbiu sem resistir e segue encarando – de modo tático, estratégico ou domesticado – com a ostensiva otimização de conteúdos jornalísticos.

¹⁷¹ Folha volta a publicar no Facebook. Disponível em: <https://www1.folha.uol.com.br/poder/2021/07/folha-volta-a-publicar-conteudo-no-facebook.shtml>. Acesso em: 07 jul. 2023.

As ações com indícios de resistência, usos táticos e subversões citadas aqui são importantíssimas, mas as maiores rupturas, transgressões e hackeamentos se originam da possibilidade de projetarmos modelos e algoritmos alternativos que não perpetuam padrões de opressão. Movimentos profissionais e sociais, como #TechWontBuildIt, #NoTechForICE¹⁷² #Data4BlackLives¹⁷³, são fundamentais nessa jornada. O primeiro é organizado por trabalhadores da indústria de tecnologia que estão se opondo a decisões eticamente questionáveis tomadas por seus empregadores, porque, muitas vezes, a liderança das empresas de tecnologia vai sacrificar a responsabilidade ética pela promessa de um contrato lucrativo. A *hashtags* agrega iniciativas e debates. Já a campanha #NoTechForICE visa criar um futuro livre, incentivando a sociedade civil a recuperar o poder sobre as tecnologias, por meio de ações educativas e de fiscalização que apoiam trabalhadores e estudantes no encorajamento contra a influência do Vale do Silício. O terceiro movimento, o #Data4BlackLives, é composto por cientistas e ativistas comprometidos com a missão de usar dados para criar mudanças concretas e mensuráveis na vida de pessoas negras. O grupo se mobiliza em campanhas, conferências e eventos, além de produzir e divulgar relatórios que evidenciam o uso discriminatório de dados e algoritmos em todos os sistemas.

Nessa linha de ruptura, Ochigame (2021a) propõe algumas remodelagens no campo da recuperação da informação, usando modelos não ortodoxos de lógica e computação. Segundo o autor, a criação desses sistemas alternativos não é o grande problema, o maior desafio é fazer com que modelos experimentais sejam adotados por grandes públicos. De acordo com Ochigame (2021a), os buscadores comerciais são muito homogêneos não porque representam um único modo de fazer motores de busca, mas porque se ancoram no contexto político-econômico da atualidade. Uma das possibilidades propostas pelo pesquisador é romper com os critérios da popularidade e similaridade usando perspectivas plurais e críticas com o potencial de desafiar, complicar ou desestabilizar os leitores em ambiente digital.

A maioria das ferramentas de pesquisa são baseadas em princípios de correspondência de palavras-chave ou conceitos que classificam os resultados por popularidade e semelhança. Tais ferramentas podem refletir preconceitos e desigualdades, reproduzindo padrões de exclusão e marginalização. Como pode ser a pesquisa se, em vez disso, privilegiar valores do pluralismo e da crítica epistemológica? (OCHIGAME, 2021a, p. 177).

Ochigame (2021a) criou o modelo experimental “Relata”, um sistema de busca alternativo ao *Google Scholar* para procurar referências bibliográficas da área de Humanas,

¹⁷² Disponível em: <https://notechforice.com/about/>. Acesso em: 04 jul. 2023.

¹⁷³ Disponível em: <https://d4bl.org/>. Acesso em: 04 jul. 2023.

especialmente antropologia sociocultural, a partir de parâmetros e relações que vão além da popularidade e similaridade. A versão piloto começou com um banco de dados de 214 trabalhos e a construção desse índice contou com a curadoria humana de membros da *Society for Cultural Anthropology*. O Relata¹⁷⁴ faz as pesquisas a partir da perspectiva crítica, que oferece uma visão crítica de uma abordagem teórica, metodológica ou outra de uma perspectiva particular; perspectiva de extensão, que aplica formulações teóricas ou conceituais existentes a novos contextos ou casos; perspectiva de incorporação, abordando críticas de trabalhos publicados anteriormente e/ou integrando-os em uma nova posição; perspectiva de reanálise, analisando o material empírico inicialmente reunido e apresentado por outros a partir de um novo olhar; perspectiva de refinamento, modificando, esclarecendo ou delimitando o escopo de formulações existentes à luz de novos contextos ou casos. O buscador apresenta resultados baseados em relações explicitamente identificadas pelos usuários, que, por sua vez, podem adicionar obras e perspectivas (relações).

Outro modelo experimental criado por Ochigame (2021a) é o *Atlas*¹⁷⁵, sistema de busca que questiona fronteiras da informação não muito claras nos critérios de indexação e ranqueamento das plataformas de busca comerciais. O *Atlas* escancara como os resultados de pesquisa do *Google* diferem entre locais e idiomas, questionando, assim, os filtros culturais-hegemônicos e as parcialidades. Por exemplo, a busca pela palavra “*god*” pode retornar com padrões parecidos em países com base religiosa cristã, mesmo que eles estejam geograficamente distantes. Ochigame (2021a) sugere uma subversão de reagrupar países fora do padrão ortodoxo de mapa *mundi*, tomando como ponto de partida os padrões de resposta para “*god*” (criando um “mapa-mundi *god*”). Nessa experiência, Austrália e Serra Leoa poderiam estar no mesmo continente, assim como Argentina e Espanha. A experimentação demonstra que países de maioria cristã tendem a mostrar imagens de um Deus cristão, os países de maioria muçulmana tendem a mostrar imagens de Alá (em árabe escrito) e os países de maioria budista tendem a mostrar imagens de Buda. Isso ocorre porque o *Google* pressupõe o idioma padrão – a língua colonial – a partir do local de onde a busca é feita, mas, afinal, se alguém estiver na Índia fazendo uma busca, mas a língua daquela pessoa for sânscrito, há uma fronteira linguística não reconhecida pelo motor de busca.

A subversão está no fato de que buscadores alternativos poderiam privilegiar o pluralismo, alterando a estrutura homogeneizante da plataformização do texto protagonizada pelo *Google*. Ochigame (2021a) defende que é preciso fazer um esforço curatorial para

¹⁷⁴ Disponível em: <https://relata.mit.edu/>. Acesso em: 10 jul. 2023.

¹⁷⁵ Disponível em: <https://searchatlas.org/>. Acesso em: 10 jul. 2023.

mapear gradativamente outros parâmetros possíveis para busca e produzir metadados legíveis por máquina sobre a semântica de conteúdos. Sabemos que o *Google* tem capacidade semântica, todavia, a entrega da busca ainda não usa de forma crítica essas relações semânticas, ficando mais focado em popularidade e semelhança para movimentar a máquina da publicidade direcionada. Segundo o pesquisador, precisamos de projetos de busca críticos que reconheçam qualquer quantificação de “relevância” como um ato interpretativo, normativo e politicamente consequente, afastando-se da aura mitológica criada pelo *Google*. “Vamos experimentar algoritmos, interfaces e táticas para reindexar o mundo novamente” (OCHIGAME, 2021a, p. 132). Uma busca crítica, afinal, esforçar-se-ia ativamente para aumentar a visibilidade dos movimentos contra-hegemônicos, tradições intelectuais e de perspectivas historicamente marginalizadas.

Nenhum dos exemplos que citamos aqui sai do escopo sugerido por Introna e Nissenbaum (2006), que tiveram a lucidez de sugerir, há tantos anos, três pilares para frear o poderio das plataformas de busca e equilibrar os sistemas de seleção de informação em ambiente digital: 1) divulgação completa e verdadeira das regras subjacentes (ou algoritmos) que regem indexação, busca e classificação; 2) apoio público para o desenvolvimento de mecanismos de busca mais igualitários e inclusivos com tecnologias de metabusca que aumentariam a transparência e o acesso; 3) construção de sistemas sistemas que refletem melhor valores sociais importantes. O que talvez esteja fora deste escopo e nas mãos de produtores de texto tem relação com apresentar um contrafluxo aos modelos formulaicos de escrita agenciados pelas plataformas de busca: subverter também é retomar as rédeas da produção, impedir o enredamento da técnica aos processos que nós deveríamos dominar, em especial, a escrita.

Para finalizar esta seção, pegamos emprestados os questionamentos de (GROHMANN, 2020, p.11): “Que modos outros há de fazer circular a vida em sociedade? Como modificar chaves de sentido em um domínio quase completo da racionalidade neoliberal?”. Apontamos aqui alguns caminhos que combinam regulação de plataformas *online*, por meio de políticas públicas, mudanças na forma como usamos a internet e inovação de *design* de sistemas. Constatamos que é quase impossível pensar em solução que não inclua alguma camada de seleção algorítmica, mas demonstramos que há oportunidades para contestação e transgressão, porque, se deixarmos a tarefa de ordenar a *Web* apenas nas mãos de interesses comerciais, ampliar-se-ão as assimetrias de poder existentes.

Carta ao leitor,

Escrever esta tese foi um processo terapêutico. Estou me curando da escrita performática que, durante anos, o trabalho no jornalismo hard news me exigiu. Poder escrever, analisar, pesquisar sobre as reconfigurações de práticas de escrita é um modo de me limpar, deixando alguma contribuição para a reflexão sobre a plataformização do texto. Como jornalista, lembro-me perfeitamente de embarcar por inteiro na onda da visibilidade, em busca de page views, compartilhamentos e espraiaamentos. E como eu embarquei! Escrevi tantos textos para alcançar os números que, ao me desligar desse tipo de processo produtivo, descobri que estava doente. Eu estava doente pela audiência das notícias.

Vale dizer que nunca foi um trabalho acrítico, pelo contrário, questionava-me rotineiramente sobre o que significava escrever para performar, vencer na busca orgânica, alcançar o topo, ser a notícia mais lida da hora, do dia, do mês. Esse ranqueamento mexia comigo na instância do ego profissional perante colegas e superiores, quando meus textos reinavam no retorno de audiência de forma astronômica – resultados que não rendiam nada para mim individualmente.

Também havia a instância da angústia pessoal de imaginar o que eu poderia fazer para sair daquela redoma da produtividade capitalista. Não restava outra escolha senão me curar disso. E não quer dizer que todos os produtores de texto que se encaixam no espectro dos performáticos estão adoentados, mas eu estava. Foi uma sensação muito libertadora a minha primeira escrita longe das sombras da audiência. Era pensar na hierarquia da informação, as palavras, a coerência, a costura do texto, a consideração pelas fontes e o afeto pelo meu leitor. E isso tudo já era muita coisa!

Longe de mim registrar essa confissão como uma “carta da oprimida” de “jornalista-escrava-de-robôs” (para citar aqui uma expressão da Natália Viana, diretora-executiva da Agência Pública). Pelo contrário, é o fechamento de um ciclo de cura e a esperança de que minha tese sirva, sem grandes utopias, para documentar as práticas da performance e dialogar com todos os profissionais do texto envolvidos nessas dinâmicas. Meu trabalho é nosso trabalho!

Parei para pensar sobre minha história com o Google e lembrei que ele esteve presente como objeto ou parte do processo de pesquisa no meu TCC da graduação, na dissertação e, agora, na tese. Lá se vão 14 anos intrigada com alguns funcionamentos da Google/Alphabet Inc. Nesse tempo, tive a oportunidade de discutir e praticar o texto para ambientes digitais com dezenas e centenas de alunos. Muitos deles estiveram angustiados pela opressão de escrever nos padrões de otimização.

A angústia está na falta de formação sobre o tema, na banalização das técnicas que engendram o texto e suas criatividades, na desmotivação em escrever para algoritmos indexadores em vez de escrever para gente, na vontade de subverter sem conseguir enxergar um caminho, porque o mercado de trabalho e as lógicas imperialistas das plataformas não deixam respiro.

Acolho e empodero esses alunos na medida do possível. Convido a pensar em disrupção, mas, confesso, que há certa impotência no meu fazer, porque tantos outros estudantes estão em busca de aprender a fazer o fluxo das plataformas e de ter a chance de sobreviver trabalhando. É uma contradição imensa! Para esses alunos e colegas que não querem romper, vou oferecer iniciações, técnicas, ferramentas, repertórios. Para os que estão no ponto da exaustão em busca da subversão, ofereço diálogo. Mais uma vez, meu trabalho é nosso trabalho!

Luana Cruz

6 CONSIDERAÇÕES FINAIS

Neste trabalho, verificamos a reconfiguração de práticas de escrita e edição – que adquirem novos pontos de atenção a exemplo de títulos, palavras-chave e cabeçalhos escaneáveis – a partir da inserção da lógica das plataformas de busca na “produção, circulação e leitura” de texto em ambientes digitais. Concluimos que a incorporação de técnicas em resposta às demandas algorítmicas de ranqueamento do *Google* é comum e praticada de modo ambivalente por produtores e editores de texto, que buscam ranqueamento e performance. A ambivalência está no fato de que esses produtores se sentem, ao mesmo tempo, na condição de ter que equilibrar seu trabalho entre a relevância para públicos e relevância para algorítmicos, mas muitos deles criam táticas de produção que parecem vencer a caixa-preta do *Google* em nome da performance e visibilidade de conteúdos.

Durante o percurso, caracterizamos sistematicamente o *Google* como plataforma de busca, o que permitiu a compreensão de movimentos e associações que nos deixam enxergar plataformização do texto, especialmente, ao perceber a indissociabilidade das estratégias comerciais e datificação com a seleção/classificação dos textos eleitos pela plataforma. A compreensão dos pilares que sustentam essa plataforma – programabilidade de algoritmos de rastreamento, indexação e classificação; conectividade de leitores que buscam e clicam capitaneados pelos critérios da similaridade e popularidade; coleta de dados que fazem girar a lógica da propaganda direcionada – pode parecer óbvia para alguns. No entanto, não havia registros organizados do funcionamento político e técnico do *Google* em língua portuguesa, nem mesmo a conexão, sob o olhar das Linguagens, dessa tecnologia como responsável pela plataformização do texto. O momento político brasileiro, de discussão da regulação de plataformas, é terreno fértil para apresentar essa pesquisa que documenta a caracterização infraestrutural do *Google*. Ainda mais porque a *Google/Alphabet Inc.* se posicionou como uma das principais lobistas na contra-argumentação da regulação no Brasil. Concluimos, portanto, pela necessidade de regulação do *Google*, a partir de um olhar público que reflita melhor os valores sociais na seleção. Não podemos deixar o mercado decidir sobre texto e informação, bens cuja circulação precisaria exigir limitações éticas das plataformas *online*.

A análise rigorosa de terceiros sobre a engrenagem da plataforma de busca é necessária para garantir a proteção da liberdade de expressão, porque o *Google* tem ampla confiança do consumidor e pode estar menos inclinado a se envolver em mudanças que arriscam mal-entendidos públicos. Mudar para quê? Se muita gente já confia no *Google* como

é, melhor deixar como está. Há ironia aqui, claro. Afinal, no percurso da pesquisa, argumentamos que a percepção dessa tecnologia como oráculo tem grande risco às pluridiversidades textuais, ao direito à informação, à construção do conhecimento coletivo e à sobrevivência do mercado multilateral de produção de texto.

Os testes que fizemos, no Capítulo 5, mostraram inúmeras possibilidades de adequação de textos à lógica de ranqueamento, desde a conformação de títulos com termos-chave até a inserção de *links* ou de palavras de transição entre sentenças. As simulações deixaram claro que o *Google* mudou o jeito de escrever daqueles que pretendem alcançar visitividade em ambiente digital, porque consolidou como quase inevitável seu modelo de indexação e classificação que tensiona produtores de texto a redigir pensando na relevância mitológica algorítmica. O *Google* pode ter mudado também nosso jeito de ler, pois fomenta a informação topográfica e ranqueada. As dinâmicas que se circunscrevem no *ranking* do *Google* convocam leitores e criadores de conteúdo ao pensamento hierárquico. O mais grave é que não está explícito o modelo de decisão para ranqueamento, pois se baseiam no *feedback* do leitor para gerar indutivamente tal modelo, o que cria um problema fundamental de interpretabilidade. Aliás, insistimos muito nessa perspectiva de não objetividade do *Google* quando listamos, no Capítulo 4, as diretrizes documentadas pela plataforma.

Demonstramos também em nossa trajetória de pesquisa pontos paradoxais. Um deles é que o *Google* está investindo cada vez mais em inteligência semântica e uso de linguagem natural, como no caso do *Core Update BERT*, mas o produtor de texto parece compreender a escrita para algoritmos como algo robotizado, porque algoritmos leem HTMLs. O caminho não seria escrever para humanos e tornar as práticas editoriais parametrizadas dispensáveis? Concluimos que, enquanto for necessário pensar a produção textual relevantemente para o leitor e ainda organizada na configuração do HTML, haverá uma sensação de atravessamento de robôs. Mesmo que seja uma escrita criativa e única, será necessário algum trabalho de edição (mesmo que transparente ao leitor e diretamente feita no HTML) para a elegibilidade algorítmica. Ressaltamos que a edição de textos é que dá o tom da parametrização para performance. Portanto, o protagonismo da edição se firma principalmente no “[...] sentido mais geral de tratamento, interferência, e/ou manipulação das materialidades discursivas com o objetivo de preparar um produto simbólico para sua circulação social”. (MUNIZ JR., 2018, p. 74).

Outro ponto paradoxal é o fato de que os *Core Updates* são publicizados pelo *Google* como melhoramentos, mas de certo modo se configuram como respostas aos usos táticos e estratégicos que acabam sendo descobertos por especialistas, principalmente, os de

otimização. A governança é, de algum modo, responsiva a esses usos, por isso ela muda tanto. Isso ajuda a sustentar o caráter contingencial, ou seja, no dinamismo que o *Google* consegue imprimir nos resultados de seleção e classificação a partir do processo dinâmico de verificação por aprovação coletiva. A governança sustenta também o caráter ambivalente da tecnologia, extremamente útil pelo serviço de busca e devoradora de fatias de publicidade direcionada.

Um de nossos objetivos de pesquisa foi discutir as consequências da plataformização do texto. Percebemos que um dos grandes custos do viés de mecanismos de busca mediando a circulação de texto é que esses seletores restringirem sistematicamente o escopo do que os leitores podem receber e quais *sites* podem ser vistos, assim, perderemos o valor geral da *Web* como um espaço público, plural e como uma fonte de informação amplamente inclusiva. Se o critério central de seleção e classificação é a popularidade, o produtor de texto trabalha em busca desse marcador de desempenho. Portanto, a natureza contingente do texto plataformizado surge a partir das parametrizações que visam ao ranqueamento, como demonstramos no Capítulo 5 e retomamos aqui. Com a ajuda de *plugins* – atuando como bússolas editoriais –, produtores e editores de texto escolhem estrategicamente a palavra-chave; planejam textos na lógica da engenharia reversa; contabilizam caracteres de títulos, metadescrições e URLs para serem visíveis na SERP; inserem *links* internos e externos e fazem *lobby* por *links*; hierarquizam o texto com cabeçalhos; entre outras ações. Cada uma dessas práticas entrelaça texto e tecnologia à medida que são convertidas em parametrizações de código HTML para otimizar um texto, sinal de que os processos de escrita e edição à luz de lógicas algorítmicas são emergentes e recorrentes nos modos de construir texto na atualidade.

Apresentamos como implicações da reconfiguração da escrita e da edição o uso generalizado de técnicas de SEO, as quais relativizamos como ações táticas, estratégicas ou domesticadas, a depender do ponto de vista. Concluimos que as técnicas de SEO são também condutoras da plataformização das cadeias editoriais. Isso porque precisamos inserir agentes técnicos, como *plugins*, e explorar *affordances* específicas do CMS e das plataformas de busca que permitam o ranqueamento de textos. Concluimos que os produtores aplicam técnicas de otimização de buscas e, assim, apropriam-se de modo potencialmente vicioso do processo de plataformização.

Outra implicação importante é a homogeneização dos textos decorrente dos esquemas estéticos impostos pelo ranqueamento *Google*, o que nos remete a uma cultura escrita em movimento cercada de mudanças nos modos de inscrição do texto, na difusão da

escrita, nas estruturas fundamentais dos suportes e na percepção das formas textuais (RIBEIRO, 2018).

Além dessa retomada que fizemos para responder à pergunta de pesquisa e apresentar conclusões, optamos por reunir, a seguir, o que consideramos contribuições relevantes, sob pena de o leitor atento suspeitar de repetição. No entanto, vimos que seria importante sistematizar esses achados em sinal de conclusão e colaboração com o campo dos Estudos de Linguagem.

6.1 TEXTO PLATAFORMIZADO

Plataformas *online* podem ser “espaços de escrita” (BOLTER, 1991) pensadas como campo físico e visual em que “acontece” o texto, afinal todas as formas de escrita são espaciais e exigem um lugar em que se inscreva/escreva. No entanto, podem ser também “equipamentos de leitura” (FURTADO, 2003), compreendendo leitura como um “entrelaçamento entre as competências de ler e navegar” (COSCARELLI, 2016). Ainda mais porque leitura e navegação são partes da dinâmica das plataformas *online*, já que o tempo gasto pelo leitor no ambiente plataformizado permite coleta dos dados de uso, comportamentos de consumo e outros elementos que contribuem para sustentar a datificação.

Acontece que o texto que se produz na *Web* plataformizada não é apenas fragmentado, não linear, hiperconectado, ele é feito, quase sempre, para atender a requisitos de desempenho mensuráveis. Ademais às questões da hipertextualidade, percebemos a crescente consciência de produtores de texto em ambiente digital sobre mecanismos de desempenho na circulação. Esses profissionais incluem como parte do pensar texto a seguinte pergunta: o que o meu leitor buscaria no *Google* para encontrar o conteúdo sobre o qual estou escrevendo? Com isso, acabam por se tornar profissionais automatizados e competitivos, dotados, em certa medida, de uma consciência da relação da produção com a performance algorítmica. Quando falamos em circulação com intuito de desempenho, é porque esses produtores e editores já perceberam que o modo como os leitores chegam a um texto ou produto cultural é tão importante quanto à informação/arte inscrita neles.

Por tudo isso, propusemos a noção de “texto plataformizado”, concebendo-o como imanente e contingente, que se constrói de modo hipertextual pela natureza do meio e do processo de leitura, mas em atenção ao desempenho para a visibilidade. E quando nos referimos à visibilidade, destacamos que a conquista dela não é um caminho desprezioso

ou intuitivo, mas uma modulação de materialidades das plataformas e dos textos para performar.

6.2 CULTURA DIGITAL PÓS-PLATAFORMAS

A atual reconfiguração de práticas na produção de texto é muito radical porque nos leva a um cenário de simbiose quase irreversível com as plataformas online e seus algoritmos. Podemos pensar em outra cultura digital pós-plataformas, pois aquela cultura digital baseada na emergência do ciberespaço e dos hipertextos, não é suficiente para olhar criticamente para o que está acontecendo. É preciso considerar a interferência dos meios e suas materialidades.

Quando compreendemos as interações digitais de forma mais materializada, percebemos que nem todas as pessoas vão se apropriar da mesma forma das funcionalidades disponíveis nesses ambientes. Como argumenta d'Andréa (2020), são muitos os usos táticos, lúdicos ou políticos possíveis nas diferentes plataformas. Essas ações são, em grande medida, contingentes, mas que parecem se repetir em alguns casos, a exemplo das técnicas de SEO, práticas incorporadas que demonstram as dimensões de atravessamentos dos algoritmos na cultura escrita. Esses algoritmos, por sua vez, são carregados de vieses e automatizações equivocadas. Manter aparência de tecnicidade ajuda os provedores a fugir de acusações de manipulação de resultados. Imperdoável é ignorar que os algoritmos são programados por seres humanos, com suas escolhas subjetivas ou guiadas por instituições. Tudo isso reforça que a tecnologia não é isenta, imparcial ou objetiva, pelo contrário, está coberta de ideologias e expressões de poder, além da natureza ambivalente.

Por todo esse contexto, propusemos a definição de uma “cultura digital pós-plataformas”, pensando ser esta mais atravessada pela comunicação mediada por algoritmos e pela tradução de tudo em dados numéricos, conforme alertou Ribeiro (2018) citando Felinto (2006). Segundo a autora, ao lermos autores entusiastas da tecnologia e ciberespaço, talvez tenhamos ficado com a impressão de que a cultura digital era tudo que sempre sonhávamos: “[...] um mundo de compartilhamento, cooperação, diálogo, leituras móveis, andanças, liberdade e troca” (RIBEIRO, 2018, p. 15). No entanto, percebemos que essa cultura digital pós-plataformas tem como fator central embeber nossas vivências sociais com uma experiência tecnológica financeirizada pelas *Big Five*. Na cultura digital pós-plataformas, a dinâmica que se estabelece não é nova, é somente outra. Afinal, a financeirização também

permeava outros momentos da história da escrita e da leitura. Há, portanto, uma tensão constante para troca desse lugar de poder temporário.

6.3 MITO DA RELEVÂNCIA

Há uma percepção senso comum de que algoritmos de plataformas de busca, como *Google*, estão selecionando e classificando conteúdos e *sites* mais relevantes na SERP. Contestamos essa percepção, mostrando por meio da análise documental do *Google* que a plataforma está, na verdade, escolhendo definir relevância ao criar métodos de contar e agrupar citações, além de atribuir valores, como qualidade, autoridade e reputação. Todavia, a escolha não é objetiva, ela é normativa e baseada em valores culturais e políticos. É que não existe uma métrica independente capaz de dizer quais seriam os resultados de pesquisa mais relevantes para qualquer busca, afinal, “[...] relevante é um julgamento fluido e carregado de significados, tão aberto à interpretação quanto alguns dos termos avaliativos sobre os quais estudiosos da mídia já se debruçaram, como ‘notável’ ou ‘popular’” (GILLESPIE, 2018, p. 103).

Por isso, concluímos que a narrativa simbólica documentada pelo *Google* – em seus termos de governança – criam o “mito da relevância”, que envolve eloquentemente produtores de texto e desenvolvedores *Web*. Propusemos a concepção de que o mito da relevância trata de acomodar os profissionais do texto ao mundo da escrita em ambiente digital, configurando-se como uma estratégia aos negócios do *Google*. Todas as normativas que parecem apoiar a prática de produção de textos não têm a obrigação de explicar. A função do mito da relevância não é tornar inteligível, mas, sim, albergar profissionais do texto em uma realidade técnica e algorítmica (encaixapretada), que parece muito assustadora e que é, na verdade, dotada de subjetividades escancaradas. O mito da relevância tem vida própria porque interfere em práticas de escrita das mais diversas e os profissionais do texto nem precisam, necessariamente, conhecer os documentos do *Google* que normatizam a relevância. O mito da relevância dá respostas, mesmo que incorpóreas e tendenciosas, a esses profissionais, portanto, desempenha uma função social, fixando ritos e modelando a produção textual.

6.4 PLATAFORMIZAÇÃO DO TEXTO

Há muitos profissionais do texto submetendo suas produções a camadas técnicas – parametrizações, aconselhamentos automatizados de *plugins* e sistemas algorítmicos – com objetivo de performar em plataformas de buscas, uma dinâmica que ilustra o fenômeno da plataformação do texto. Observamos que a dependência de uma plataforma *online*, como *Google*, para geração de tráfego para *sites* e visibilidade para conteúdos acaba por moldar a “produção, circulação e leitura” a partir dos critérios privados e empresariais dessa plataforma. Sendo assim, propusemos a concepção da “plataformização do texto” como fenômeno composto pelo atravessamento de critérios algorítmicos e normatizações definidas pelas plataformas de busca na escrita e edição em ambiente digital. A plataformação do texto é o que faz existir o texto incrementado em performance e desempenho e amarrado a retórica do mito da relevância construída pelo *Google*.

Se pensarmos em um pequeno espectro de intencionalidades de escrita como o jornalista que escreve para informar, o redator SEO que escreve para atrair, o *copywriter* que escrever para converter, o professor que escreve para ensinar, o escritor que escreve para fabular (VECCHIO, 2022), o divulgador científico que escreve para popularizar, o *creator (influencer)* que escreve para encantar, o publicitário que escrever para persuadir, quase não haverá fuga das dinâmicas de plataformas de busca se o ambiente de produção for a *Web*. A plataformação do texto penetra e afeta todos esses setores. A complexa interação entre CMS e *Google* é parte do processo definido por Helmond (2015) como plataformação da *Web*. Quando responsáveis por *websites* modulam conteúdos e funcionalidades, refinando apresentação visual e construção em APIs, ele estão, muitas vezes, providenciando a plataformação do texto. Portanto, a visibilidade das produções, de cada profissional citada no espectro anterior, dependerá da seleção das plataformas de busca a partir dos usos que esses criadores fazem, principalmente das práticas que eles incorporam em busca do desempenho. Isso ocorre porque as lógicas de funcionamento, governança, construção de “métricas” e conjunto sociotécnico de algoritmos são responsáveis por novos modos de escrever, editar e ler textos em ambiente digital.

Sabemos que não esgotamos os estudos sobre o *Google* como plataforma, nem mesmo essa visão de mediação da “produção, circulação e leitura”. Sobre esse sistema

complexo, nossa pesquisa tem limitações ao tratar de leitura, porque os métodos escolhidos olham exclusivamente para a produção. Conseguimos observar os movimentos de busca como movimentos de leitura. A intenção de busca, leitura-escritura, leitura-navegação e interação pós-SERP são importantes às engrenagens do *Google*, mas ainda há um desafio metodológico em compreender a leitura mediada por plataformas *online*. As possibilidades de pesquisa poderiam incluir estudos de recepção.

A nossa compreensão da caixa-preta do *Google* é também limitada, primeiramente pelos nossos conhecimentos técnicos e formação sem enfoque no desenvolvimento de algoritmos e, em segunda instância, pela opacidade da plataforma. Isso não se configura exatamente como um problema, porque revela o quanto os sistemas das *Big Five* são intencionalmente encaixapretados, mas vamos considerar como uma limitação.

Sugerimos estudos futuros sobre a plataformização do texto por áreas (jornalismo, *marketing*, ciência, literatura etc.), a fim de reunir particularidades; trabalhos que ouçam produtores e editores envolvidos com o texto plataformizado com intenção de coletar percepções (ou não percepções) sobre as implicações do *Google* na rotina de trabalho; pesquisas que sistematizem modos de regulação dos mecanismos de pesquisa muito além da divulgação completa e verdadeira das regras subjacentes (ou algoritmos). Sugerimos, também, uma pesquisa a partir do olhar de avaliadores de pesquisa, entendendo esse processo de calibragem do *Google* e, claro, problematizando ainda mais precarização desses trabalhadores. E, claro, sugerimos estudos que aprofundem as relações entre IA e plataformas de busca.

Por fim, ainda há a possibilidade de projetos inspirados na remodelação de sistemas de busca – como o fez Ochigame (2021a) –, mas pensados conjuntamente com modelos de negócios implementáveis; estudos que enfoquem metodologias de ensino e letramento crítico para futuros produtores de texto, a fim de empoderá-los em busca das subversões; trabalhos que olhem mais profundamente para a edição de texto concomitante com a edição HTML (desenvolvimento *Web*), colocando esses dois trabalhos como complementares. Espero que este trabalho contribua para reflexões em espaços acadêmicos e profissionais, lugares de onde venho e para onde sempre **PARTO**.

REFERÊNCIAS

- ANDERSON, Chris. **The long tail**: why the future of business is selling less of more. San Francisco: Wired, 2004.
- ANDERSON, Clarence W. Beyond journalism in the present tense. **NiemanLab**, 2014. Disponível em: <http://www.niemanlab.org/2014/12/beyond-journalism-in-the-present-tense/>. Acesso em: 14 jun. 2023.
- ASSIS, Juliana Alves. Gêneros textuais, tecnologia e textualização. **Scripta**, Belo Horizonte, v. 6, n. 11, p. 134-151, 2002.
- BAKHTIN, Mikhail. Os gêneros do discurso. *In*: BAKHTIN, Mikhail. **Estética da criação verbal**. vol. 4. São Paulo: WMF Martins Fontes, 1992. p. 261-306.
- BENJAMIN, Walter. Teoria do conhecimento, teoria do progresso. *In*: BENJAMIN, Walter. **Passagens**. Belo Horizonte: Editora UFMG; São Paulo: Imprensa Oficial do Estado de São Paulo, 2006. p. 499-530.
- BENJAMIN, Walter. **A tarefa do tradutor**. Tradução de João Barrento. Lisboa: Universidade Nova de Lisboa, 2008.
- BOGOST, Ian; MONTFORT, Nick. **Platform studies**: frequently questioned answers. Irvine: Digital Arts and Culture, 2009.
- BOLTER, Jay David. **Writing space**: the computer, hypertext, and the history of writing. Mahwah: Erlbaum, 1991.
- BOLTER, Jay David; GRUSIN, Richard A. Remediation. **Configurations**, v. 4, n. 3, p. 311-358, 1996.
- BRUNO, Fernanda. Algorítmicas e Negacionismo: a pandemia, o comum, o futuro. *In*: DUARTE, Luisa; GORGULHO, Victor (org.). **No tremor do mundo**: ensaios e entrevistas à luz da pandemia. Rio de Janeiro: Cobogó, 2020. p. 241-254.
- BRIN, Sergey; PAGE, Lawrence. The anatomy of a large-scale hypertextual web search engine. **Computer networks and ISDN systems**, v. 30, n. 1-7, p. 107-117, 1998.
- BUCHER, Taina. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. **Information Communication and Society**, v. 20, n. 1, p. 30-44, 2017.
- BUENO, Thaisa; REINO, Lucas Santiago Arraes. SEO no jornalismo: títulos testáveis e suas implicações. **Estudos em Jornalismo e Mídia**, v. 16, n. 2, p. 98-112, 2019.
- JUANALS, Brigitte. O livro e a tela – acesso ao conhecimento: da página do livro à página-tela. Entrevista concedida a Geneviève Vidal. *In*: RIBEIRO, Ana Elisa et al. **O hipertexto em tradução**. Belo Horizonte: UFMG, 2007. p. 71-81.

CAMPOS, Haroldo de. Da transcrição: poética e semiótica da operação tradutora. *In*: OLIVEIRA, Ana Cláudia de; SANTAELLA, Lúcia (org.). **Semiótica da Literatura**. São Paulo: Educ, 1987. p. 53-74.

CARVALHO, Juliano Maurício de. O dono do texto: dilemas do jornalismo em plataformas digitais. *In*: MARTINS, Gerson Luiz; REINO, Lucas Santiago Arraes; BUENO, Thaísa (org.). **Performance em ciberjornalismo**. Tecnologia, inovação e eficiência. Campo Grande: Editora UFMS, 2017. p. 79-94.

CASTRO, Julio Cesar Lemes de. Plataformas algorítmicas: interpelação, perfilamento e performatividade. **Revista Famecos**, Porto Alegre, v. 26, n. 3, p. 1-24, 2019.

CERTEAU, Michel de. **A invenção do cotidiano**: 1. Artes de fazer. 16 ed. Petrópolis: Vozes, 2009.

CHARAUDEAU, Patrick. **Discurso das mídias**. 2. ed. São Paulo: Contexto, 2010.

CHARTIER, Roger. **A aventura do livro**: do leitor ao navegador. São Paulo: Editora UNESP; Imprensa Oficial do Estado, 1999.

CHARTIER, Roger. **Os desafios da escrita**. São Paulo: Editora UNESP, 2002. p. 77-100.

CHAUÍ, Marilena. **Boas-Vindas à filosofia**. São Paulo: WMF Martins Fontes, 2013.

COSCARELLI, Carla Viana. Entendendo a leitura. **Revista de estudos da linguagem**, Belo Horizonte, v. 10, n. 1, p. 7-27, 2002.

COSCARELLI, Carla Viana. Textos e hipertextos: procurando o equilíbrio. **Linguagem em (Dis)curso**, Tubarão, v. 9, p. 549-564, 2009.

COSCARELLI, Carla Viana. Navegar e ler na rota do aprender. *In*: COSCARELLI, Carla Viana. **Tecnologias para aprender**. São Paulo, Parábola Editorial, 2016. p. 61-80.

COSCARELLI, Carla Viana. A leitura em múltiplas fontes: um processo investigativo. **Ensino e Tecnologia em Revista**, Curitiba, v. 1, n. 1, p. 67-79, 2017.

COSTA VAL, Maria da Graça. Repensando a textualidade. *In*: AZEREDO, José Carlos de. **Língua portuguesa em debate**: conhecimento e ensino. Rio de Janeiro: Vozes, 2000. p. 34-51.

CHRISTOFOLETTI, Rogério; VIEIRA, Livia de Souza. Métricas, ética e “cultura do clique” no jornalismo online brasileiro: o caso de resistência do não fo. de. **Dispositiva**, v. 4, n. 1, p. 74-87, 2015.

COMISSÃO EUROPEIA. The Digital Services Act package, 2022. Disponível em: <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>. Acesso em: 14 set. 2022.

CORRÊA, Elizabeth Saad; BERTOCCHI, Daniela. O algoritmo curador – O papel do comunicador num cenário de curadoria algorítmica de informação. *In: ENCONTRO ANUAL DA COMPÓS*, 21., Juiz de Fora, 2012. **Anais...** Juiz de Fora: Compós, 2012.

CRUZ, Luana Teixeira de Souza. **Rota hipertextual baseada em tags**: discussão de processos de produção e leitura como sistemas complexos no contexto da Web Semântica. 2014. 126 f. Dissertação (Mestrado em Estudos de Linguagens) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2014.

CRUZ, Luana Teixeira de Souza; RIBEIRO, Ana Elisa. Redação Web: novos fluxos editoriais e breve estudo de caso de inbound marketing. *In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO*, 39., São Paulo, 2016. **Anais...** São Paulo: Intercom, 2016. p. 1-16.

CRUZ, Luana Teixeira de Souza; VECCHIO, Pollyanna Mattos Moura. Plataformização da *Indústria Cultural*: produções contingentes no *Amazon Kindle Unlimited*. *In: CONGRESSO INTERNACIONAL DE ARTE, CIÊNCIA E TECNOLOGIA*, 7., virtual, 2022.

CUPANI, Alberto. Tecnologia e poder. *In: CUPANI, Alberto. Filosofia da tecnologia: um convite*. Florianópolis: Editora da UFSC, 2011. p. 151-168.

DALBEN, Sílvia. **Cartografando o jornalismo automatizado**: redes sociotécnicas e incertezas na redação de notícias por “robôs”. 2018. 117 f. Dissertação (Mestrado em Comunicação Social) – Universidade Federal de Minas Gerais, Belo Horizonte, 2018.

DARDOT, Pierre; LAVAL, Christian. **A nova razão do mundo**. São Paulo: Boitempo, 2016.

DARTON, Robert. **The case for books**: past, present, and future. Nova York: PublicAffairs, 2009.

D’ANDRÉA, Carlos Frederico de Brito. **Pesquisando plataformas online**: conceitos e métodos. Salvador: EDUFBA, 2020.

DIGITAL MARKETS ACT (DMA). 2022. Disponível em: <https://eur-lex.europa.eu/eli/reg/2022/1925/oj>. Acesso em: 14 set. 2022.

FURTADO, José Afonso. O papel e o pixel. **Ciberdifusão**, Coimbra, jun. 2003.

FOLHA DE S.PAULO. Webinar Web 3.0 e Metaverso. 2022. Disponível em: <https://www1.folha.uol.com.br/seminariosfolha/2022/06/web-30-e-metaverso-sao-temas-de-seminario-folha.shtml>. Acesso em 13 jul. 2022.

FOLHA DE S.PAULO. **Manual da redação**: as normas de escrita e conduta do principal jornal do país. São Paulo: Publifolha, 2022.

FONSECA, Gregório de Almeida; D’ANDRÉA, Carlos Frederico de Brito. Governança e mediações algorítmicas da plataforma YouTube durante a pandemia de COVID-19. **Revista Dispositiva**, Belo Horizonte, v. 9, p. 6-26, 2020.

FONSECA, Gregório de Almeida. **A “vacina” nas plataformas:** (des) informação científica no YouTube antes e durante a pandemia de Covid-19 no Brasil. 2022. 229f. Tese (Doutorado em Comunicação Social) – Universidade Federal De Minas Gerais, Belo Horizonte, 2022.

GABRIEL, Martha. **Marketing na era digital:** conceitos, plataformas e estratégias. São Paulo: Novatec Editora, 2010.

GARCÍA-OROSA, Berta; CANAVILHAS, João; VÁZQUEZ-HERRERO, Jorge. Algorithms and communication: a systematized literature review. **Comunicar Media Education Research Journal**, v. 31, n. 74, p. 09-21, 2022.

GILLESPIE, Tarleton. Algorithm. *In*: PETERS, Benjamin. **Digital keywords:** a vocabulary of information society and culture. Princeton: Princeton University Press, 2016. p. 18-30.

GILLESPIE, Tarleton. A relevância dos algoritmos. **Parágrafo**, São Paulo, v. 6, n. 1, p. 95-121, 2018.

GONZAGA-PONTES, Camila. **Aguarde mais informações.** Uma análise da webnotícia com base na releitura da estrutura da notícia de Teun van Dijk. 2012. 128 f. Dissertação (Mestrado em Estudos de Linguagens) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2012.

GOOGLE Google Search sends more traffic to the open web every year. 2021. Disponível em: <https://blog.google/products/search/google-search-sends-more-traffic-open-web-every-year/>. Acesso em: 12 jul. 2023.

GOOGLE. Como resultados são gerados automaticamente. 2022a. Disponível em: <https://www.google.com/intl/>. Acesso em: 08 abr. 2022.

GOOGLE. Search quality evaluator guidelines. 2022b. Disponível em: <https://guidelines.raterhub.com/searchqualityevaluatorguidelines.pdf>. Acesso em: 16 set. 2022.

GOOGLE. Guia do Google de introdução à otimização para motores de busca. 2022c. Disponível em: <https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=pt-br>. Acesso em: 15 dez. 2022.

GOOGLE. Como a busca do Google funciona. 2022d. Disponível em: <https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/>. Acesso em: 15 dez. 2022.

GOOGLE. Google news webinar for LATAM publishers. 2022e. Disponível em: <https://publishersonair.withgoogle.com/events/latam-google-news-Webinar-danny-sullivan>. Acesso em: 08 abr. 2022.

GOOGLE. Relatório de transparência. 2023a. Disponível em: <https://transparencyreport.google.com/about?hl=pt-br>. Acesso em: 09 jan. 2023.

GOOGLE. Guia detalhado sobre como a Pesquisa Google funciona. 2023b. Disponível em: <https://developers.google.com/search/docs/fundamentals/how-search-works?hl=pt-br>. Acesso em: 09 jan. 2023.

GOOGLE. Melhorando a Busca com testes rigorosos. 2023c. Disponível em: <https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/rigorous-testing/>. Acesso em: 13 jan. 2023.

GOOGLE. Como resultados são gerados automaticamente. 2023d. Disponível em: <https://www.google.com/intl/pt-BR/search/howsearchworks/how-search-works/ranking-results/#meaning>. Acesso em: 13 jan. 2023.

GOOGLE. Avaliação da experiência na página para melhorar o conteúdo da Web. 2023e. Disponível em: <https://developers.google.com/search/blog/2020/05/evaluating-page-experience?hl=pt-br#page-experience-ranking>, Acesso em: 24 fev. 2023.

GOOGLE. Termo de serviços e política de privacidade. 2023f. Disponível em: <https://policies.google.com/terms?hl=pt-BR>. Acesso em: 08 mar. 2023.

GOOGLE. Patentes. 2023g. Disponível em: <https://patents.google.com/patent/US6285999B1/en>. Acesso em: 08 mar. 2023.

GOOGLE. Synonym identification based on co-occurring terms. 2023h. Disponível em: <https://patents.google.com/patent/US8538984B1/en?q=US+Patent+8%2c538%2c984>. Acesso em: 08 jun. 2023.

GOOGLE. Como o Google apoia o jornalismo e ajuda os brasileiros a acessar informações relevantes. 2023i. Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/apoiando-o-jornalismo/>. Acesso em: 06 jun. 2023.

GOOGLE. Learn from others experiences with more perspectives on Search. 2023j. Disponível em: <https://blog.google/products/search/google-search-perspectives/>. Acesso em: 12 jul. 2023.

GOOGLE. Method for node ranking in a linked database. 2023k. Disponível em: <https://patents.google.com/patent/US6285999B1/en>. Acesso em: 12 jul. 2023.

GOOGLE. Producing a ranking for pages using distances in a web-link graph. 2023l. Disponível em <https://patents.google.com/patent/US9165040B1/en>. Acesso em: 12 jul. 2023.

GOOGLE. Google Search Status Dashboard. 2023m. Disponível em: <https://status.search.google.com/products/rGHU1u87FJnkP6W2GwMi/history?hl=pt-br>. Acesso em: 12 jul. 2023.

GOOGLE algorithm update history: a history of major Google algorithm updates from 2000-present. **Moz**, s.d. Disponível em: <https://moz.com/google-algorithm-change>. Acesso em: 08 jun. 2023.

GROHMANN, Rafael do Nascimento. A comunicação na circulação do capital em contexto de plataformação. **Liinc em Revista**, Belo Horizonte, v. 16, n. 1, p. 5145, 2020a.

GROHMANN, Rafael do Nascimento. Plataformização do trabalho: entre dataficação, financeirização e racionalidade neoliberal. **Revista Eletrônica Internacional de Economia**

Política da Informação, da Comunicação e da Cultura, São Cristóvão, v. 22, n. 1, p. 106-122, 2020b.

GROHMANN, Rafael do Nascimento. O que é circulação na comunicação? Dimensões epistemológicas. **Revista Famecos**, Porto Alegre, v. 27, p. 01-13, 2020c.

HARAWAY, Donna. Situated knowledges: the science question in feminism and the privilege of partial perspective. **Feminist Studies**, Washington, v. 14, n. 3, p. 575-599, 1988.

HELMOND, Anne. The platformization of the Web: making Web data platform ready. **Social Media and Society**, London, v. 1, n. 2, 2015.

HISSA, Cássio Eduardo Viana. **Entrenotas: compreensões de pesquisa**. Belo Horizonte: Editora UFMG, 2012.

INTRONA, Lucas D.; NISSENBAUM, Helen. Shaping the Web: why the politics of search engines matters. **The information society**, S.l., v. 16, n. 3, p. 169-185, 2006.

JENKINS, Henry. **Cultura da convergência**. São Paulo: Aleph, 2008.

JENKINS, Henry; FORD, Sam; GREEN, Joshua. **Cultura da conexão: criando valor e significado por meio da mídia propagável**. São Paulo: Aleph, 2014.

JURNO, Amanda. C. **Facebook e a plataformação do jornalismo: uma cartografia das disputas, parcerias e controversas entre 2014 e 2019**. 2020. 225 f. Tese (Doutorado em Comunicação Social) – Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

JURNO, Amanda. A mediação jornalística x A mediação algorítmica. *In*: BENFIELD, Dalida; MORESCHI, Bruno; PEREIRA, Gabriel; YE, Katherine. **Afetando tecnologias, maquiando inteligências**, 2021. Disponível em: <https://book.affecting-technologies.org/a-mediacao-jornalistica-x-a-mediacao-algoritmica/>. Acesso em: 18 ago. 2022.

KALSING, Janaína. **Jornalistas metrificados e a plataformação do jornalismo**. 240 f. 2021. Tese (Doutorado em Comunicação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.

KARHAWI, Issaaf. TikTok e pânico moral. **COM+**, São Paulo, 2021. Disponível em: <http://www.commais.org/2021/07/tiktok-e-panico-moral/>. Acesso em: 31 mar. 2023.

KITCHIN, Rob. Thinking critically about and researching algorithms. **Information, communication & society**, London, v. 20, n. 1, p. 14-29, 2016.

KRESS, Gunther.; VAN LEEUWEN, Theo. Front Pages. (the critical) analysis of newspaper layout. *In*: BELL, A.; GARRET, P. (org.). **Approaches to media discourse**. Hoboken: Blackwell Publishing, 1998. p. 186-219.

LANDOW, George P. **Hypertext 3.0: Critical theory and new media in an era of globalization**. Baltimore: JHU Press, 2006.

LEAL, Bruno Souza. Do texto à textualidade na comunicação: contornos de uma linha de investigação. *In*: LEAL, Bruno Souza; CARVALHO, Carlos Alberto; ALZAMORA, Geane. **Textualidades midiáticas**. Belo Horizonte: PPGCom/UFMG, 2018. p. 17-34.

LEAVER, Tama; HIGHFIELD, Tim; ABIDIN, Crystal. **Instagram: visual social media cultures**. Hoboken: John Wiley & Sons, 2020.

LIBERATO, Yara; FULGÊNCIO, Lúcia. **É possível facilitar a leitura: um guia para escrever claro**. 2. ed. São Paulo: Contexto, 2018.

LORUSSO, Silvio. **Emprecariado: todo mundo é empreendedor. Ninguém está a salvo**. São Paulo: Clube do Livro do Design, 2023.

MAINGUENEAU, Dominique. Ethos, cenografia, incorporação *In*: SILVEIRA, Germínia M. M. L. Ethos no discurso: trilhando um percurso teórico. **Nova Revista Amazônica**, v. 2, n. 1. 2014, p. 72-83.

MANOVICH, Lev. **The language of new media**. Cambridge; London: MIT Press, 2001.

MARCUSCHI, Luiz Antônio. Linearização, cognição e referência: o desafio do hipertexto. *In*: COLÓQUIO DA ASSOCIAÇÃO LATINO-AMERICANA DE ANALISTAS DO DISCURSO, 4., Santiago, Chile, 1999. **Anais...** Santiago, Chile: Aled, 1999.

MARCUSCHI, Luiz Antônio *et al.* **Gêneros textuais: definição e funcionalidade**. *In*: DIONÍSIO, A.; MACHADO, A. R.; BEZERRA, M. A. (org.). **Gêneros textuais e ensino**. Rio de Janeiro: Lucerna, 2002. p. 19-36.

MARCUSCHI, Luiz Antônio. A questão do suporte dos gêneros textuais (parte 2). **DLCV**, João Pessoa, v. 1, n. 1, p. 09-40, 2003. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/dclv/article/view/7435>. Acesso em: 04 jul. 2023.

MARTINS, Gerson Luiza, REINO, Lucas S. Arraes, BUENO, Thaís (org.). **Performance em ciberjornalismo**. Tecnologia, inovação e eficiência. Campo Grande: Ed. UFMS, 2017.

MCKENZIE, Donald Francis. **Bibliografía e a sociología dos textos**. São Paulo: EDUSP, 2018.

MEDEIROS, Nuno. Acções prescritivas e estratégicas: a edição como espaço social. **Revista Crítica de Ciências Sociais**, Coimbra, Portugal, n. 85, p. 131-146, 2009.

MEDINA, Eden. Rethinking algorithmic regulation, **Kybernetes**, S.l., v. 44, n. 6/7, p. 1005-1019, 2015.

METAXA, Danaë; TORRES-ECHEVERRY, Nicolás. Google's role in spreading fake news and misinformation. **SSRN.com**, S.l., p. 01-31, out. 2017.

MOTTA, Luiz Gonzaga. Análise pragmática da narrativa jornalística. *In*: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 28., Rio de Janeiro, 2005. **Anais...** Rio de Janeiro: Intercom, 2005. p. 05-09.

MULDOON, James. **Platform socialism**: how to reclaim our digital future from Big Tech. London: Pluto Press, 2022.

MUNIZ JR., José de Souza. **Tinha um editor no meio do caminho**: questões contemporâneas de edição, preparação e revisão textual. Divinópolis: Artigo A, 2018.

MUNIZ JR., José de Souza. Edição. *In*: CABRAL, Cleber Araújo; RIBEIRO, Ana Elisa (org.). **Tarefas da Edição**. Belo Horizonte: DEDC, LED, Impressões de Minas, 2020. p. 68-72. Disponível em: <https://www.letras.bh.cefetmg.br/wp-content/uploads/sites/193/2019/10/Tarefas-da-Edic%cc%a7a%cc%83o-arquivo-digital-07-10-20.pdf>. Acesso em: 03 jul. 2023.

NAYAK, Pandu. Understanding searches better than ever before. **Google The Keyword**, 25 out. 2019. Disponível em: <https://blog.google/products/search/search-language-understanding-bert/>. Acesso em: 12 jun. 2023.

NIEBORG, David. B.; POELL, Thomas. The platformization of cultural production: Theorizing the contingent cultural commodity. **New Media and Society**, London, v. 20, n. 11, p. 4275-4292, 2018.

NIELSEN, Jakob. **Usabilidade na web**. São Paulo: Elsevier Brasil, 2007.

NGUYEN, George. BERT is rolling out to Google search in over 70 languages. **Search Engine Land**, 09 dez. 2019. Disponível em <https://searchengineland.com/bert-is-rolling-out-to-google-search-in-over-70-languages-326146>. Acesso: 12 jun. 2023.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. Seminário Regulação de Plataformas Digitais no Brasil – Parte 1 (Manhã), São Paulo, 01 set. 2022. YouTube: nic.br. Disponível em: <https://www.youtube.com/watch?v=iGqx8Z96eUc&list=PLQq8-9yVHyObJyyjdRDttwq0YtTN-sNUv>. Acesso em: 14 set. 2022.

O DILEMA das Redes Sociais. Direção: Jeff Orlowski. Estados Unidos: **Netflix**, 2020. 94 min.

OCHIGAME, Rodrigo. **Remodeling Rationality**: an inquiry into unorthodox modes of logic and computation. 2021a. 235 f. Tese (Doutorado em História, Antropologia, Ciência, Tecnologia e Sociedade) – Massachusetts Institute of Technology, Massachusetts, 2021a.

OCHIGAME, Rodrigo. A política dos buscadores. seminário automação e digitalização no capitalismo contemporâneo. 2021b Disponível em: <https://www.youtube.com/watch?v=8vcGD0dgMH4>. Acesso em: 03 out. 2022.

OCHIGAME, Rodrigo; HOLSTON, James. Filtering dissent social media and land struggles in Brazil. Berkeley: **Department of Anthropology**, 2016. Disponível em: <https://escholarship.org/uc/item/0sw0m8ww>. Acesso em: 19 set. 2022.

OLIVER, Julian, SAVICIC Gordan, VASILIEV, Danja. The critical engineering manifesto. **The Critical Engineering Working Group**, Berlin, out. 2011-2021. Disponível em: <https://criticalengineering.org/pt>. Acesso em: 03 out. 2022.

O'REILLY, Tim. Web 2.0 Compact Definition: trying again. **Radar**, 10 dez. 2006. Disponível em: <http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html>. Acesso em: 18 mar. 2022.

PIERRE, Lévy. **Cibercultura**. São Paulo: Editora 34, 2010.

PLANTIN, Jean Christophe *et al.* Infrastructure studies meet platform studies in the age of Google and Facebook. **New Media and Society**, London, v. 20, n. 1, p. 293-310, 2016.

POELL, Thomas; NIEBORG, David; VAN DIJCK, José. Plataformização. **Fronteiras - estudos midiáticos**, São Leopoldo, v. 22, n. 1, p. 02-10, 2020.

POSTMAN, Neil. Informing ourselves to death. *In*: POSTMAN, Neil. **The nature of technology**. Leiden: Brill, 2013. p. 05-14.

POSSENTI, Sírio. O que é um bom texto hoje. Belo Horizonte, 2020. Palestra.

RIBEIRO, Ana Elisa. Leituras sobre hipertexto: trilhas para o pesquisador. *In*: SIMPÓSIO NACIONAL DE LETRAS E LINGUÍSTICA E I SIMPÓSIO INTERNACIONAL DE LETRAS E LINGUÍSTICA, 11. e 1., Uberlândia, 2006. **Anais...** Uberlândia: UFU, 2006.

RIBEIRO, Ana Elisa. Tecnologia e poder semiótico: escrever, hoje. **Texto Livre: Linguagem e Tecnologia**, Belo Horizonte, v. 8, n. 1, p. 112-123, 2015.

RIBEIRO, Ana Elisa. **Textos multimodais: leitura e produção**. São Paulo: Parábola Editorial, 2016.

RIBEIRO, Ana Elisa. **Escrever, hoje: palavra, imagem e tecnologias digitais na educação**. 1 ed. São Paulo: Parábola Editorial, 2018a.

RIBEIRO, Ana Elisa. **Livro: edição e tecnologias no século XXI**. Belo Horizonte: Moinhos, 2018b.

RIBEIRO, Ana Elisa. Livro. *In*: CABRAL, Cleber Araújo; RIBEIRO, Ana Elisa (org.). **Tarefas da Edição**. Belo Horizonte: DEDC, LED, Impressões de Minas, 2020. p. 85-88. Disponível em: <https://www.letras.bh.cefetmg.br/wp-content/uploads/sites/193/2019/10/Tarefas-da-Edic%cc%a7a%cc%83o-arquivo-digital-07-10-20.pdf>. Acesso em: 11 abr. 2022.

RIBEIRO, Ana Elisa. Audiência Pública realizada pela Comissão de Educação, Ciência e Tecnologia da Assembleia Legislativa de Minas Gerais, Belo Horizonte, 12 jul. 2023. YouTube: Assembleia de Minas Gerais. Disponível em: <https://www.youtube.com/watch?v=4k0Vm9YieGs>. Acesso em: 12 ago. 2023.

RIEDER, Bernhard; MATAMOROS-FERNÁNDEZ, Ariadna; COROMINA, Òscar. From ranking algorithms to “ranking cultures” Investigating the modulation of visibility in YouTube search results. **Convergence**, S.l., v. 24, n. 1, p. 50-68, 2018.

RIEDER, Bernhard; RÖHLE, Theo. Métodos digitais: dos desafios à bildung. *In*: OMENA, Janna Joceli (org.). **Métodos digitais: teoria-prática-crítica**. Lisboa: Icnova, 2019. p. 19-36.

RIEDER, Bernhard. **Engines of order: a mechnology of algorithmic techniques**. Amsterdam: Amsterdam University Press, 2020.

ROGERS, Richard. O fim do virtual: os métodos digitais. **Lumina**, Juiz de Fora, v. 10, n. 3, p. 01-34, 2016.

SALGADO, Luciana. **Ritos genéticos no mercado editorial: autoria e práticas de textualização**. 2007. 305 f. Tese (Doutorado em Estudos da Linguagem) – Universidade Estadual de Campinas, Campinas, 2007.

SALGADO, Luciana. O autor e seu duplo nos ritos genéticos editoriais. **Eutomia**, Recife, v. 1, n. 01, p. 525-546, 2008.

SANTAELLA, Lúcia. **Matrizes da linguagem e pensamento: sonora, visual, verbal**. 3. ed. São Paulo: Iluminuras/Fapesp, 2005.

SANTAELLA, Lucia. **Navegar no ciberespaço: o perfil cognitivo do leitor imersivo**. São Paulo: Paulus, 2004.

SARDÁ, Thaís; LUPINACCI, Ludmila; BARBOSA, Camila C; ZAGO, Gabriela; VALIATI, Vanessa; ARAÚJO, Willian. A buzzfeedização do jornalismo: 5 coisas que você precisa saber sobre o caso Zero Hora. *In*: SIMPÓSIO NACIONAL DA ABCIBER, 8., São Paulo, 2014. **Anais...** São Paulo: ABCiber, 2014.

SCORE 100/100 with Rank Math post tests. **Rank Math**, 2023. Disponível em: <https://rankmath.com/kb/score-100-in-tests/>. Acesso em 29 jun 2023.

SILVA, Daniel Reis. **Relações Públicas, ciência e opinião: lógicas de influência na produção de (in)certezas**. 2017. 336 f. Tese (Doutorado em Comunicação) – Universidade Federal De Minas Gerais, Belo Horizonte, 2017.

SILVERMAN, David. **Interpretação de dados qualitativos: método para análise de entrevistas, textos e interações**. Tradução de Magda Franca Lopes. Porto Alegre: Artmed, 2009a.

SILVERMAN, David. **Um livro bom, pequeno e acessível sobre pesquisa qualitativa**. Porto Alegre: Bookman Editora, 2009b.

SLAWSKI, Bill. The Google Hummingbird update and the likely patent behind hummingbird. **SEO by the sea**, 27 set. 2013. Disponível em <https://www.seobythesea.com/2013/09/google-hummingbird-patent/#:~:text=The%20hummingbird%20Patent%20tells%20us,tend%20to%20appear%20together%20in>. Acesso em: 08 jun. 2023.

SOARES, Magda. Novas práticas de leitura e escrita: letramento na cibercultura. **Educação & Sociedade**, Campinas, v. 23, p. 143-160, 2002.

SOARES, Verônica; CRUZ, Luana Teixeira de Souza. Textos de ciência em ambiente digital: percurso exploratório em Nexa e Vox sobre vacinas. **Estudos em Jornalismo e Mídia**, Florianópolis, v. 17, n. 2, p. 175-184, 2020.

SCHIFFRIN, André. **O dinheiro e as palavras**. São Paulo: BEÏ Comunicação, 2011.

SCHOLZ, Trebor. **Cooperativismo de plataforma: contestando a economia do compartilhamento corporativo**. São Paulo: Fundação Rosa Luxemburgo; Editora Elefante; Autonomia Literária, 2016.

SRNICEK, Nick. **Platform capitalism**. Londres: Polity, 2016.

THE UNITED STATES DEPARTMENT OF JUSTICE. Antitrust Division. **United States et al versus Google LLC**. 2022. Disponível em: <https://www.justice.gov/atr/case-document/file/1536456/download>. Acesso em: 04 out. 2022.

THOMPSON, John B. **A mídia e a modernidade**. Petrópolis: Vozes, 1998.

THOMPSON, John. B. The New Visibility. **Sage**, London, v. 22, p. 31-51, 2005.

THOMPSON, John. B. **Mercadores de Cultura: o mercado editorial do século XXI**. São Paulo: Editora Unesp, 2013.

TOONEN, Edwin. Google BERT: a better understanding of complex queries. **Yoast SEO for everyone**, 04 nov. 2019. Disponível em: <https://yoast.com/google-bert/>. Acesso em: 12 jun. 2023.

TRAQUINA, Nelson. **Teorias do Jornalismo**. Por que as notícias são como são. 3. ed. Florianópolis: Insular, 2012.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. Escola de Ciência da Informação. O papel dos algoritmos e das plataformas digitais em contextos sociopolíticos. S.l., 2022. YouTube: ECI UFMG. Disponível em: <https://www.youtube.com/watch?v=nzRdD6XiGA4>. Acesso em: 13 jul. 2022.

USZKOREIT, Kakob. Transformer: a novel neural network architecture for language understanding. **Google Research**, 31 ago. 2017. Disponível em: <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>. Acesso em: 15 maio 2023.

VAIDHYANATHAN, Siva. **A Googlelização de tudo**. São Paulo: Editora Cultrix, 2011.

VAN DEN BOSCH, Antal; BOGERS, Toine; DE KUNDER, Maurice. Estimating search engine index size variability: a 9-year longitudinal study. **Scientometrics**, S.l., v. 107, n. 2, p. 839-856, 2016.

VAN DIJCK, José. Confiamos nos dados? As implicações da datificação para o monitoramento social. **MATRIZES**, São Paulo, v. 11, n. 1, p. 39-59, 2017.

VAN DIJCK, José. A sociedade da plataforma: entrevista com José van Dijck. **Digilabour**, 06 mar. 2019. Disponível em: <https://digilabour.com.br/a-sociedade-da-plataforma-entrevista-com-jose-van-dijck/>. Acesso em: 18 ago. 2022.

VAN DIJCK, José.; POELL, Thomas.; WAAL, Martijn. **The Platform Society: public values in a connective world**. Londres: Oxford Press, 2018.

VANDENDORPE, Christian. Variedades do hipertexto. *In*: RIBEIRO, Ana Elisa *et al.* **O hipertexto em tradução**. Belo Horizonte: UFMG, 2007. p. 60-69.

VECCHIO, Pollyanna de Matos Moura. **O direito de publicar: autopublicação, fast-publishing e tecnologias digitais no mercado editorial brasileiro**. 2022. 256f. Tese (Doutorado em Estudos de Linguagens) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2022.

VIEIRA, Lívia de Souza. **Métricas editoriais no jornalismo online: ética e cultura profissional na relação com audiências ativas**. 2018. 393 f. Tese (Doutorado em Jornalismo) – Universidade Federal de Santa Catarina, Florianópolis, 2018.

VIANA, NATÁLIA. Eu, uma escrava dos robôs. **Agência Pública**, 13 mar. 2023. Disponível em: [https://mailchi.mp/apublica.org/newsletter-xeque-na-democracia-030?e=9647d7e595 &utm_source=substack &utm_medium=email](https://mailchi.mp/apublica.org/newsletter-xeque-na-democracia-030?e=9647d7e595&utm_source=substack&utm_medium=email). Acesso em: 20 mar. 2023.

VILLAS BÔAS, Luciana. **A República de chinelos: Bolsonaro e o desmonte da representação**. São Paulo: Editora 34, 2022.

VOLPATO, Bruno. Google anuncia BERT, seu novo algoritmo de pesquisa. **Resultados Digitais**, 28 out. 2019. Disponível em: <https://resultadosdigitais.com.br/marketing/google-bert/>. Acesso em: 12 jun. 2023.

WINNER, Langdon. Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. **Science, technology, & human values**, S.l., v. 18, n. 3, p. 362-378, 1993.

WINQUES, Kérley; LONGHI, Raquel Ritter. Dos meios às mediações (algorítmicas): mediação, recepção e consumo em plataformas digitais. **MATRIZES**, São Paulo, v. 16, n. 2, p. 151-172, 2022.

YOAST SEO. The Flesch reading ease score: why and how to use it. **Yoast SEO for everyone**, 20 maio 2023. Disponível em: <https://yoast.com/flesch-reading-ease-score/>. Acesso em: 29 jun. 2023.

ZICHERMANN, Gabe; CUNNINGHAM, Christopher. **Gamification by design: Implementing game mechanics in web and mobile apps**. Sebastopol: O'Reilly Media Inc., 2011.

ZILLER, Joana. Segunda Kitnet Mais Bonita da Internet: táticas de Certeau nos vídeos do YouTube. *In*: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 35., Fortaleza, 2012. **Anais...** Fortaleza: Intercom, 2012.

ZILLER, JOANA *et al.* Lesbianidades em rede: visibilidades e invisibilidades no YouTube. *In*: SIMPÓSIO INTERNACIONAL LAVITS, 6., Salvador, 2019. **Anais...** Salvador: Lavits, 2019.

ZIMMER, Michael. Privacy on planet Google: using the theory of contextual integrity to clarify the privacy threats of Google's quest for the perfect search engine. **Journal of Business & Technology Law**, S.l., v. 3, n. 1, p. 108-126, 2008.

ZIMMER, Michael. Web search studies: Multidisciplinary perspectives on web search engines. **International handbook of internet research**. Springer, p. 507-521, 2009.