



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE CIÊNCIA POLÍTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA POLÍTICA**

ULISSES MATHEUS BRAGA DE FREITAS MELO

**FEITA SOB MEDIDA: A ESTRUTURA DE UMA NOTÍCIA FALSA E SEU PAPEL
NO CONVENCIMENTO DO ELEITOR**

Recife
2022

ULISSES MATHEUS BRAGA DE FREITAS MELO

**FEITA SOB MEDIDA: A ESTRUTURA DE UMA NOTÍCIA FALSA E SEU PAPEL
NO CONVENCIMENTO DO ELEITOR**

Dissertação de Mestrado apresentada como um dos requisitos para a obtenção do título de Mestre em Ciência Política pelo Programa de Pós-Graduação em Ciência Política da Universidade Federal de Pernambuco.

Área de concentração: Democracia e Instituições

Orientadora: Prof.^a Dr.^a Nara de Carvalho Pavão

Recife
2022

Catalogação na fonte
Bibliotecária Maria do Carmo de Paiva, CRB4-1291

M528f Melo, Ulisses Matheus Braga de Freitas.

Feita sob medida : estrutura de uma notícia falsa e seu papel no
convencimento do eleitor / Ulisses Matheus Braga de Freitas Melo. – 2022.
104 f. : il. ; 30 cm.

Orientadora: Profª. Drª. Nara de Carvalho Pavão.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, CFCH.
Programa de Pós-Graduação em Ciência Política, Recife, 2022.
Inclui referências e apêndices.

1. Ciência Política. 2. Redes sociais on-line. 3. Desinformação. 4. Fake news. 5. Eleitores. I. Pavão, Nara de Carvalho (Orientadora). II. Título.

320 CDD (22. ed.)

(BCFCH2022-040)

ULISSES MATHEUS BRAGA DE FREITAS MELO

**FEITA SOB MEDIDA: A ESTRUTURA DE UMA NOTÍCIA FALSA E SEU PAPEL
NO CONVENCIMENTO DO ELEITOR**

Dissertação de Mestrado apresentada como um dos requisitos para a obtenção do título de Mestre em Ciência Política pelo Programa de Pós-Graduação em Ciência Política da Universidade Federal de Pernambuco.

Aprovada em 22/02/2022

BANCA EXAMINADORA

participação por videoconferência

Prof^a. Dr^a. Nara de Carvalho Pavão (Orientadora)

Universidade Federal de Pernambuco

participação por videoconferência

Prof. Dr. Davi Cordeiro Moreira (Examinador Interno)

Universidade Federal de Pernambuco

participação por videoconferência

Prof. Dr. Juliano Mendonça Domingues da Silva (Examinador Externo)

Universidade Católica de Pernambuco

Dedico este trabalho à memória das mais de 600 mil vítimas da COVID-19 no Brasil. Também a todos os cientistas que lutaram e lutam diariamente, até mesmo contra os poderes instituídos, para levar informações necessárias à população, e aos pesquisadores que produziram a saída para a maior pandemia dos últimos cem anos.

AGRADECIMENTOS

À Edjane e Djair, meus pais, que buscaram me incentivar a estudar desde cedo e aceitaram o caminho que escolhi trilhar, sem eles e a sua criação não estaria aqui hoje.

À Nielline, noiva e companheira, por estar ao meu lado nos momentos mais tensos do mestrado e pela paciência por escutar muitos dos questionamentos que estão presentes nessa dissertação dezenas de vezes com paciência e atenção.

Aos meus amigos, Luiz, que já leu – e corrigiu – tantas vezes este e outros trabalhos meus, e Luan, pela disponibilidade durante o processo de seleção, sem ele não conseguiria a aprovação para o mestrado.

Agradeço também à prof. Dr^a. Nara Pavão, sem ela esse trabalho não teria sido realizado, meu caminho pela Ciência Política passa por ela.

Registro minha gratidão ao prof. Dr. Dalson Figueiredo, que mudou a forma como eu entendia a produção de conhecimento e me apresentou a importância da transparência e replicação para o método científico.

Agradeço a todos que fazem parte do Centro de Pesquisa sobre Comportamento Político e Opinião Pública, sem a atenção e as dicas que vocês me deram este trabalho nunca chegaria tão longe.

À todos que lutam pela democracia e combatem os autoritarismos diários pelo quais passamos nos últimos anos.

RESUMO

O espalhamento de desinformação *on-line* se tornou um problema para as democracias contemporâneas, a crença e o compartilhamento de notícias falsas são práticas incentivadas por diversos setores da elite política e conseguem interferir diretamente no comportamento dos indivíduos, principalmente daqueles partidários. Este trabalho, contudo, busca compreender quais são as características estruturais do texto desinformativo, e quais os possíveis efeitos que tais particularidades podem ter no consumo desse tipo de material. A pesquisa analisou centenas de notícias falsas e verdadeiras através de diversas técnicas de análise de texto automatizada fornecidas pelos softwares Iramuteq e ALT, sendo observadas as diferenças estatísticas entre o jornalismo tradicional e o desinformativo. Os resultados encontrados demonstram uma maior presença de nomes de lideranças e ideologias políticas, bem como a maior citação de instituições entre as notícias falsas, termos associados à cognição quente pela literatura especializada, um forte ativador do raciocínio motivado. Além disso, as desinformações são menores e significativamente mais legíveis que as notícias tradicionais, característica ligada ao uso de processamento heurístico de informações. As notícias falsas compartilhadas *on-line*, portanto, se enquadram nos diversos pressupostos teóricos, apontados pelas principais linhas de pesquisa contemporânea, como fortes incentivadores de consumo e da crença de desinformação. Tais descobertas demonstram o potencial crítico que esse tipo de comunicação política possui, e como até mesmo sua estrutura pode ser parte desse processo.

Palavras-Chave: notícias falsas; comportamento político; estrutura léxica; análise de texto automatizada.

ABSTRACT

The spread of disinformation *on-line* has become a problem for contemporary democracies, the belief and sharing of fake news are practices encouraged by various sectors of the political elite and it can directly interfere in the behavior of individuals, especially those who are partisans. This work, however, seeks to understand what are the structural characteristics of misinformative text, and what are the possible effects that such particularities can have on the consumption of this type of material. The research analyzed hundreds of false and true news through several automated text analysis techniques, by using Iramuteq and ALT software, observing the statistical differences between traditional and misinformative journalism. The results demonstrate a big amount of names of leaders and political ideologies, as well as citation of institutions among the fake news, terms associated with hot cognition by the specialized literature, a strong activator of motivated reasoning. Furthermore, disinformation is smaller and significantly more readable than traditional news, a characteristic linked to the use of heuristic information processing. Thus, *on-line* shared fake news fits into several theoretical assumptions, pointed out by the main lines of contemporary research, as strong encouragers of consumption and the belief of misinformation. These findings demonstrate the critical potential that this type of political communication has, and how even its structure can be part of this process.

Keywords: fake news; political behavior; lexical structure; automated text analysis.

SUMÁRIO

1	INTRODUÇÃO	10
2	A DESINFORMAÇÃO.....	10
2.1	DELIMITANDO O(S) CONCEITO(S).....	14
2.2	A DESINFORMAÇÃO E A POLÍTICA	17
3	DEMANDA E OFERTA: COMO O CONSUMO DE DESINFORMAÇÃO SE RELACIONA COM A SUA ESTRUTURA TEXTUAL?	23
3.1	AS REDES SOCIAIS, NOSSAS IDENTIDADES E EMOÇÕES	24
3.2	A RACIONALIDADE MOTIVADA E A COGNIÇÃO QUENTE	28
3.3	A PREGUIÇA E A SIMPLICIDADE	31
3.4	AS PARTICULARIDADES DO CONTEÚDO POLÍTICO	33
3.5	DISCERNINDO MENTIRAS E VERDADES	34
3.6	DISCERNINDO NOTÍCIAS FALSAS E VERDADEIRAS.....	36
4	DADOS E MÉTODO	40
4.1	BANCOS DE DADOS	40
4.2	METODOLOGIA E TÉCNICAS APLICADAS.....	41
4.2.1	Análise de Conteúdo e Análise de Texto Automatizada	41
4.2.2	Análise de conteúdo automatizada a partir do Iramuteq.....	43
4.2.3	Análise da legibilidade	46
5	RESULTADOS	48
5.1	ANÁLISE DE ESPECIFICIDADES E PRIMEIROS TESTES	48
5.2	LEGIBILIDADE	60
5.3	TEMAS	67
6	LIMITAÇÕES E AGENDA DE PESQUISA	77
7	CONCLUSÃO	79
	REFERÊNCIAS	81
	APÊNDICE A – ANÁLISE ESTATÍSTICA EXPLORATÓRIA (NOTÍCIAS COM TEMÁTICA POLÍTICA)	88
	APÊNDICE B – ANÁLISE ESTATÍSTICA EXPLORATÓRIA (NOTÍCIAS SEM TEMÁTICA POLÍTICA)	90
	APÊNDICE C – AFC DO BANCO DE DADOS GERAL.....	91
	APÊNDICE D - ANÁLISE EXPLORATÓRIA DAS NOTÍCIAS VERDADEIRAS VIA CHD	92

APÊNDICE E – ANÁLISE FATORIAL POR CORRESPONDÊNCIA (NOTÍCIAS VERDADEIRAS)	96
APÊNDICE F – ANÁLISE EXPLORATÓRIA DAS NOTÍCIAS FALSAS VIA CHD	97
APÊNDICE G – ANÁLISE FATORIAL POR CORRESPONDÊNCIA (NOTÍCIA FALSA).....	100
APÊNDICE H – FASES DA COLETA E TRATAMENTO DOS DADOS NO IRAMUTEQ.....	101
APÊNDICE I – TRATAMENTO DOS DADOS E APLICAÇÕES DE TESTES NO ALT (LEGIBILIDADE.COM).....	103
APÊNDICE J – TERMOS LIGADOS AO NOTICIÁRIO POLÍTICO SELECIONADOS PARA CRIAÇÃO DE SUBCORPUS	104

1 INTRODUÇÃO

Qual é a estrutura textual de uma notícia falsa? O modo como a desinformação é escrita tem alguma relação com o seu consumo e o compartilhamento *on-line*? Este trabalho busca responder a essas e outras perguntas, demonstrando como o formato das *fake news* se adaptou à forma como processamos informações na era da internet, sendo mais chamativas, simples e “viralizáveis” que as matérias e artigos jornalísticos tradicionais. As descobertas feitas aqui aprofundam a compreensão sobre esse novo formato de comunicação política e permitem o debate sobre o desenvolvimento de uma nova estratégia de combate à desinformação a partir do olhar atento do leitor.

O compartilhamento em massa de desinformação vem sendo objeto de diversas análises desde os pleitos eleitorais de 2016 nos EUA e Reino Unido. Essas duas eleições ficaram marcadas pela grande circulação de notícias falsas durante o período de campanha (ALLCOTT; GENTZKOW, 2017; KUMAR; SHAH, 2018), e pela utilização desse tipo de material no pronunciamento de lideranças políticas. Tal modelo de interação entre a elite partidária e sua militância deu tão certo que foi replicado em outros países, como o Brasil, alcançando um sucesso tão grande quanto (DOURADO, 2020; RECUERO; GRUZD, 2019).

De fato, observa-se que tal formato de comunicação política se estabeleceu em diversas democracias nos últimos anos, tanto que o termo “*fake news*” é cada vez mais o centro de atenção e debate por parte da população e mídia¹. Os casos brasileiro e norte-americano, inclusive, demonstram que o uso desse expediente comunicacional pode se manter mesmo após o período eleitoral, fazendo parte do dia a dia da elaboração e divulgação de políticas públicas (SOARES *et al.*, 2021) e gestão de instituições (PAVARIN, 2019). As notícias falsas, portanto, acabaram se tornando uma ferramenta fundamental de gestão de comunicação de campanhas e governos em diversas democracias pelo mundo e são peças fundamentais para o processo de polarização afetiva crescente nesses regimes (IYENGAR *et al.*, 2019).

Entretanto, apesar da exploração considerável que a academia vem fazendo sobre esse tema nos últimos anos (LAZER *et al.*, 2018; NYHAN, 2020), abordando a facilidade com que esse tipo de material “viraliza” nas redes (VOUSOUGHI *et al.*, 2018) e os processos mentais que utilizamos ao consumi-los (FLYNN *et al.*, 2017; PENYCOOK *et al.*, 2020), ainda existem

¹ Os dados disponibilizados pelo Google Trends demonstram como o termo *fake news* passou a ser utilizado fortemente a partir de outubro de 2016. Ver mais em: <https://trends.google.com.br/trends/explore?date=all&geo=US&q=rumour,fake%20news,misinformation,disinformation>.

poucos trabalhos que buscam estudar as particularidades textuais da desinformação. A grande maioria das pesquisas que abordam o conteúdo das notícias falsas como objeto de estudo tentam promover modelos preditivos capazes de apontar automaticamente a falsidade do material, sem se questionar muito sobre a influência que as variáveis textuais possuem para o consumo e o compartilhamento *on-line* (MONTEIRO *et al.*, 2019; CHOY; CHONG, 2018; RASHKIN *et al.*, 2017; FREIRE; GODSCHIMIDT, 2019).

Este estudo, portanto, busca avançar sobre esse campo ainda pouco explorado, principalmente dentro da ciência política. Ele almeja contribuir para o melhor entendimento do fenômeno político-comunicacional recente, que é a desinformação *on-line*, verificando, inclusive, possíveis particularidades do conteúdo político neste formato, como a maior legibilidade de notícias falsas que abordam a política como tema principal. Análises como esta podem apontar para novos caminhos a serem trilhados no combate à desinformação *on-line*, apresentando características estruturais significativas.

Para tanto, será utilizada a metodologia de análise de texto automatizada (OLIVEIRA, 2019) através dos *softwares* Iramuteq (RATINAUD, 2009) e ALT (legibilidade.com) na base de dados Fake.Br (MONTEIRO *et al.*, 2019). Esta estratégia metodológica permite, através do tratamento estatístico do conteúdo, a observação de tendências gramaticais, temáticas e de legibilidade, que estão atreladas ao formato de comunicação presente nas desinformações e até mesmo à forma como processamos mentalmente as notícias que consumimos. Para verificar a existência de diferenças significativas serão aplicados testes ligados à contagem de palavras, concentração de termos e categorias gramaticais (Análise de Especificidades), eixos temáticos (Classificação Hierárquica Descendente) e facilidade de leitura (Índices de Legibilidade Flesch-Kincaid e Coleman-Liau). O trabalho também disponibilizará material para a replicação dos testes realizados em um banco de dados *on-line*².

A pesquisa espera encontrar diferenças significativas entre a estrutura das notícias falsas e verdadeiras por três motivos centrais. Primeiro, baseando-se na literatura que trata dos modelos de processamento de informação por parte dos consumidores de desinformação, tanto no que se refere ao raciocínio motivado (FLYNN *et al.*, 2017), como a que aborda o modelo heurístico de processamento de comunicações, como um fator importante para o consumo e compartilhamento deste tipo de material (PENNYCOOK; RAND, 2021). Segundo, os trabalhos que demonstram que as mentiras (VRIJ, 2008) e a desinformação (HORNE; ADALI, 2017)

² Os dados utilizados e os testes feitos neste trabalho estão disponíveis na plataforma GitHub no seguinte link: https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo.

possuem um formato de escrita diferente das informações verdadeiras. Terceiro, pelo desapego ao uso do padrão jornalístico de escrita por parte dos redatores das notícias falsas (GOMES; DOURADO, 2019).

As hipóteses apresentadas nas próximas páginas defendem que as desinformações têm uma presença maior de termos politicamente delicados e divisórios, bem como mais adjetivações morais e conceitos sociopolíticos em seu texto, características que ativam a cognição quente do leitor e, consequentemente, incentivam o uso de raciocínio motivado (LOGUE; TABER, 2005). Além disso, se espera que esse tipo de material contenha menos palavras, substantivos e verbos, e seja mais legível que as notícias verdadeiras, promovendo uma leitura simples, capaz de ativar com mais facilidade o processamento heurístico (PENNYCOOK; RAND, 2021; AGNHOTRI; BHATTACHARYA, 2016). Por outro lado, pronomes pessoais, adjetivos e verbos devem apresentar uma maior concentração entre as informações falsas (MONTEIRO *et al.*, 2019). Por fim, acredita-se que todas essas diferenças estruturais sejam ressaltadas em material com temática política devido a uma possível especialização dos redatores desse conteúdo.

O trabalho acabou por encontrar resultados que corroboram com parte dessas hipóteses. As desinformações observadas apresentaram, em média, menos palavras, preposições, adjetivos numéricos e substantivos que as verdadeiras, e um número consideravelmente maior de verbos, artigos definidos e adjetivos. Também foi encontrada uma maior presença de adjetivos morais e conceitos sociopolíticos entre as informações falsas. Outra variável que apresentou um resultado significativo foi a legibilidade, o conteúdo falso possuía - de forma estável - uma escrita mais simples e acessível que o verdadeiro. Por fim, a pesquisa também conseguiu verificar que estas características lexicais são ressaltadas em notícias com temática política, ou seja, a diferença estrutural entre as informações falsas e verdadeiras com conteúdo político é maior que a presente em materiais similares que abordam outros temas (esportes, celebridades, saúde etc.).

A apresentação da pesquisa que se segue está dividida em sete partes: a primeira aborda os conceitos ligados ao fenômeno do compartilhamento de desinformação na internet e delimita os termos referentes ao objeto de pesquisa; a segunda revisa os estudos mais promissores sobre as motivações para a crença, o consumo e o compartilhamento de informações falsas, bem como sua estrutura; a terceira parte detalha a metodologia, as técnicas e as bases de dados utilizadas; a quarta apresenta os resultados obtidos; a quinta discute a relação destes dados com a literatura sobre o consumo de desinformação política e apresenta

novas discussões a partir dos resultados; a sexta levanta possíveis limitações ligadas à metodologia e a própria área de estudo; por fim, a conclusão reúne os principais achados do trabalho e projeta uma nova agenda de pesquisa a partir do que foi discutido.

2 A DESINFORMAÇÃO

2.1 DELIMITANDO O(S) CONCEITO(S)

Antes de promover qualquer avanço na revisão dos estudos sobre a estrutura da desinformação *on-line* é preciso estabelecer, para efeitos de delimitação, do que se trata esse termo e em que contexto está o debate sobre sua aplicação. Isso porque há, no momento, uma verdadeira cacofonia sobre a utilização de conceitos que façam referência às informações duvidosas.

Por muito tempo a palavra “boato” dominou o ambiente midiático e acadêmico, sendo tratada como um fenômeno ligado ao jornalismo marrom³ (DOURADO, 2020) e tendo pouco espaço no estudo da política. O uso do termo “boato” geralmente se aplica a uma notícia que, embora não corresponda à realidade, é de difícil comprovação e, além disso, de grande interesse, o que determina sua intensa circulação entre as pessoas (HOYLER, 1966). Ao se referir a uma informação como um “boato”, normalmente, remete-se a um acontecimento que pode vir a ser comprovado, mas que, no momento da fala ou da comunicação, está circulando entre as pessoas sem a devida checagem.

Com o tempo, a criação de novos meios de comunicação mudou esse cenário. A expansão da internet trouxe consigo novos formatos de interação social, possibilitando o surgimento de modalidades inéditas de informações enganosas – como o *spam*⁴ e o *hoax*⁵. Ainda assim, o foco dessas mensagens, geralmente espalhadas via *e-mails*, eram golpes financeiros e propagandas comerciais. O noticiário político continuava consideravelmente afastado deste ambiente.

Após essa fase inicial, marcada por sites com elementos relativamente estáveis e interações por *e-mails* e mídias físicas, outras formas de comunicação surgiram na internet. Essa nova fase, que ficaria conhecida como *Web 2.0*, trouxe consigo um alto grau de

³ O termo jornalismo - ou imprensa - marrom é derivado da expressão inglesa *yellow press*, que se refere à mídias sensacionalistas que muitas vezes ultrapassam os limites da ética jornalística criando boatos e inflando histórias. Ver mais em: <https://super.abril.com.br/mundo-estranho/como-surgiu-a-expressao-imprensa-marrom/>.

⁴ “O SPAM é uma mensagem eletrônica que chega ao usuário sem a sua permissão ou sem seu desejo em recebê-lo. Geralmente são recebidas por e-mail, mas também podem circular pelas redes sociais ou comentários de blogs. O SPAM tem um fundo geralmente comercial, mas também pode assumir um viés criminoso.”. Ver mais em: <https://www.techtudo.com.br/noticias/noticia/2016/07/o-que-e-spam.html>.

⁵ “O *hoax*, como é chamado qualquer boato ou farsa na Internet, pode espalhar vírus entre os seus contatos.” Por algum tempo, essa era a forma mais comum de se referir a informações falsas compartilhadas online, principalmente via e-mail. Nas primeiras redes sociais de sucesso o termo também era popularmente utilizado. Ver mais em: <https://www.techtudo.com.br/artigos/noticia/2013/05/o-que-e-hoax-e-como-fugir-das-farsas-da-internet.html>.

interatividade, colaboração e produção/consumo de conteúdo pelos próprios usuários (MARTINO, 2015, p. 13). Um modelo específico atingiu um nível de sucesso sem precedentes: a mídia social (NEWMAN *et al.*, 2019). A partir de ascensão das mídias sociais, as interações entre os indivíduos passaram por uma grande mudança, afinal elas tornavam possíveis a troca de informações diárias. Bastava abrir a *timeline* do *Facebook* ou *Twitter*, por exemplo, para saber o que seus amigos, celebridades ou políticos haviam publicado. A gestão das redes sociais passou a chamar a atenção dos organizadores de campanhas eleitorais.

As eleições brasileiras de 2010 e 2014 são exemplos interessantes de como campanhas de difamação já eram extremamente populares nesses ambientes⁶, e como em pouco tempo as estratégias foram se modificando. Ainda em 2010 os principais mecanismos utilizados para espalhar desinformação eram os *blogs*⁷, utilizados como fontes de informações falsas e teorias da conspiração. No pleito de 2014, contudo, o cenário já era diferente: o *Facebook* havia se tornado a principal ferramenta de comunicação política virtual e diversas estratégias de desinformação eram ali promovidas (GOMES; DOURADO, 2019). Esses eventos já demonstravam que as redes sociais haviam se tornado um campo de batalha, envolvendo narrativas e acusações falsas entre as militâncias e apoiadores políticos de todos os espectros ideológicos.

Porém, foi após as eleições estadunidenses e a votação do *Brexit*, ambas em 2016, que o cenário envolvendo informações enganosas sobre política mudou. A partir desse momento as *fake news* passaram a ser o centro de um debate que tratava desde a política eleitoral até algoritmos de redes sociais (LAZER *et al.*, 2018; ALLCOTT; GENTZKOW, 2017; GOMES; DOURADO, 2019).

O uso do termo “*fake news*”, inclusive, surge inicialmente com o advento do compartilhamento em massa de *links* de sites de conteúdo duvidoso nas redes sociais (ALLCOTT; GENTZKOW, 2017; KUMAR; SHAH, 2019). Naquele momento, muitos autores conceituaram as notícias falsas como informações fabricadas que imitavam o conteúdo da mídia de notícias na forma, mas não em processo organizacional ou intenção (LAZER *et al.*, 2018). Para alguns pesquisadores esse era, inclusive, o principal fator responsável por dissociar o novo fenômeno político das *fake news* das outras estratégias de difamação. O caráter pretensamente

⁶ O site Boatos.org compilou os 10 maiores boatos espalhados durante a campanha eleitoral de 2014. Ver mais em: <https://www.boatos.org/politica/especial-os-10-maiores-boatos-espalhados-durante-eleicoes-2014.html>.

⁷ Uma investigação feita em 2018 pela BBC Brasil apontou como funcionavam blogs em defesa da - à época candidata - Dilma Rousseff durante a campanha eleitoral de 2010: <https://www.bbc.com/portuguese/brasil-43118825>.

jornalístico da mentira era considerado um ponto fundamental de distinção desse novo formato envolvendo as informações falsas (GOMES; DOURADO, 2019).

Entretanto, com a popularização do termo na mídia, o conceito começou a ser usado como sinônimo de postagens virais de todo o tipo, seja das que careciam de checagem ou daquelas já comprovadamente falsas (TANDOC Jr. *et al.*, 2018). Além disso, com o tempo, atores políticos passaram a se utilizar do termo “*fake news*” para atacar mídias tradicionais quando estas apontavam alguma questão controversa sobre os seus mandatos e campanhas⁸, o que provocou, de certa forma, a derrocada do uso do termo por parte da academia. Os que defendem essa abordagem afirmam que não há uma delimitação conceitual que permita a operacionalização cuidadosa do conceito, sem que este caia em vieses e politização (WARDLE, 2018).

Esse movimento fez com que parte da literatura passasse a utilizar – e ressignificar – dois conceitos: *misinformation* e *disinformation*⁹. Hoje, muitos pesquisadores associam *misinformation* à informação incorreta distribuída sem a intenção de enganar; e *disinformation* à informação incorreta disseminada com a intenção de enganar o receptor do conteúdo (WARDLE, 2018). Ambos os conceitos não se utilizam de critérios formais de aparência nem dos espaços de compartilhamento, focando no conteúdo e na intenção dos criadores e compartilhadores da informação. Dessa forma, tanto a *misinformation* como a *disinformation* não se limitam ao conceito primevo de *fake news* (LAZER *et al.*, 2018), podendo incluir características não jornalísticas ao material enganoso.

Para manter uma certa coerência conceitual, este trabalho irá unificar esses significados. Tanto o conceito de “notícias falsas” (*fake news*), como o de “desinformação”, serão tratados como sinônimos de “informações comprovadamente falsas”. Não serão observados os critérios ligados à intenção do criador do conteúdo falso, mas apenas ao conteúdo em si. Esta classificação mais simplificada facilita a operacionalização das bases de dados que serão utilizadas neste trabalho, sendo mais objetiva que outros conceitos que exigem suposições sobre a origem e a intencionalidade dos criadores da desinformação.

⁸ Donald Trump usou o termo de forma recorrente para atacar parte da mídia (ver mais em: <https://oglobo.globo.com/mundo/fake-news-sete-ataques-do-governo-trump-imprensa-20978675>). Jair Bolsonaro também adotou esse expediente, inclusive contra instituições como o STF, ver mais em: <https://www.correiobrasiliense.com.br/politica/2021/07/4940501-bolsonaro-chama-nota-do-stf-de-fake-news-e-diz-que-corte-cometeu-crime.html>

⁹ Ambos os termos são traduzidos para o português como “desinformação”.

2.2 A DESINFORMAÇÃO E A POLÍTICA

A desinformação política não é algo novo no noticiário, seja por campanhas difamatórias ou boatos não comprovados, mentiras sobre instituições políticas e figuras de destaque foram utilizadas largamente na história. O Brasil, em especial, possui diversos exemplos de informações falsas espalhadas intencionalmente com o intuito de garantir uma certa narrativa política, tal como o Plano Cohen¹⁰ em 1937, os boatos sobre uma possível revolução comunista em 1964¹¹, ou ainda a estratégia de desinformação promovida antes do atentado terrorista do Riocentro¹².

Todavia, estudos sobre desinformação política começaram a ser desenvolvidos em maior escala apenas nos últimos anos. Apesar de análises seminais como o trabalho de Kuklinski (*et al.*, 2000), grande parte da produção acadêmica envolvendo a crença e a distribuição de informações falsas foi produzida na última década¹³. Essa ascensão ocorreu, em grande parte, como uma resposta ao crescimento do compartilhamento de informações falsas ligadas às novas tecnologias comunicacionais que surgiram e se desenvolveram neste período (como o *spam*, o *hoax* e os comentários difamatórios em sites comerciais).

Tanto a popularização do uso da internet, durante a primeira década dos anos 2000, quanto o consumo em massa de redes sociais geraram grandes comoções por parte de acadêmicos e jornalistas. Inicialmente, muito se falava sobre uma possível “democratização” do conhecimento (LAZER *et al.*, 2018; MARTINO, 2015, p. 10) e da revolução que ocorreria na forma como interagimos socialmente; por diversas vezes, esses comentários eram preenchidos de expectativas otimistas sobre nossa relação com esse amálgama de dados e informações. Essa excitação inicial, contudo, foi seguida por crises de expectativas envolvendo questionamentos sobre a qualidade dos conteúdos consumidos e possíveis problemas sociais advindos da forma como nos relacionamos com as redes (LAZER *et al.*, 2018; BRADY *et al.*, 2020).

¹⁰ Ver mais sobre o Plano Cohen em: <https://cpdoc.fgv.br/producao/dossies/AEraVargas1/anos30-37/GolpeEstadoNovo/PlanoCohen>.

¹¹ Ver mais sobre a ascensão do anticomunismo no pré-golpe de 1964: <https://apublica.org/2019/04/1964-o-brasil-nao-estava-a-beira-do-comunismo-diz-historiador/>.

¹² Pouco antes do atentado, dizeres com a sigla da VPR (Vanguarda Popular Revolucionária), organização de esquerda extinta em 1973, haviam sido pichadas nas proximidades do Riocentro. Ver mais em: <http://www.fgv.br/cpdoc/acervo/dicionarios/verbete-tematico/riocentro-atentado-do>.

¹³ Um bom exemplo de como essa produção vem crescendo nos últimos anos pode ser visto no visualizador de ngram do Google Books: https://books.google.com/ngrams/graph?content=fake+news%2Cmisinformation%2Cdisinformation%2CChoxes%2Crumor&year_start=1950&year_end=2019&corpus=26&smoothing=2.

As discussões sobre o consumo e a crença em informações falsas e teorias da conspiração passaram a ocupar o centro do debate político, institucional e acadêmico, principalmente a partir de 2016. Esse movimento ocorreu em grande parte devido ao sucesso das campanhas eleitorais de Donald Trump e do *Brexit* naquele ano e, posteriormente, à eleição de líderes populistas em todo o mundo, como Jair Bolsonaro (Brasil), Narendra Modi (Índia)¹⁴ e Rodrigo Duterte (Filipinas)¹⁵. Todos esses pleitos foram marcados pela distribuição em massa de informações falsas durante as respectivas campanhas eleitorais, principalmente nas redes sociais (ALLCOTT; GENTZKOW, 2017; DOURADO, 2020, 2021; BURSZTYN; BIRNBAUM, 2019; RECUERO; GRUZD, 2019; RESENDE *et al.*, 2018, 2019).

No entanto, talvez a grande quebra de paradigma entre os modelos de desinformação anteriores e essa nova leva de cascadas de informações falsas foi o papel decisivo das elites políticas na distribuição e compartilhamento desse tipo de conteúdo. Tanto Donald Trump, como Nigel Farage e Jair Bolsonaro marcaram os pleitos anteriormente citados por afirmações públicas envolvendo desinformação. O que antes era visto apenas como um comportamento da militância, passou a ser discurso oficial de campanha utilizado por lideranças políticas.

A produção acadêmica que surgiu logo após esses eventos, portanto, apontava para o uso em larga escala de informações falsas como ferramenta política nas mais diversas redes sociais, mas principalmente no *Twitter*, *Facebook* e *Whatsapp*. Alguns autores afirmavam que o barateamento da produção jornalística e a exigência de uma maior velocidade na apuração de conteúdo eram alguns dos principais fatores estruturais que colocavam em risco a qualidade da mídia tradicional, o que abria espaço para veículos hiperpartidários e disseminadores de notícias falsas (LAZER *et al.*, 2018). Outros estudos também descobriram que notícias falsas se espalham mais rapidamente e atingem um maior número pessoas nas redes sociais que as notícias verdadeiras (VOUSOUGHI *et al.*, 2018; ZHAO *et al.*, 2020). Essa literatura desenhava um cenário um tanto quanto problemático, demonstrando como o ambiente *on-line* era propício – e até mesmo estimulador – da adoção da desinformação como parte de uma estratégia de comunicação política.

No Brasil, diversos trabalhos também se voltaram para o problema das informações falsas. Alguns estudos catalogaram as várias formas de desinformação que estiveram circulando durante a campanha eleitoral de 2018 no *Twitter*, apontando o então candidato Jair Bolsonaro

¹⁴ Sobre o uso de desinformação online nas eleições indianas, ver mais em: <https://www.bbc.com/news/world-asia-india-47797151>.

¹⁵ Sobre o uso de desinformação online nas eleições filipinas, ver mais em: <https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war>.

como principal beneficiário desta circulação (DOURADO, 2020, 2021). Outros buscaram analisar como a estrutura de disseminação de informações falsas nas redes sociais já estava consolidada antes mesmo do pleito eleitoral (RECUERO; GRUZD, 2019; RESENDE *et al.*, 2018, 2019). Nesse último caso, as análises encontraram uma atividade considerável em pequenos grupos disseminadores de desinformação nas redes sociais já no início de 2018. Um evento chave para a descoberta dessa estrutura de compartilhamento foi o movimento realizado por caminhoneiros em março de 2018, pois, já naquele momento, havia uma distribuição em massa de materiais panfletários e, muitas vezes, desinformativos tanto no *Twitter* quanto – e principalmente – no *Whatsapp* (RESENDE *et al.*, 2018; PINHEIRO-MACHADO, 2019, p. 122).

Outras descobertas recentes envolvem as principais temáticas e informações debatidas entre os eleitores nas redes sociais (CHAGAS; CARREIRO, 2021; CHAGAS *et al.*, 2019; DOURADO 2020, 2021). Esses estudos encontraram fortes tendências antipetistas, anticomunistas, anticorrupção e antiesquerdistas, principalmente em redes ligadas ao candidato Jair Bolsonaro, através de um discurso que incutia diversos sentimentos entre os usuários, como, por exemplo, o medo de uma “venezuelização” do Brasil (CHAGAS *et al.*, 2019), o avanço de uma suposta ameaça comunista (CHAGAS; CARREIRO, 2021), ou mesmo a criação e a divulgação de promessas de campanha e frases fora de contexto de diversos candidatos (DOURADO, 2020, 2021). Eles também encontraram indicativos de uma certa centralização da fabricação dos materiais compartilhados, identificando alguns perfis extremamente ativos em mais de um desses grupos (CHAGAS *et al.*, 2019; PAVARIN, 2019).

É fato que a estratégia envolvendo a distribuição em massa de desinformação e a repercussão pela elite política desse material ainda é um sucesso, tanto que os líderes eleitos mantiveram tais métodos de desinformação após o período da campanha eleitoral. O mercado das notícias falsas é barato e – econômica e politicamente – lucrativo, e vem se tornando mais estável do que análises iniciais previam (ALLCOTT; GENTZKOW, 2017), mesclando-se com o jornalismo hiperpartidário ascendente e recebendo apoio do poder estabelecido (RECUERO *et al.*, 2021; SOARES *et al.*, 2021).

Os incentivos para o uso de tais estratégias são vários. As informações falsas parecem garantir o engajamento da militância e dos apoiadores políticos, e, além disso, as cascatas de desinformação permitem ao movimento político manipular o debate público ao seu favor e possibilitam o engajamento em torno de declarações absurdas – na maioria das vezes cheias de

desinformação, que ficaram conhecidas por *cortinas de fumaça*¹⁶. Ainda que não convençam o eleitor, as informações falsas garantem narrativas de defesa e ataque contra opositores estabelecendo uma agenda política específica, independentemente da grande mídia, algo considerado impensável anos atrás (McCOMBS, 2014).

A convicção desses atores políticos nessa estratégia parece tanta que, até mesmo quando a pandemia de COVID-19 se alastrou pelo mundo, Donald Trump e Jair Bolsonaro julgavam melhor adotar teorias da conspiração como modelos de políticas públicas (CARVALHO *et al.*, 2021). Nos EUA e em parte da Europa, as notícias falsas sobre as vacinas e o apoio de setores da extrema-direita à desinformação *antivaxx* podem ter atrapalhado seriamente a vacinação¹⁷. No caso específico dos Estados Unidos, a baixa taxa de imunização da população acabou por gerar uma “pandemia de não-vacinados”.

O mesmo fenômeno se repetiu no Brasil, onde informações falsas sobre as vacinas foram usadas recorrentemente por políticos de extrema-direita – entre eles o presidente da república – para recomendar o uso de tratamento medicamentoso não comprovado no combate da COVID-19 (SOARES *et al.*, 2021), o que gerou o uso dessas drogas por parte dos seus apoiadores¹⁸. Jair Bolsonaro, em especial, ainda atacou os governadores que proibiram aglomerações e fecharam o comércio durante o recrudescimento da pandemia¹⁹, afirmou que as máscaras – utilizadas para a proteção individual contra a COVID²⁰ – faziam mal²¹, encabeçou

¹⁶ A tese das cortinas de fumaça vem sendo levantadas desde o início do governo Bolsonaro, já que o mesmo sempre comete uma gafe ou divulga algum impropério quando surgem questionamentos ou informações negativas sobre o seu governo, aparentemente na tentativa de manipular indiretamente o noticiário do dia. Ver mais em: <https://noticias.uol.com.br/colunas/leonardo-sakamoto/2020/03/05/do-pib-pifio-a-golden-shower-7-cortinas-de-fumaca-do-governo-bolsonaro.htm>.

¹⁷ Nos EUA, por exemplo, cidades que votaram mais em Trump nas eleições de 2020 têm percentuais bem menores de vacinação que cidades que foram pró-Biden. Ver mais em: <https://acasignups.net/21/08/11/us-covid19-vaccination-levels-county-over-time>. Além disso, os estados com menores taxas de vacinação nos EUA possuem governadores republicanos. Ver mais em: <https://www.uol.com.br/vivabem/noticias/bbc/2021/07/28/pandemia-entre-nao-vacinados-como-eua-podem-estar-perdendo-de-novo-controle-sobre-o-coronavirus.htm>.

¹⁸ Pesquisa de julho de 2020 do DataPoder360 demonstrou que os apoiadores do presidente afirmam que utilizariam os medicamentos indicados por Jair Bolsonaro caso contraíssem a doença. Ver mais em: <https://www.poder360.com.br/poderdata/43-tratariam-covid-19-com-cloroquina-entre-bolsonaristas-73-tomariam/>

¹⁹ 16 governadores chegaram a enviar uma carta aberta ao presidente da república repudiando tais ataques: Ver mais em: https://www.em.com.br/app/noticia/politica/2021/03/29/interna_politica,1251834/governadores-fazem-manifesto-contra-agressoes-e-fake-news-zema-nao-assina.shtml.

²⁰ Ver mais sobre a capacidade protetiva das máscaras contra a COVID-19 em: <https://www.correiobraziliense.com.br/ciencia-e-saude/2021/06/4928711-protecao-de-mascaras-contra-a-covid-19-e-superior-a-90.html>.

²¹ O presidente citou uma pesquisa feita no Facebook como parâmetro para a não utilização de máscaras em crianças em uma de suas *lives*. Ver mais em: <https://g1.globo.com/bemestar/coronavirus/noticia/2021/02/26/bolsonaro-usa-enquete-alema-distorcida-para-criticar-uso-de-mascaras.ghtml>.

agressões verbais contra o então governador de São Paulo, João Doria²² e mentiu sobre a vacina adquirida pelo gestor estadual, a CoronaVac²³, o que pode ter gerado uma diminuição na intenção de se vacinar pela população²⁴.

A desinformação, portanto, tornou-se parte do discurso oficial de campanhas eleitorais e governos estabelecidos no mundo contemporâneo. O Brasil, em especial, passa por um intenso debate sobre quais os limites da liberdade de expressão da elite política, principalmente no que se refere ao poder de difamar instituições, adversários e mesmo acadêmicos em prol de um projeto/narrativa de poder²⁵.

As consequências de tal estratégia comunicacional para a confiança nas instituições que perfazem o modelo democrático liberal contemporâneo ainda são nebulosas. A interferência das cascatas de desinformação na opinião pública e no comportamento do eleitorado parecem não ser tão fortes para mudar significativamente eleições (FLYNN *et al.*, 2017; NYHAN, 2020), entretanto, grandes grupos extremamente engajados – como vistos durante a pandemia de COVID-19 – podem vir a colocar em risco a democracia, principalmente se tiverem apoio de partidos e instituições, uma vez que exemplos não faltam²⁶.

Sabendo disso, este trabalho busca verificar a existência de um modelo específico de comunicação que esteja presente na desinformação. Parte-se do pressuposto de que o formato como essas elites políticas interagem com sua militância pode esclarecer questões ligadas ao processamento de informações por parte dos eleitores. Além disso, seguimos os fortes indícios de que os produtores desse material buscam ativar afetos e partidarismos desses grupos através de uma linguagem sensacionalista.

²² O presidente chegou a chamar o governador do estado de São Paulo de ‘moleque’ e outras ofensas. Ver mais em: <https://www.folhape.com.br/politica/em-entrevista-bolsonaro-ataca-stf-maia-e-xinga-doria-de-moleque-e/169156/>.

²³ Por mais de uma vez o presidente atacou a compra do imunizante, se utilizando algumas vezes de informações falsas para tanto. Ver mais em: <https://saude.ig.com.br/2021-01-18/relembre-10-vez-que-bolsonaro-atacou-a-coronavac.html>.

²⁴ Entre agosto e dezembro de 2020, período em que o presidente duvidou da eficácia das vacinas compradas pelo governo de São Paulo (<https://oglobo.globo.com/brasil/veja-10-vezes-em-que-bolsonaro-criticou-coronavac-24843568>), o apoio à vacinação contra a Covid-19 caiu em 16% segundo o Datafolha (<https://g1.globo.com/jornal-nacional/noticia/2021/07/13/adesao-dos-brasileiros-a-vacinacao-contra-covid-atinge-nivel-recorde-segundo-o-datafolha.ghtml>).

²⁵ O presidente chegou a acusar o então diretor do INPE, Ricardo Galvão, de “trabalhar para uma ONG” após este indicar o aumento anormal de incêndios na floresta amazônica. Ver mais em: https://www.terra.com.br/noticias/ciencia/sustentabilidade/demitido-por-bolsonaro-ex-diretor-do-inpe-recebe-premio_3199c18662345d453d71d6e29b404c3drcwjve4v.html.

²⁶ A invasão ao Capitólio realizada por apoiadores de Donald Trump, após a sua derrota nas eleições de 2020, foi motivada por seitas conspiratórias como a *Qanon*. Ver mais em: <https://brasil.elpais.com/internacional/2021-02-13/historia-da-conspiracao-que-levou-ao-assalto-do-capitolio-assim-funciona-o-qanon.html>.

A descoberta de particularidades da estrutura lexical das notícias falsas pode vir a permitir comparações entre esse formato e outros tipos de discursos políticos e midiáticos. A partir da observação detalhada será possível apontar métodos particulares de convencimento e arregimentação de setores engajados, e a partir disso abrir um leque de possibilidades para a agenda de pesquisa futura.

Sendo assim, a observação e análise da estrutura das notícias falsas se faz cada vez mais necessária em um ambiente político que se utiliza – da direita à esquerda – desse expediente. Descrever um modelo de transmissão de ideias e informações pode vir a esclarecer questões profundas sobre como nós nos relacionamos com a linguagem e como a temática política nos afeta diariamente.

Entretanto, antes de tratar do formato em que a desinformação é apresentada, é necessário entender como funcionam a crença, o consumo e o compartilhamento desse tipo de conteúdo, ou seja, o que é fundamental para que as pessoas deem crédito e compartilhem material desinformativo. Para tanto, serão apresentadas as principais linhas teóricas que abordam a forma como nosso cérebro processa informações, e as possíveis explicações sobre o modo como interagimos com as notícias.

3 DEMANDA E OFERTA: COMO O CONSUMO DE DESINFORMAÇÃO SE RELACIONA COM A SUA ESTRUTURA TEXTUAL?

O ambiente político e comunicacional atual não esclarece por si só porque nós, como cidadãos, consumimos e compartilhamos desinformação. O que, antes de mais nada, explica esse comportamento? Partidarismo? Desconfiança na política e nas instituições? Analfabetismo informacional? Entender melhor o porquê deste consumo e como ele constrói um modelo de desinformação capaz de “viralizar” nas redes sociais é essencial para qualquer pesquisa que busque compreender a relação entre os fatores psicológicos das notícias falsas e a sua estrutura lexical. Este modelo de estudo possibilita um aprofundamento teórico sobre as variáveis que influenciam no êxito de uma desinformação política e os seus possíveis efeitos, modificando o foco da academia sobre esse fenômeno político comunicacional.

Este formato de observação científica se torna ainda mais importante quando tomamos conhecimento de que a análise cuidadosa de métricas de “viralização” de postagens e temas debatidos na internet se tornou um expediente fundamental para a assessoria de movimentos populistas (EMPOLI, 2020). Há um verdadeiro mercado ligado às notícias falsas que demanda certo tipo de narrativa e abordagem, mas que, para serem efetivas, tanto no que se refere à obtenção do clique, quanto no convencimento do eleitor, precisam se enquadrar em um modelo específico. Este trabalho, portanto, parte do pressuposto de que existe uma compreensão por parte dos redatores de desinformação (principalmente daqueles que residem em mídias hiperpartidárias) de que a forma com a qual a notícia é apresentada textualmente é importante para o consumo e compartilhamento da informação e, consequentemente, para a sua “viralização” nas redes.

A seguir serão expostas algumas das respostas possíveis que a literatura desenvolveu nos últimos anos para o fenômeno da desinformação. Primeiro, serão apresentadas as particularidades sociológicas e antropológicas das redes sociais e como elas influenciam diretamente na forma como nos comunicamos e interagimos com outras pessoas *on-line*, depois o trabalho irá revisar as duas principais explicações do porquê nós consumimos e compartilhamos desinformações: o raciocínio motivado e o processamento heurístico, correlacionando esses dois fatores com a estrutura textual das notícias falsas. Por fim, serão revisados os estudos que analisam os processos psicológicos existentes por trás da criação de mentiras e a literatura recente que estuda a estrutura das desinformações *on-line*.

3.1 AS REDES SOCIAIS, NOSSAS IDENTIDADES E EMOÇÕES

Um dos fatores estruturais importantes para o estudo do comportamento político das pessoas ao consumir uma informação falsa é o ambiente em que ela se encontra, ou seja, o local onde o indivíduo acessou a desinformação. O meio de comunicação não se reduz a um suporte para a mensagem, mas interfere diretamente no seu conteúdo (MARTINO, 2015, p. 199). Talvez por isso, desde o surgimento das mídias sociais, muito se tem falado sobre o papel dos seus algoritmos na criação de bolhas comunicacionais (RECUERO; GRUZD, 2019; RECUERO; SOARES; ZAGO, 2021).

Esse debate sobre como essa coleta de dados privados, utilizada pelas redes sociais *online* para a distribuição de informações e recomendações de círculos de amizades, são capazes de influenciar a vida das pessoas já era algo bastante presente na esfera pública. Trabalhos como os de Pariser (2012) e O’Neil (2016) levantaram questões sobre como a geração de dados sobre os nossos gostos e identidades podem produzir efeitos não imaginados em torno da opinião pública, segurança, privacidade e liberdade de pensamento. Esses questionamentos, entretanto, se acentuaram bastante após o pleito eleitoral estadunidense de 2016. Muito se falou sobre a forma como as informações foram distribuídas para os usuários de mídias sociais, questionamentos sobre o sucesso de campanhas difamatórias foram apontados por candidaturas derrotadas²⁷. O *Twitter*, *Facebook* e *Youtube* se tornaram alvos de investigações e denúncias²⁸. A situação só piorou quando escândalos envolvendo a venda de informações pessoais para agências de *marketing* político vieram à tona, colocando os CEO’s de grandes empresas na posição de investigados²⁹.

É verdade que, de forma recorrente, muitas destas mídias foram (e são) coniventes com agentes de desinformação³⁰. Isso ocorre, provavelmente, devido ao sucesso de

²⁷ Muitos críticos apontam que redes sociais como o Facebook tem ciência dos núcleos que espalham todo o tipo de desinformação em suas páginas, mas é leniente pois esse tipo de conteúdo gera grande engajamento. Ver mais em: <https://exame.com/revista-exame/o-maior-eleitor-de-trump/>.

²⁸ Acusações recentes vêm apontando que o próprio Facebook tinha ciência dos problemas ligados aos algoritmos utilizados pela plataforma e o incentivo a discursos extremistas. Ver mais em: <https://www.uol.com.br/tilt/noticias/redacao/2021/10/26/documentos-poder-algoritmos-facebook.htm>.

²⁹ Um deles foi Mark Zuckerberg, dono e criador do Facebook, que foi inquirido pelo senado. Ver mais em: <https://g1.globo.com/economia/tecnologia/noticia/mark-zuckerberg-depoe-ao-senado-sobre-uso-de-dados-pelo-facebook.ghtml>.

³⁰ Apesar de existirem alguns esforços localizados, como a criação de novas ferramentas de detecção de desinformação e a exclusão de certos perfis, ainda há certa leniência das redes sociais diante de desinformações publicadas pela elite política. Por exemplo, o presidente Bolsonaro teve, pela primeira vez, uma postagem sua ocultada apenas em maio de 2020, diante do seu negacionismo para com a pandemia de COVID-19. Ver em: <https://g1.globo.com/economia/tecnologia/noticia/2020/05/12/bolsonaro-compartilha-conteudo-falso-e-instagram-oculta-postagem-do-presidente.ghtml>.

compartilhamento das notícias falsas nesses ambientes (VOUSOUGHI *et al.*, 2018) e também pela pressão de setores políticos beneficiados por essa leniência. A suposição de que os custos políticos e econômicos do combate a estes atores talvez não valham a pena para os grandes conglomerados de mídias virtuais parece válida diante de descobertas recentes³¹.

Entretanto, estudos vêm apontando para um problema diferente do que foi sugerido inicialmente (NYHAN, 2020; BRADY, 2020). A questão envolvendo o consumo de informação política nas redes sociais não parece ser ligada unicamente a algoritmos maliciosos ou a presença de robôs que compartilham um material político específico, mas também às ativações psicológicas que este tipo de ambiente é capaz de gerar nos internautas. Isso se tornou cada vez mais claro com a ascensão do compartilhamento de desinformação em massa via *Whatsapp* (RESENDE *et al.*, 2018, 2019; BURSZTYN; BIRNBAUM, 2019), uma mídia social que não possui algoritmos capazes de influenciar na distribuição de conteúdo.

As redes sociais formadas em ambientes virtuais, de uma forma geral, são ambientes que incentivam certos comportamentos nos usuários, já que elas promovem a reunião de pessoas que possuem pensamentos, gostos, interesses e até mesmo ideologias parecidas. Essas reuniões geralmente são centradas em volta de um tema ligado a uma certa identidade (cultural, étnica, de gênero ou mesmo política) que é reafirmada pelos membros do grupo através de uma autoexposição contínua entre os usuários, seja através de postagens, compartilhamentos ou de apoios explícitos (curtidas, likes etc.). O modelo de grupos das redes sociais incentiva – de forma recorrente – a postagem enaltecedora da causa pela qual o grupo se une, e pune críticas ou questionamentos feitos a essa mesma causa, ou a como ela se organiza.

Isso acontece porque os seres humanos, mesmo em ambientes não-virtuais, são motivados a trocar informações de uma forma que satisfaça as suas necessidades de pertencimento a uma coletividade, e essa demanda por aceitação ativa tendências apoio às características que são o motivo central de existência desses agrupamentos (BRADY *et al.*, 2020). Aliado a esse comportamento está o processo de detração a opositores, que funciona como uma forma de reiterar a identidade e a superioridade dos valores do grupo (VAN BAVEL; PEREIRA, 2018).

As mídias sociais, ainda por cima, geram relações menos personalizadas, mais flexíveis e dinâmicas, diferente das redes sociais que nós possuímos (MARTINO, 2015, p. 56).

³¹ Denúncias feitas pela ex-funcionária do Facebook Francis Haugen demonstram que a empresa tem ciência de que seus algoritmos podem estar incentivando o consumo de desinformação. Ver mais em: <https://theintercept.com/2021/10/08/4-grandes-mentiras-que-o-facebook-contou-para-voce/>.

O ambiente *on-line*, muitas vezes, conecta pessoas que não se conhecem no mundo real, cuja ligação é centrada em torno de um determinado assunto ou comportamento. Essas conexões mais fracas acabam por incentivar a reiteração da identidade que une os participantes de um determinado grupo, já que não há outros fatores que condicionem a aproximação desses indivíduos (BRADY *et al.*, 2020). As interações sociais que criamos com amigos e família, a proximidade das casas, o dia a dia no trabalho, na escola, ou mesmo a torcida por um time local são – muitas vezes ao mesmo tempo – a causa da manutenção de laços afetivos. Nesse ambiente se torna mais difícil se desligar de um certo círculo de pessoas por discordâncias ideológicas ou partidárias, algo bem diferente do que é visto nos grupos *on-line*.

Estudos vêm demonstrando, inclusive, que o medo do ostracismo, de ser ignorado por outros membros de um determinado grupo, parece ser um fator importante no endosso a comportamentos nas mídias sociais. Esse fenômeno pode ocorrer mesmo em ambientes onde as pessoas mal se conhecem devido ao temor da possibilidade de uma rejeição inicial (GARRET *et al.*, 2020).

Essas descobertas indicam que, apesar das ligações mais frágeis, a reiteração da identidade e o medo da exclusão social podem provocar uma maior desindividualização dos membros da comunidade, reduzindo a capacidade de autoavaliação de cada um (BRADY *et al.* 2020). Isso faz com que os participantes desindividualizados percam as amarras referentes ao respeito a grupos adversários e particularizem ataques realizados à sua comunidade. Este efeito pode ser um incentivo para o compartilhamento de desinformações vistas como absurdas³² por pessoas que não fazem parte de um determinado grupo. Diversos trabalhos já apontavam como o partidarismo poderia interferir na forma como vemos o mundo, julgamos políticas públicas e notícias (VAN BAVEL; PEREIRA, 2018; DRUCKMAN *et al.*, 2013; BOLSEN *et al.*, 2014), entretanto, as mídias sociais possuem o potencial de estimular ainda mais essas características, independente da presença de algoritmos ou não.

É neste ponto que surge um aprofundamento do problema. O partidarismo e a polarização política vem sendo alimentado através da discussão de diversas pautas morais e identitárias nos últimos anos. Temas divisórios envolvendo questões de liberdade sexual, educação de menores de idade, liberação do uso de armas de fogo e fraudes eleitorais vem sendo, cada vez mais, segmentados entre ideologias políticas opostas (FUKS; MARQUES,

³² Alguns boatos espalhados nas eleições brasileiras de 2018 ficaram conhecidos por remeter a situações absurdas, como a distribuição de mamadeiras eróticas pelo candidato Fernando Haddad. Ver mais em: <https://rollingstone.uol.com.br/noticia/4-fake-news-completamente-absurdas-das-eleicoes-de-2018-mamadeira-erotica-jesus-e-travesti-e-mais/>.

2020). É verdade que essas pautas carregam consigo muito potencial de engajamento, tanto no mundo real como no virtual, entretanto, nas redes sociais elas têm um espaço de destaque (DOURADO, 2020) e isso ocorre principalmente porque tais debates carregam consigo uma grande carga emocional e moral (BRADY *et al.*, 2017).

Tendo em vista esse cenário, diversas análises vêm ressaltando o papel das emoções no consumo e compartilhamento de matérias e postagens na internet. Berger e Milkman (2012) já haviam demonstrado como os artigos do *New York Times* com maior carga emocional “viralizaram” mais do que outros. Ainda nesse estudo, os autores destacaram as emoções positivas e algumas negativas (como a raiva e a rejeição) como fatores importantes para popularidade de uma matéria de jornal.

Outras descobertas, contudo, indicam o papel não apenas da carga emotiva, mas também de termos ligados à moral, no sucesso de compartilhamento de postagens nas redes sociais (BRADY *et al.*, 2017). Geralmente, esse tipo de conteúdo mais normativo e emocionalmente carregado acaba servindo como um demarcador dos limites de atuação de um grupo e do status de pertencimento de um indivíduo dentro da rede social (BRADY *et al.*, 2020). Além disso, o simples fato de valorizarmos e confiarmos em nossas emoções pode estar ligado ao maior compartilhamento de desinformação nas redes sociais. Os experimentos realizados por MARTEL (*et al.*, 2020) demonstram que, quando somos incentivados a confiar em nossas emoções, ficamos mais suscetíveis a acreditar em notícias falsas.

Sabendo disso, este trabalho parte do pressuposto de que tais características textuais devem estar mais presentes em desinformações, já que estas também são mais “viralizáveis” que as notícias tradicionais (VOUSOUGHI *et al.*, 2018; ZHAO e ZHAO, 2020). Portanto, espera-se que esse conteúdo trate mais de pautas consideradas sensíveis por grupos politicamente engajados que se fazem presentes nas mídias sociais, como direitos sexuais, políticas de gênero, educação de crianças e adolescentes, fraudes eleitorais e ameaças bélicas/militares:

H1: Desinformações possuem mais termos ligados a temas politicamente delicados e divisórios.

Também serão testados, observando-se os limites metodológicos deste trabalho, se as descobertas de BRADY (*et al.*, 2017) são reproduzidas nas postagens desinformativas brasileiras. Espera-se que as notícias falsas contenham mais adjetivações morais, ou seja, termos utilizados no enaltecimento e acusação de figuras e instituições (bom, mau, honesto,

corrupto, bandido, falida). Este trabalho acredita que essa análise seja capaz de operacionalizar as premissas do trabalho citado, já que nele a variável *moral* é caracterizada pelo poder de contágio social ligado a avaliações de normas sociais e questões sobre a justiça (BRADY *et al.*, 2017). Por esse motivo, tendo em vista o caráter metodológico desta pesquisa, espera-se que as:

H2: Desinformações possuem mais adjetivações morais em seu texto (bom, mau, justo, incrível, honesto, corrupto e etc.).

3.2 A RACIONALIDADE MOTIVADA E A COGNIÇÃO QUENTE

Desde antes de 2016 a literatura especializada em comportamento político e opinião pública já tratava o partidarismo, ou a identidade partidária, como um dos fatores mais prováveis de explicação para a crença em desinformação. Uma das primeiras descobertas nessa área foi publicada em Kuklinski (*et al.*, 2000), na pesquisa que demonstrou que pessoas mal-informadas se comportavam de forma significativamente diferente das pessoas sem informações sobre um determinado tema político. Segundo o estudo, as pessoas mal-informadas³³, ou seja, aquelas que acreditavam em desinformações ou teorias da conspiração, possuíam uma crença mais forte do que as pessoas que tinham um conhecimento preciso sobre o mesmo assunto. Da mesma forma, esses indivíduos mal-informados dificilmente mudavam de opinião ou se corrigiam diante de uma informação corretiva, um comportamento significativamente diferente das pessoas que não possuíam muitas informações sobre o mesmo tema, que se atualizavam diante das novas referências (KUKLINSKI *et al.*, 2000).

Segundo os autores, essa atitude é justificada por certos processos mentais e cognitivos que possuímos. Estes processos se dão da seguinte forma: primeiramente, construímos nossas inferências sociais baseando-se em experiências pessoais, informações prévias, convivências e preceitos culturais. No entanto, essas concepções acabam por gerar um forte impulso em direção à consistência, ou seja, desejamos inconscientemente que nossas crenças e atitudes sejam coerentes entre si. Isso faz com que sejamos excessivamente confiantes nas nossas próprias opiniões e julgamentos (KUKLINSKI *et al.*, 2000).

Essa interpretação do processamento informacional é chamada de “Teoria do Raciocínio Motivado” (KUNDA, 1990). Ela ganhou muita força nas últimas duas décadas,

³³ Para garantir a coesão conceitual neste trabalho, adaptarei os conceitos utilizados por KUKLINSKI *et al.* (2000). No trabalho original as pessoas ‘mal-informadas’ eram aquelas que acreditavam em informações erradas, enquanto as pessoas desinformadas eram as que não possuíam muita informação sobre o tema.

sendo uma das principais linhas argumentativas de justificação do porquê tantas pessoas dão crédito às teorias da conspiração e desinformação. Essa hipótese parte do pressuposto de que, quando processamos informações, utilizamos basicamente de duas motivações principais para avaliar evidências e formar impressões (KUNDA, 1990). A primeira delas é a motivação – ou objetivo – de precisão, e, quanto a utilizamos, tendemos a buscar e considerar cuidadosamente as evidências relevantes, a fim de alcançar uma conclusão correta ou, pelo menos, mais precisa (TABER; LODGE, 2006). Geralmente nos inclinamos a privilegiar a precisão ao processarmos informações quando as conclusões advindas podem custar a nossa reputação, seja porque esses raciocínios serão avaliados por terceiros (como em uma prova e apresentação de um trabalho) ou quando as conclusões são importantes para pessoas próximas (projetos comunitários, planejamento de uma viagem).

A segunda motivação é a direcional. Diferentemente do objetivo de precisão, ela reflete o desejo de chegar a uma conclusão específica, geralmente priorizando um fim que seja consistente com nossas crenças prévias (JERIT; ZHAO, 2020). Psicologicamente, trata-se de um raciocínio bem parecido com o anterior, tanto que, normalmente, não notamos que estamos sendo direcionais (KUNDA, 1990). Contudo, a grande particularidade do processamento direcional ocorre no momento em que estamos coletando informações para chegar a uma conclusão. Quando somos direcionais escolhemos – inconscientemente – memórias ou dados que coadunam com as nossas crenças anteriores, isso pode ocorrer por vários motivos, como pressões advindas de um grupo ao qual o indivíduo se vê como pertencente, crenças culturais, ideologias políticas ou mesmo preconceitos muito arraigados (KUNDA, 1990).

Quando esse processamento direcional ocorre tendemos a raciocinar motivadamente, ou seja, passamos a utilizar diversos vieses (o de confirmação é um dos mais comuns) na escolha das informações que fundamentam a nossa reflexão sobre um determinado tema ou objeto. Esse comportamento acaba por diminuir consideravelmente a precisão das nossas conclusões, além de retroalimentar o sentimento de busca por informações consistentes com as respostas enviesadas que obtivemos.

O raciocínio motivado foi largamente estudado nos últimos anos. Pesquisas demonstraram que ele possui um papel muito importante nas crenças políticas errôneas das pessoas (FLYNN *et al.*, 2017), ativando um maior ceticismo para informações que sejam dissonantes com os pressupostos individuais (TABER; LODGE, 2006), sendo utilizado pelos indivíduos em situações onde a polarização política é mais saliente (DRUCKMAN *et al.*, 2013) ou quando um determinado tema gera tensões políticas (TABER; LODGE, 2006;

DRUCKMAN, 2012), além de ser ativado quando partidarismo político é forte (BOLSEN *et al.*, 2014). Essas descobertas tornaram o raciocínio motivado uma das principais explicações para a crença em desinformação (JERIT; ZHAO, 2020; BAPTISTA; GRANDIM, 2020).

Outros estudos, inclusive, ressaltam a importância do conteúdo consumido para a ativação do raciocínio motivado. A hipótese de cognição quente (LODGE; TABER, 2005; TABER; LODGE, 2006) indica que conceitos sociopolíticos³⁴ influenciam no processamento de informações a partir do raciocínio motivado. Ela parte do pressuposto de que esses termos são carregados de afeto e que os sentimentos atrelados aos conceitos são capazes de afetar a memória de longo prazo, interferindo no processo de julgamento de uma informação do início ao fim³⁵, desde a codificação da informação, sua recuperação e compreensão, à sua expressão como preferência ou escolha (LODGE; TABER, 2005).

Essa abordagem defende que esse modelo de processamento de informações, quando ativado de forma recorrente, tende a polarizar cada vez mais indivíduos que já possuem sentimentos relativamente fortes contra (ou a favor) de uma liderança ou movimento político, reativando o raciocínio motivado sempre que o modelo de apresentação das informações gire em torno desses conceitos. Essa crença é derivada dos resultados de experimentos que apontaram para a tendência de indivíduos com certa sofisticação política de sofrerem mais com o efeito da cognição quente (LODGE; TABER, 2005).

Portanto, sabendo do sucesso político e midiático que as notícias falsas têm e do poder de influência que elas exercem em setores engajados, este trabalho espera encontrar uma maior concentração de termos sociopolíticos entre as desinformações. A maior presença dessas palavras explicaria – em parte – como esse tipo de conteúdo ativa o raciocínio motivado. Sendo assim:

H3: Desinformações possuem mais conceitos sociopolíticos (nomes de lideranças, grupos, partidos e ideologias políticas).

³⁴ Os autores afirmam que toda liderança política, grupos, questões, símbolos e ideias pensadas e avaliadas no passado tornaram-se afetivamente carregadas – positiva ou negativamente – e isso afeta diretamente a memória de longo prazo. (LODGE e TABER, 2005).

³⁵ Outros autores chamam esse efeito no processamento de informações de “heurística do como eu me sinto” (*how-do-i-feel heuristic*) e “heurística da simpatia” (*likability heuristic*). (ver mais em LODGE e TABER, 2005).

3.3 A PREGUIÇA E A SIMPLICIDADE

Apesar do raciocínio motivado ter sido considerado nos últimos anos uma variável fundamental para a compreensão do consumo de desinformação *on-line*, novas pesquisas vêm apontando para outras características que podem ser fatores importantes no que diz respeito ao consumo e o compartilhamento desse conteúdo. Um desses comportamentos é a preguiça (ou desatenção).

O termo preguiça vem sendo utilizado por alguns autores para descrever uma maneira específica de processar e consumir informações (PENNYCOOK; RAND, 2018, 2021; PENNYCOOK *et al.*, 2021). Essa linha de abordagem segue a teoria psicológica do raciocínio dual, de acordo com a qual possuímos duas formas genéricas de processar informações: a de tipo 1, mais intuitiva e que vem à mente diretamente como uma resposta a um estímulo; e a de tipo 2, caracterizada pela deliberação e análise³⁶.

O processamento de tipo 1 por ser mais intuitivo e demandar menos esforço reflexivo é ligado à um comportamento preguiçoso pelos autores (PENNYCOOK; RAND, 2018), entretanto ele é a forma mais comum de raciocínio, pois, além de ser mais simples, sua realização também se dá de forma mais rápida e, na maioria das vezes, apresenta relativa eficiência. Porém, o sistema 1 também está ligado ao uso de heurísticas, atalhos mentais, para a tomada de decisões, e isto pode nos dizer muito sobre o consumo de desinformação (PENNYCOOK; RAND, 2021).

Esses pressupostos se relacionam com as descobertas recentes que demonstraram que boa parte dos indivíduos que compartilham notícias falsas suspeitam de sua imprecisão (PENNYCOOK *et al.*, 2021). Na verdade, a crença na veracidade de uma notícia está mais relacionada com a precisão do seu título do que com a concordância partidária do texto com os pressupostos políticos do leitor (PENNYCOOK *et al.*, 2021), um indicativo de que o principal fator para acreditarmos em uma notícia continua sendo sua precisão, mesmo em ambientes polarizados.

Essas revelações publicadas recentemente indicam que a teoria do raciocínio motivado não explica completamente o compartilhamento de desinformações nas redes sociais. Na

³⁶ Para exemplificar a diferença entre esses dois tipos de processamento, os autores apresentam o seguinte teste de reflexão cognitiva: “Um taco e uma bola custa \$1,10 no total. O taco custa \$1,00 a mais que a bola. Quanto custa a bola?”. A resposta intuitiva (processamento de tipo 1) e incorreta “10 centavos” vêm à mente intuitivamente para a maioria das pessoas, enquanto a resposta correta - 5 centavos - surge (para a maioria) apenas com um processo de deliberação adicional (PENNYCOOK; RAND, 2018, 2021).

verdade, segundo o estudo de PENNYCOOK *et al.* (2021), apenas cerca de $\frac{1}{3}$ das pessoas que estavam dispostas a compartilhar algum tipo de desinformação em sua *timeline* acreditavam na precisão delas. Todo esse comportamento ficou visível a partir da realização de diversos experimentos realizados por esses pesquisadores. Neles, as pessoas que consumiam e compartilhavam notícias falsas eram questionadas posteriormente ao ato de compartilhar sobre a precisão das informações que elas ajudaram a divulgar (PENNYCOOK *et al.*, 2021).

Esses mesmos experimentos apontaram que incentivos de precisão e atenção podem ser formas eficazes de coibir o compartilhamento de desinformação (PENNYCOOK *et al.*, 2021). Apoiar, mesmo que levemente, a reflexão analítica (sistema 2) pode coibir o compartilhamento de informações falsas. Esse efeito é um indicativo de que as pessoas falham em discernir a verdade da falsidade mais por causa da falta de reflexão e preguiça do que por um raciocínio politicamente motivado (PENNYCOOK e RAND, 2021).

As descobertas demonstram a importância da heurística e da economia reflexiva no compartilhamento de notícias falsas. A heurística, é importante lembrar, é uma forma de processamento de informações extremamente útil para qualquer ser humano, pois ela possibilita a análise de uma quantidade de dados muito maior do que o processamento analítico e em um período menor de tempo (MONDAK, 1993). Por isso, é de se esperar que as pessoas utilizem mais processamento heurístico quando elas têm de lidar com uma maior quantidade de dados e que busquem textos mais simples e menores em ambientes sobrecarregados de informações. Esta demanda gera uma cadeia de produção voltada para um modelo de mensagens curtas e “viralizáveis”, portanto, espera-se que os produtores informações falsas políticas criem conteúdo com a seguinte característica:

H4: Desinformações possuem, em média, menos palavras em seu corpo textual.

Além disso, partindo dos pressupostos anteriormente expostos, trabalhos recentes indicam que a maior legibilidade de um texto pode ativar o processamento heurístico em revisões de produtos em sites *on-line* (AGNHOTRI; BHATTACHARYA, 2016) e na utilização de códigos de programação (ALARCON *et al.*, 2017), espera-se, portanto, que o mesmo ocorra no consumo de notícias. Além disso, as *timelines* das redes sociais como o *Twitter* e *Facebook*, ou mesmo um grupo de *Whatsapp*, que são preenchidas de informações curtas e atualizadas a cada segundo, são verdadeiros modelos de ativação dessa forma de processamento de dados (PENNYCOOK *et al.*, 2021).

Sendo assim, fundamentando-se no pressuposto de que a legibilidade de um texto está relacionada ao modelo heurístico de processamento de informações, é possível supor uma estrutura textual de ativação do sistema 1. Portanto, espera-se que uma linguagem mais acessível³⁷ seja atrativa a pessoas que usam o modelo de processamento heurístico. A literatura relaciona a simplicidade textual com a atratividade e confiabilidade da informação (AGNHOTRI; BHATTACHARYA, 2016; ALARCON *et al.*, 2017). Índices de legibilidade são utilizados historicamente como uma forma de operacionalizar a simplicidade textual de um *corpus* e podem indicar estas características a partir da contagem das sílabas e letras em uma determinada sentença (COLEMAN; LIAU, 1975; FLESCH, 1979).

Tendo em vista que este trabalho já previu uma presença maior de termos emocionais e de temas sensíveis nas hipóteses 1 e 2, adicionaremos previsões acerca da complexidade textual da desinformação. Acredita-se que uma escrita mais simples facilitaria a priorização do raciocínio heurístico sobre uma determinada informação. Parte-se do pressuposto, novamente, que essa estrutura linguística promoveria a ativação do processamento do sistema 1, portanto:

H5: Desinformações possuem uma escrita mais legível que informações verdadeiras.

3.4 AS PARTICULARIDADES DO CONTEÚDO POLÍTICO

Este trabalho também levanta a possibilidade de interferência da temática política nas características das notícias falsas e verdadeiras. Essa hipótese é construída com base nos conteúdos abordados pelo noticiário político, que geralmente já é preenchido de termos sociopolíticos, o que por si só é um fator ativador de raciocínio motivado (LOGGE; TABER, 2005).

Além disso, acredita-se que parte das desinformações compartilhadas em redes sociais advém de centros de distribuição específicos, mesmo em mídias sociais estruturalmente pulverizadas como o *Whatsapp* (CHAGAS *et al.*, 2019; PAVARIN, 2019). Trabalhos também vêm apontando para o papel das mídias hiperpartidárias (RECUERO *et al.*, 2021), perfis de influenciadores (RECUERO; GRUZD, 2019) e até mesmo números de telefone possivelmente ligado a setores da elite política na distribuição de informações falsas (PAVARIN, 2019).

Essas descobertas podem indicar uma certa especialização por parte dos produtores desse tipo de desinformação. Sabendo disso, este trabalho espera encontrar características

³⁷ Veremos mais adiante de forma detalhada (no capítulo que discorre sobre a metodologia do trabalho) como os Índices de Legibilidade contabilizam a complexidade de um texto em específico.

ressaltadas nas notícias falsas que tratem de assuntos ligados à política nacional e internacional³⁸.

H6: Desinformações com temática política possuem características gramaticais e de legibilidade mais ressaltadas que desinformações não-políticas.

3.5 DISCERNINDO MENTIRAS E VERDADES

Pesquisas e teorias que interligam o uso de certas palavras com as atitudes e comportamentos individuais existem há bastante tempo, sendo encontrados nas mais diversas áreas do conhecimento. Trabalhos advindos da psicologia (UNDEUTSCH, 1967; VRIJ, 2008; PENNEBAKER *et al.*, 2003; TAUSCZIK; PENNEBAKER, 2010), análises promovidas por especialistas em comunicação (CARPENTER, 1981; KNAPP *et al.*, 1974) e saúde (WEINTRAUB, 1981, 1989) reúnem diversos modelos e pressupostos sobre como certas estratégias verbais (e textuais) carregam consigo significados que vão além da semântica tradicional. Essa linha teórica acredita que a organização sintática e gramatical das palavras pode expor dificuldades de raciocínio devido ao processamento de informações, rejeição a um certo objeto ou, até mesmo, estratégias de convencimento e “viralização” de um conteúdo midiático.

Estudos partem do pressuposto de que as pessoas se expressam de forma diferente quando mentem e dizem a verdade. As explicações do porquê dessa diferença variam bastante, Vrij (2008) reuniu algumas delas no trabalho *Detecting Lies and Deceit*. Segundo o autor – que revisou 69 estudos sobre o tema – as particularidades ligadas ao uso de palavras entre mentirosos podem estar ligadas ao: 1) Estado emocional do emissor, principalmente quando este sente medo ou culpa, que pode ser decorrente do ato de faltar com a verdade; 2) Esforço cognitivo, pois fabricar uma mentira demanda uma maior reflexão e trabalho criativo; 3) Tentativa de controle, partindo do pressuposto que ao mentir as pessoas tenderiam a omitir informações que não as comprometam; 4) Dissociação do alvo da mentira, pois as pessoas que mentem tendem a se afastar do objeto da mentira; e 5) Falta de convicção, já que as mentiras são baseadas em fatos inexistentes.

Mas como esses estados psicológicos nas falas das pessoas? Quais termos indicam o que os indivíduos estão sentindo, mas tentam esconder, ao mentir? Segundo Pennebaker (*et al.*, 2003), certos tipos de palavras operacionalizam melhor os sentimentos e desafios psicológicos

³⁸ Os critérios para a seleção do noticiário político estão elencados na seção 3.1, referente ao banco de dados usados nesta pesquisa.

que o processo de mentir infringe em um indivíduo. Esses termos são categorias importantes, já que sua utilização está correlacionada ao modo como nos relacionamos socialmente e pensamos, sendo, inclusive, processados em regiões cerebrais específicas. Os autores as chamam de *partículas*.

As partículas são, na verdade, uma categoria de palavras que incluem pronomes, artigos, preposições, conjunções e termos auxiliares, que tem por função conectar as *palavras de conteúdo* (substantivos, verbos e adjetivos). O uso de partículas pode oferecer indicativos interessantes ao pesquisador, apontando dificuldades no processo criativo da mentira ou, até mesmo, do sentimento de rejeição relativo ao objeto que está sendo caluniado. As partículas se apresentam, portanto, como uma variável que deve ser observada com atenção, pois elas representam mais da metade das palavras faladas por um indivíduo, mesmo existindo apenas cerca de 200 delas na língua inglesa (PENNEBAKER *et al.*, 2003).

Experimentos iniciais que seguiam essa linha de análise apontaram que pessoas que são induzidas a mentir sobre seus pensamentos e comportamentos tendem a usar menos pronomes pessoais no singular do que aqueles que falam a verdade (NEWMAN *et al.*, 2002), elas também tendem a utilizar mais termos autorreferenciais quando estão sendo honestas consigo mesmas. Segundo esses estudos, comunicações enganosas apresentam um menor número de pronomes de primeira pessoa do singular (PENNEBAKER *et al.*, 2003).

Por fim, outros experimentos apontaram que as mentiras contêm um número menor de palavras excluientes (*exclusive words*) – preposições e conjunções como, “mas”, “exceto”, “sem”, “entretanto”, “no entanto” e “excluir” – do que relatos verdadeiros (PENNEBAKER *et al.*, 2003; HANCOCK *et al.*, 2007; NEWMAN *et al.*, 2002), esses achados são indicadores de que o ato de mentir exige um esforço criativo por parte do emissor que acaba comprometendo a complexidade da narrativa.

Tendo em vista essas diferenças, pesquisadores passaram a se utilizar do método e das premissas já desenvolvidas pelos estudos anteriores para avançar na análise das estruturas léxicas da mentira, observando também o texto escrito. Os trabalhos de Gurrea (2015) e Mihalcea e Strapparava (2009), ainda bastante exploratórios, tendem a confirmar os resultados encontrados nos trabalhos que analisaram a palavra falada. Ambos os estudos indicam uma maior utilização de pronomes pessoais de primeira pessoa nos textos verdadeiros e pronomes de terceira pessoa nos textos falsos. Gurrea (2015), que analisou um *corpus* em espanhol, ainda encontrou um maior detalhamento de locais e horários nos textos verdadeiros, resultado este

que condiz com outros estudos realizados em sites *on-line* (OTT *et al.*, 2011), como veremos a seguir.

Tabela 1 – Presença de variáveis gramaticais significativas nos trabalhos revisados sobre o ato de mentir

Características	Mentiras	Verdades
Pronomes de 1 ^a pessoa	–	+
Pronomes de 3 ^a pessoa	+	–
Termos excluientes	–	+
Termos emocionalmente negativos	+	–
Autorreferências	–	+
Locais (precisão)	–	+
Tempo (precisão)	–	+

Fonte: produção do autor, 2022.

3.6 DISCERNINDO NOTÍCIAS FALSAS E VERDADEIRAS

Diante da ascensão da utilização da internet como forma de comunicação, novos trabalhos passaram a ter como objeto a interação textual entre os usuários da rede. Pesquisas começaram a comparar textos publicados em páginas da internet, como as avaliações de prestações de serviços em sites de hotelaria (OTT *et al.*, 2011). Essa análise, em específico, encontrou uma maior presença de substantivos, adjetivos, preposições, conjunções determinantes e de coordenação em comentários verdadeiros, enquanto comentários falsos possuíam mais verbos, advérbios, pronomes e predeterminantes.

Após os pleitos eleitorais norte-americano e britânico de 2016, diversos pesquisadores passaram a analisar o conteúdo das notícias falsas compartilhadas, principalmente, em redes sociais. A maioria desses trabalhos, diferentemente dos realizados no início do século, buscava encontrar variáveis léxicas capazes de distinguir o conteúdo falso e verdadeiro e, a partir daí, construir modelos computacionais capazes de reconhecer uma notícia falsa pelo texto (RASHKIN *et al.*, 2017; CHOY; CHONG, 2018; MONTEIRO *et al.*, 2019; HORNE; ADALI, 2017).

Esses trabalhos encontraram distinções significativas no conteúdo das notícias falsas e verdadeiras. Foi observado que as notícias falsas possuíam títulos maiores e mais explicativos que as verdadeiras (HORNE; ADALI, 2017), e que o corpo da desinformação era

significativamente menor (HORNE; ADALI, 2017; BAPTISTA, 2020). Também foi possível verificar tendências ligadas à legibilidade dos textos que reportavam informações falsas, sendo esse material mais simples e acessível que as notícias verdadeiras (HORNE; ADALI, 2017). Outros trabalhos se aprofundaram na análise das particularidades léxicas das notícias falsas, encontrando a predominância de pronomes pessoais, verbos e advérbios. Enquanto as notícias verdadeiras tendiam a ter um tamanho maior e mais termos comparativos.

Tabela 2 – Variáveis gramaticais e lexicais encontradas nos trabalhos revisados

Trabalho	Maior presença em Notícias Verdadeiras	Maior Presença em Notícias Falsas
Horne e Adali (2017)	palavras, termos técnicos, pontuação, aspas, substantivos, termos analíticos, substantivos (título)	advérbios, redundância, pronomes pessoais, comprimento (título), letras maiúsculas (título), substantivos próprios (título)
Rashkin (<i>et al.</i> , 2017)	termos comparativos e números	pronomes de 1 ^a e 2 ^a pessoa, palavras no superlativo, subjetivas e advérbios modais
Baptista (2020)	palavras	simplicidade, informalidade, redundância, pronomes pessoais e demonstrativos, verbos
Choy e Chong (2018)		verbos, adjetivos, nomes, números, palavras fortes (positivas ou negativas), comprimento (título)
Pérez-Rosas (<i>et al.</i> , 2020)	pronomes de 2 ^a pessoa do singular, termos negativos, expressões de relatividade	advérbios, verbos, pontuação
Monteiro (<i>et al.</i> , 2019)	caracteres, frases, tamanho médio das frases, palavras ligadas à incerteza e a não imediatismo.	erros gramaticais, verbos

Fonte: produção do autor, 2022.

Uma das pontuações interessantes de parte dessa linha de análise está relacionada com o uso da heurística. Esta é a conclusão de alguns dos trabalhos que estudaram o conteúdo das desinformações compartilhadas nas redes sociais (HORNE; ADALI, 2017; BAPTISTA, 2020) e eles defendem que o processamento de informação pelo leitor pode ser menos analítico caso ele tome por base, por exemplo, apenas os títulos das notícias falsas.

Essas descobertas estão diretamente relacionadas com a teoria do raciocínio dual e o papel da heurística no compartilhamento de notícias falsas (PENNYCOOK, RAND, 2017; 2021), já que elas apontam para uma maior simplicidade na escrita de desinformações, o que

ativaria o sistema de tipo 1, mais intuitivo e rápido. Além disso, a literatura voltada para a análise de conteúdo das notícias falsas vem apontando para o uso de manchetes impressionantes, que contenham relatos exagerados de eventos e provoquem uma variedade de sentimentos – como a alegria, o entusiasmo, a raiva e a tristeza (BAPTISTA; GRADIM, 2020). Esse expediente está de acordo com as duas principais teorias que explicam a “viralização” e a crença em notícias falsas, seja por influência de um raciocínio motivado por vieses políticos, seja por falta de atenção. Novamente, termos emotivos têm uma importância central no consumo não apenas de desinformação, mas de qualquer conteúdo de mídia, principalmente na era das redes sociais (BRADY *et al.*, 2020; BERGER; MILKMAN, 2012).

Em resumo, a literatura especializada aponta uma diferença significativa entre a estrutura textual de notícias falsas e verdadeiras. Essa linha de descobertas indica o que Gomes e Dourado (2019) já suspeitavam: o consumo de desinformação na internet pouco tem a ver com uma estratégia de mimetização do modelo jornalístico tradicional pelas *fake news*. A relação entre demanda e oferta nas notícias falsas segue uma lógica diferente da do mercado jornalístico. Enquanto este exige um modelo de redação cuidadosa, revisão ortográfica e metodologias de investigação, a desinformação segue tendências da escrita sensacionalista, utilizando-se de elementos da publicidade. Trata-se, de fato, de dois mercados diferentes, com demandas distintas e ofertas especializadas (ALLCOTT; GENTZKOW, 2017).

Sendo assim, tendo em vista os novos critérios de observação apresentados nas últimas páginas, a reflexão sobre a estrutura da desinformação pode ser expandida para análises de cunho mais gramatical e menos conteudístico (temático). Essa observação possibilitará a visualização dos modelos de escrita tanto das desinformações como das notícias verdadeiras, verificando-se as particularidades gramaticais de cada grupo.

Como foi visto, trabalhos anteriores relatam diferenças significativas de tamanho, legibilidade e de presença de formas gramaticais entre as notícias falsas e verdadeiras. A literatura parece ser relativamente coesa no que se refere ao maior uso em desinformações de: 1) Pronomes pessoais, o que pode ser um indicativo de informalidade, ou uma tentativa de aproximar o leitor do autor da desinformação (BAPTISTA, 2020); 2) Adjetivos, o que seria coerente com julgamentos prévios e categorizações a opositores e aliados políticos, algo muito visto em mídias hiperpartidárias; e 3) Verbos, o que estaria relacionado ao fato da desinformação se preocupar menos com a contextualização da informação e mais com a mensagem em si (esse comportamento, inclusive, permite a utilização de notícias verdadeiras fora de contexto).

H7: Desinformações possuem mais pronomes pessoais, adjetivos e verbos que informações verdadeiras.

Também se espera a maior presença em informações verdadeiras de: 4 - Substantivos, que – em oposição a maior presença de verbos nas informações falsas – é um possível indicativo de contextualização e desenvolvimento da informação; e 5 - Numerais, que apontam um maior detalhamento da informação, essa variável também pode estar relacionada à citação de locais e horários no texto (OTT *et al.*, 2011; GURREA, 2015).

H8: Desinformações possuem menos substantivos e numerais que informações verdadeiras.

4 DADOS E MÉTODO

4.1 BANCOS DE DADOS

Tendo ciência das descobertas feitas pela literatura especializada, o próximo passo deste trabalho é encontrar meios para testar as hipóteses que foram levantadas aqui. Tentaremos, portanto, replicar parte desses estudos a partir de novas abordagens e técnicas diferentes das aplicadas nos trabalhos que foram revisados. Será utilizada a base de dados advinda do projeto Fake.Br (MONTEIRO *et al.*, 2019), que fornece mais de 7000 notícias falsas e verdadeiras, colhidas entre janeiro de 2016 e janeiro de 2018, disponibilizadas publicamente para replicação.

O banco Fake.Br delimitou suas notícias reais e falsas por correspondência temática, ou seja, para cada notícia falsa encontrada os pesquisadores buscaram outra verdadeira sobre o mesmo tema. Para conseguir esses resultados os autores utilizaram as notícias falsas como “guia” para o levantamento dos dados, pesquisando publicações verdadeiras – em mídias confiáveis – que possuíssem a mesma pauta da desinformação (MONTEIRO *et al.*, 2019).

Essa característica em particular do banco de dados possibilitará uma análise mais minuciosa sobre as particularidades envolvendo as abordagens temáticas em notícias falsas e verdadeiras. Mesmo tendo conteúdo correspondentes, espera-se que a forma como os autores de desinformações tratam o objeto da notícia seja significativamente diferente dos redatores de informações verdadeiras.

Importante dizer que os autores do trabalho original que levantaram os dados para o Fake.Br (MONTEIRO *et al.*, 2019), buscavam testar, através de estratégias de processamento de linguagem natural (PNL), modelos preditivos capazes de identificar notícias falsas automaticamente a partir de seu conteúdo³⁹. Os resultados encontrados, através da “tokenização” dos textos⁴⁰, apontavam certas particularidades que distinguiam as notícias falsas das verdadeiras como uma maior presença de frases, adjetivos e advérbios por segmento de texto nas informações precisas, bem como um menor número de verbos e erros gramaticais.

Ressalta-se ainda que o Fake.Br possui um corpo de notícias com os mais variados temas (política, celebridades, sociedade, ciência e tecnologia, economia e religião). Por esse

³⁹ No trabalho original, os pesquisadores testaram diversos métodos de PNL. O sucesso dos modelos preditivos variou entre 77% e 88% de acertos (MONTEIRO *et al.*, 2019).

⁴⁰ Para evitar vieses ligados ao tamanho do texto, o que prejudicaria um modelo preditivo, os autores analisaram as características de cada banco utilizando os segmentos textuais como parâmetro de observação (tokens), e não das notícias completas (MONTEIRO *et al.*, 2019).

motivo serão selecionadas para este trabalho 800 notícias, 400 falsas e 400 verdadeiras com temática política, e também 140 notícias, 70 falsas e 70 verdadeiras, com outros assuntos⁴¹. Utilizamos como critério de seleção para o banco com temática política a citação a alguma personalidade, instituição, partido ou evento que esteve no centro do debate político nacional e internacional nos últimos 6 anos⁴².

Essa distinção, inclusive, permite a verificação de alguma particularidade, tanto na abordagem temática, quanto lexicométrica, das notícias falsas com pauta política. Espera-se que, devido à especialização técnica dos seus produtores, as desinformações políticas terão suas características ressaltadas quando comparadas ao material não-político (H6). Por fim, importa dizer que os dados foram tratados e adaptados para a leitura do Iramuteq, de acordo com o que foi estabelecido no manual de Camargo e Justo (2016)⁴³.

4.2 METODOLOGIA E TÉCNICAS APLICADAS

4.2.1 Análise de Conteúdo e Análise de Texto Automatizada

A análise de conteúdo (AC) é um conjunto de técnicas de análise das comunicações que visa obter, através de procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam inferir conhecimentos relativos às condições de produção/recepção (variáveis inferidas) dessas mensagens (BARDIN, 1977). Trata-se de um modelo metodológico relativamente antigo, os primeiros relatos de aplicação de algo parecido com essa técnica são referentes aos estudos de hinos e textos religiosos feitos desde o século XVII (BARDIN, 1977), mas foi apenas na primeira metade do século XX, com o desenvolvimento de técnicas de estudo da imprensa e da propaganda durante a 1^a Guerra Mundial, que a AC passou a ser uma metodologia voltada para o estudo amplo de formas de comunicações (OLIVEIRA, 2019).

Durante boa parte do século passado, contudo, o estudo de textos e discursos ficou atrelado ao modelo quantitativo. Influenciados pelos trabalhos de BEREKMAN (1954) muitos pesquisadores trataram a AC como uma técnica de pesquisa para a descrição objetiva, sistemática e quantitativa do conteúdo manifesto na comunicação (RICHARDSON *et al.*, 2015, p. 221). Durante as décadas de 40 e 50 as análises de conteúdo tinham a objetividade como fim

⁴¹ Textos que se centravam no dia a dia de celebridades, esportes, comportamento, anedotas, meteorologia, música, religião, entre outros foram colocados neste grupo.

⁴² Os dados colhidos nesta pesquisa, em sua maioria, foram publicados no segundo semestre de 2017.

⁴³ Os detalhes e o passo a passo das edições realizadas pelo autor para o processamento dos dados estão descritos no Apêndice D.

primordial e essa estratégia recebia diversas críticas de pesquisadores sociais de outras áreas, principalmente por negar a importância do que não estava explicitamente exposto no conteúdo estudado. Para esses críticos, o contexto político e as características dos emissários deveriam ser levados em consideração (RICHARDSON *et al.*, 2015, p. 221).

Após uma abundante produção advinda de pesquisadores franceses essa metodologia passou e levar mais em conta o estudo do ambiente social em que o conteúdo foi produzido. A análise de conteúdo hoje é considerada por alguns pesquisadores como uma técnica central para todas as ciências humanas e vem se transformando em um instrumento importante para o estudo da interação entre os indivíduos (RICHARDSON *et al.*, 2015, p. 221).

O ponto de partida desse modelo de análise é a mensagem, seja ela verbal, gestual, silenciosa, figurativa, documental ou diretamente provocada (PUGLISI; FRANCO, 2005). A partir daí a metodologia estabelece o trabalho de organização da pré-análise do objeto estudado, onde o pesquisador decide qual será o objeto de sua análise e a pergunta de pesquisa. Esta fase é seguida pela exploração do material, momento em que geralmente se estabelece um *corpus* textual de forma delimitada, havendo depois o tratamento dos resultados e, por fim, a produção de inferências sobre dados obtidos (BARDIN, 1977).

Diferentemente da análise do discurso (AD), mais voltada para o contexto em que as mensagens estão inseridas, a AC foca o texto como centro de sua pesquisa. Essa prioridade referente ao objeto é fundamental para a construção de uma metodologia capaz de ser objetiva e replicável, algo que era visto como prioridade pelos desenvolvedores deste método (OLIVEIRA, 2019).

Entretanto, com o desenvolvimento de novas tecnologias de processamento de dados, muitos programadores passaram a utilizar os critérios metodológicos da análise de conteúdo para a criação de modelos capazes de auxiliar no estudo de *corpus* textuais. Esses modelos também se utilizam da lexicometria, desenvolvida na década de 70, no processamento das informações estudadas (OLIVEIRA, 2019). Estas novas técnicas, que ficaram conhecidas como análise de texto automatizada (ATA), permitem o processamento de uma quantidade maior de dados de forma rápida e segura.

No entanto, a análise de texto automatizada possui algumas características próprias, como, por exemplo, o pressuposto de que os sentidos das trocas comunicativas podem ser encontrados via análises matemáticas sobre as palavras, que se tornam estatística através da contagem de frequência e testes multivariados (OLIVEIRA, 2019). Isso é possível devido a

dois processos necessários para o tratamento dos dados textuais, o primeiro é o *steaming*, onde as palavras são reduzidas a sua menor forma completa, retirando suas terminações, o que permite a *lematização*, que nada mais é do que a união dos termos pelo seu radical. Esta abordagem possibilita a categorização das palavras pelo seu significado, e facilita a formulação de inferências por parte do pesquisador.

A ATA, contudo, também difere da AC no sentido de que suas categorizações são, em sua maioria, feitas depois da análise estatística de *softwares* e modelos de programação. Geralmente os pesquisadores se utilizam dos resultados obtidos após o tratamento dos dados para gerar inferências, não sem antes testar a validação do tratamento realizado (IZUMI; MOREIRA, 2018). Como bem lembram Izumi e Moreira (2018), transformar palavras em números não substitui a leitura atenta e cuidadosa de documentos (por isso a necessidade da validação), mas permite a análise sistemática de grandes bases de textos sem a necessidade de mão de obra em larga escala e de um enorme montante de recursos financeiros. Dessa forma, o trabalho e a interpretação do pesquisador ainda se faz presente tanto para estudar os números advindos das análises quantitativas, quanto para compreender os resultados obtidos em técnicas que auxiliam no estudo qualitativo do *corpus*.

Além disso, este trabalho também se utiliza de critérios lexicométricos de análise de texto. Os termos e as formas gramaticais encontradas nos textos são considerados variáveis de extrema importância para o estudo em si. A lexicometria é um procedimento metodológico e tecnológico que trata, a partir de análises estatísticas, dados qualitativos sob uma perspectiva quantitativa com a finalidade de produzir caracterizações topológicas e combinatórias de elementos lexicais de um determinado *corpus* (DAMASCENO, 2008).

Diante disso, foram escolhidos dois *softwares* que reúnem diversas técnicas de análise de texto automatizada, cada um voltado para testes específicos. O primeiro é o Iramuteq (RATINAUD, 2009), *software* que utiliza programação em R para aplicar técnicas de ATA e lexicométricas. O segundo é o ATL (legibilidade.com), que é capaz de aplicar diversos testes de legibilidade em um banco de dados textual.

4.2.2 Análise de conteúdo automatizada a partir do Iramuteq

O *software* Iramuteq (RATINAUD, 2009) foi utilizado na produção de parte das análises estatísticas e de conteúdo sobre o *corpus* textual do banco de dados escolhido neste trabalho. Este programa já foi aplicado em alguns estudos sobre desinformação e notícias falsas (CHAGAS; CARREIRO, 2021; DOURADO, 2021; BAPTISTA, 2020) servindo de ferramenta

para a exploração do conteúdo compartilhado em redes sociais como *Twitter*, *Facebook* e *Whatsapp*. Trata-se de um *software* gratuito, desenvolvido sob a lógica *open source*, licenciado por GNU NPL (v2), que utiliza o *software* R e a linguagem python (www.python.org) para processar os dados (CAMARGO; JUSTO, 2013). O Iramuteq também possibilita a replicação dos dados, tanto em R como no próprio *software*, já que todo o script original é salvo automaticamente na máquina que realiza o processamento dos dados.⁴⁴ Além disso, este *software* possui um dicionário em português que vem sendo revisado constantemente (CAMARGO; JUSTO, 2013, 2016).

Esse programa permite a realização de diversas técnicas de análise de texto automatizada, possibilitando a observação de características do conteúdo do *corpus*. A inicial, e mais básica, são a contagem de palavras e a segmentação automatizada do texto (geralmente em 40 toques). Além de serem fundamentais para a realização de todas as outras técnicas de ATA que o *software* disponibiliza, esses dois processamentos de dados possibilitam que o pesquisador crie um novo *corpus* textual a partir de segmentos de texto que contém uma palavra específica. Por exemplo, é possível selecionar todos os segmentos textuais de um material que citam o nome de uma liderança ou partido político e criar um *subcorpus* a partir deste.

Outra técnica disponibilizada pelo *software* é a análise de especificidades, que permite, para além da contagem dos termos e das formas gramaticais, a comparação destas variáveis entre diferentes *corpus* textuais a partir da distribuição hipergeométrica⁴⁵ das palavras no texto. Trata-se de uma análise de contrastes, na qual o *corpus* é dividido em função de uma variável escolhida pelo pesquisador (CAMARGO; JUSTO, 2013), neste trabalho o critério de divisão é justamente a falsidade do texto. Sendo assim, através da análise de especificidades é possível comparar estatisticamente a presença de termos e das formas gramaticais em dois *corpus* diferentes.

A análise de especificidades será utilizada de forma recorrente no trabalho para os testes que irão comparar a lexicometria dos *corpus*. Como ela é capaz de apontar tendências ligadas aos termos e formas gramaticais utilizadas em um texto, espera-se com ela testar as hipóteses 1, 2, 3, 6, 7 e 8. Para evitar possíveis vieses, ligados ao tamanho dos bancos de dados

44 Os dados para verificação e replicação estão disponíveis no GitHub no seguinte link: https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo.

45 Em teoria das probabilidades e estatística, a distribuição hipergeométrica é uma distribuição de probabilidade discreta que descreve a probabilidade de sucessos em retiradas, sem reposição, de uma população de tamanho que contém exatamente sucessos, sendo cada retirada um sucesso ou um fracasso. Ver mais em: https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_hipergeom%C3%A9trica.

comparados – como bem foi apontado pelos próprios criadores do Fake.Br (MONTEIRO *et al.*, 2019) –, também aplicamos a análise de especificidades sobre segmentos de textos. Para tanto, selecionamos 15 termos – ligados ao noticiário político – e criamos *subcorpus* para cada um deles. O intuito com isso é testar se os dados encontrados na análise geral se repetem nessas observações localizadas de forma coesa, testando assim a robustez dos resultados obtidos.

Outra forma de utilizar o Iramuteq, que será vista no decorrer do trabalho, é aplicando a Classificação Hierárquica Descendente (REINERT, 1990), também chamada de “Método Reinert” (em homenagem ao seu idealizador, Max Reinert). Esta abordagem categoriza todos os termos do texto, classificando-os a partir de sua proximidade via *qui-quadrado* (χ^2), observando a identificação de co-ocorrências das palavras nos mesmos segmentos textuais (CERVI, 2018), ou seja, quanto mais vezes os termos “Lula” e “PT” estiverem presentes, ao mesmo tempo, em segmentos textuais, mais significativamente próximos eles serão. A CHD também reúne os termos em classes categorizando-os hierarquicamente segundo a sua significância estatística (p-valor). O *software* então disponibiliza a visualização desta técnica através de dois modelos: um dendrograma, que prioriza a divisão entre as classes de palavras e a hierarquia de cada termo; e via AFC (Análise Fatorial por Correspondência), que apresenta a proximidade entre cada termo a partir de um plano cartográfico.

A partir desses resultados o pesquisador pode observar o conteúdo de uma grande quantidade de dados textuais, estabelecendo os temas centrais das linhas argumentativas presentes no *corpus* (CERVI, 2018; CAMARGO; JUSTO, 2013; 2018). Esta técnica viabiliza o teste das hipóteses que abordam o conteúdo das notícias falsas e verdadeiras (H1, H2 e H3), permitindo a comparação entre os temas abordados pelos bancos de dados e os termos centrais de cada classe, possibilitando assim a verificação de tendências sensacionalistas ou acusatórias no texto. Além disso, a CHD permite a validação temática da seleção do noticiário político para os testes da H6.

4.2.3 Análise da legibilidade

Este trabalho também utilizará da análise automatizada de conteúdo para classificar a legibilidade (ou apreensibilidade)⁴⁶ dos materiais desinformativos e das informações verdadeiras. Como vimos, parte da literatura afirma que a simplicidade linguística é uma das características das notícias falsas (HORNE; ADALI, 2017) e que esta característica está ligada ao processamento heurístico de informações (AGNHOTRI; BHATTACHARYA, 2016; ALARCON *et al.*, 2017), por esse motivo utilizaremos o *software* ALT (Análise de Legibilidade Textual) para analisar os níveis de legibilidade desses conteúdos.

O ALT é um sistema que reúne índices de legibilidade para textos da língua portuguesa, tomando por base fórmulas adaptadas do inglês. Ele possui um algoritmo que conta o número de letras, sílabas, palavras, frases e palavras complexas e, a partir de modelos matemáticos (Flesch-Kincaid, Gulpease, ARI, Coleman-Liau e Gunning-fog) calcula os índices de legibilidade dos textos (legibilidade.com). Alguns dessas fórmulas, inclusive, já foram utilizadas pela literatura na análise comparativa da estrutura linguística das notícias falsas e verdadeiras⁴⁷. Como esses modelos foram criados originalmente para outras línguas (Inglês, Francês e Italiano) serão utilizadas apenas as categorizações já empregadas em trabalhos publicados que testaram estes índices de legibilidade em língua portuguesa (CAVIQUE, 2008; LYRA; AMARAL, 2012).

Dois dos índices mais utilizados pela literatura para identificar a complexidade do texto escrito giram em torno do modelo Flesch-Kincaid (FK). O primeiro está relacionado à facilidade de leitura (FLESCH, 1979), sendo calculada a partir do número médio de sílabas por palavra e o comprimento médio das frases, o resultado da fórmula varia de 0, que indica uma baixa legibilidade (alta complexidade), a 100, que indica alta legibilidade (baixa complexidade) (LYRA; AMARAL, 2012). Na mesma linha segue o índice de anos de escolaridade Flesch-Kincaid, que converte o mesmo modelo em anos de escolaridade, adaptando o texto a uma faixa de anos de estudo. No *software* ALT esse índice vai de 0, alta legibilidade (baixa complexidade) a 20 baixa legibilidade (alta complexidade).

⁴⁶ O termo legibilidade está sendo utilizado neste trabalho como uma tradução de *readability*, que se refere à dificuldade de compreensão de um texto devido à estrutura das palavras e das frases, este conceito difere de *legibility*, que se refere às características tipográficas e de layout de um texto. Em seu trabalho LYRA e AMARAL (2012) se utilizam do conceito de apreensibilidade para traduzir *readability* e legibilidade para se referir a *legibility*.

⁴⁷ Em seu trabalho HORNE e ADALI (2017) utilizam os índices Gunning-fog e Flesch-Kincaid para calcular a legibilidade dos títulos e corpo das notícias falsas e verdadeiras.

Índice Flesch-Kincaid:

$$226 - 1,03 x (nº de palavras / nº de frases) - 72 x (nº de sílabas / nº de palavras)$$

Índice de anos de escolaridade Flesch-Kincaid:

$$0,36 x (nº de palavras / nº de frases) + 10,4 x (nº de sílabas / nº de palavras) - 18$$

Também utilizaremos como parâmetro o índice Coleman-Liau (COLEMAN; LIAU, 1975), que calcula o nível de ensino necessário para o consumo e um certo texto baseado nas medidas dos comprimentos das sentenças e a média do número de caracteres por palavras (LYRA; AMARAL, 2012).

Índice de Coleman-Liau:

$$5,4 x (nº de letras / nº de palavras) - 1,4 x (nº de frases / nº de palavras) - 13,7$$

Partindo-se, portanto, do pressuposto de que esses índices conseguem apontar a complexidade dos textos estudados, tem-se uma ferramenta de teste para as hipóteses 4 e 5. Se a literatura estiver correta, as notícias falsas e desinformações compartilhadas em redes sociais tendem a ter uma complexidade menor do que as informações verídicas e isso deve ser apontado nos testes acima relatados.

Para testar as hipóteses ligadas à legibilidade levantadas neste trabalho serão selecionados 200 textos do banco Fake.Br (Monteiro *et al.*, 2019), 200 contendo notícias falsas e 200 contendo notícias verdadeiras. Esses dois *corpus* (falso/verdadeiro) serão subdivididos em *subcorpus* menores, contendo cada um 10 textos. A partir dessa base de dados serão realizados os testes na plataforma ALT, observando os valores obtidos por cada conjunto de dados nas análises de legibilidade. Essa técnica diminui os efeitos de possíveis *outliers*, e é capaz de verificar a estabilidade, dentro do banco de dados, dos resultados obtidos. O passo a passo com o tratamento dos dados e as aplicações dos testes está detalhado no Apêndice H.

5 RESULTADOS

5.1 ANÁLISE DE ESPECIFICIDADES E PRIMEIROS TESTES

O banco de dados de textos com conteúdo político reuniu 800 notícias (400 falsas e 400 verdadeiras). A análise estatística do Iramuteq identificou 508.396 palavras (ocorrências), sendo 6.697 delas (1,32% dos termos) *hepaxes*, que são termos não processados pelo *software*. O baixo percentual de *hepaxes* indica uma eficácia considerável na identificação dos termos pelo Iramuteq. A média geral de palavras por texto foi 636,29. O banco de dados com notícias verdadeiras reuniu 441.536 palavras, a média de termos por texto foi de 1103,84. Já o de notícias falsas contabilizou 66.860 palavras, e uma média de 167,57 termos por texto.

Estes dados iniciais já apontam para uma diferença essencial entre os dois tipos de notícias: o tamanho, as notícias verdadeiras tinham, em média, quase 7 vezes mais palavras que as falsas. Esse resultado está de acordo com as descobertas feitas em outros trabalhos (HORNE; ADALI, 2017; BAPTISTA, 2020), análise similar à do estudo original sobre o banco *Fake.Br* (MONTEIRO *et al.*, 2019). Ele também confirma a hipótese 4, que previa o tamanho menor das desinformações.

A partir da análise de especificidades o Iramuteq também consegue comparar, através de lei hipergeométrica, a presença dos termos ativos no banco de notícias falsas e verdadeiras. Logo de início é possível observar como a pauta envolvendo a crise diplomática entre a Coreia do Norte e os EUA foi preponderante no banco referente às notícias falsas. Destaca-se o uso de palavras escritas incorretamente, como *Coreia*⁴⁸, *Jon-un*⁴⁹, os dois termos de maior concentração neste banco. Além disso as palavras “EUA”, “Kim”, “nuclear”, “ditador”, “Vinson”⁵⁰ e “míssil” ressaltam a preponderância deste tema entre as notícias falsas presentes no banco de dados. No que se refere à política nacional os termos “corrupto”, “petista”, “fribol”, “Dório(a)”⁵¹, “Globo” e “pf” foram, proporcionalmente, mais citados entre as notícias falsas.

Na base de dados com notícias verdadeiras foi possível verificar a preponderância de termos suplementares como as preposições “em”, “de”, “por” e “segundo”⁵² e os advérbios

⁴⁸ Até 2008 a palavra *Coreia* se escrevia com acento agudo, no entanto, com o advento da reforma ortográfica daquele ano o termo perdeu esse acento, assim como outros ditongos abertos.

⁴⁹ Referente ao nome Kim Jong-un, líder norte-coreano.

⁵⁰ Referente ao nome Carl Vinson, nome do porta-aviões de 97.000 toneladas que se deslocou para o mar da Coreia durante a crise diplomática.

⁵¹ Termo referente ao nome do, à época, prefeito de São Paulo, João Doria. Como pode se observar o sobrenome foi escrito de forma errada 20 vezes no banco de notícias falsas.

⁵² Acreditamos que a maior utilização desse termo deriva da aplicação do mesmo como preposição.

“como” e “então” quando comparados a base de dados de notícias falsas, que têm substantivos e adjetivos entre as palavras com maior concentração. Outros termos que são predominantes entre as notícias verdadeiras e merecem destaque são “processo”, “decisão” e “recurso”, palavras comumente ligadas a pautas jurídicas e a palavra *rio*, provavelmente referente ao estado do Rio de Janeiro, que passou por uma intervenção federal durante o período em que as notícias foram colhidas⁵³.

Tabela 3 – Termos com maior concentração entre as notícias falsas e verdadeiras⁵⁴

Notícias Falsas		Notícias Verdadeiras	
Termos	Presença	Termos	Presença
coréia	76,4953	em	20,8046
jon_un	31,8681	afirmar	16,7174
eua	28,0622	segundo	14,0301
kim	27,0662	de	12,1053
nuclear	21,6356	processo	8,8645
corrupto	21,0928	reforma	8,6348
petista	18,6898	como	8,6159
vídeo	17,1553	g1	7,6796
através	16,4328	decisão	7,6014
friboi	16,2027	rio	7,5701
ditador	16,0228	a	7,2779
jornalista	15,6725	por	7,0194
dório	15,5746	forma	6,8449
globo	15,1201	haver	6,5751
pf	13,9991	recurso	6,2069
estar	13,7124	qual	6,1956
um	12,9801	primeiro	6,037
ataque	12,6742	então	5,8884
míssil	12,4001	nós	5,7173
vinson	12,3871	entanto	5,7116

Fonte: produção do autor, 2022.

⁵³ Inicialmente a intervenção ocorreu via GLO (Operação de Garantia da Lei e da Ordem), ainda em 2017, posteriormente foi decretada uma intervenção militar no estado durante todo o ano de 2018.

⁵⁴ Os valores representam a probabilidade de encontrar a palavra em questão no banco de dados analisado a partir da comparação entre notícias falsas e verdadeiras.

A partir da observação mais detalhada da análise de especificidades foi possível testar a H1 (Desinformações dão mais centralidade a termos ligados a temas delicados e divisórios), a H2 (Desinformações possuem mais adjetivações morais) e a H3 (Desinformações possuem mais conceitos sociopolíticos).

No que se refere a presença de termos ligados a temas delicados e divisórios, não foi possível verificar uma diferença significativa entre as duas categorias. Palavras ligadas a debates sobre liberdade sexual (*gay* e *LGBT*), interrupção da gravidez (*aborto*) e raça (*racismo*) foram as únicas significativamente presentes no material estudado⁵⁵, entretanto nenhuma delas apresentou uma concentração consideravelmente maior em um dos bancos. A maioria desses termos, inclusive, foi mais presente entre as notícias verdadeiras (*LGBT*, *aborto* e *racismo*). De fato, essas pautas não foram o centro temático de nenhuma das duas categorias – como se verá adiante nos testes realizados com a CHD – o que indica que as notícias falsas desse período evitavam abordar esses temas em seu conteúdo⁵⁶.

Entretanto, a observação das adjetivações morais apresentou resultados mais significativos. Palavras como “corrupto”, “ditador”, “honesto”, “poderoso”, “golpista”, “ladrão”, “famoso”, “vagabundo”, “pobre”, “podre”, “privilegiado”, “terrível”, “sujo” e “criminoso” tiveram uma concentração maior entre as notícias falsas, enquanto os termos “íntegro”, “mal” e “importante” estiveram mais concentrados nas informações verdadeiras. Esse resultado indica o maior uso de adjetivações que carregam algum valor moral entre as notícias verdadeiras, com destaque para os termos “corrupto” e “honesto”, que apresentaram uma presença 21 e 10 vezes maiores nas desinformações. Esse tipo de conteúdo, importa lembrar, foi apontado como uma variável significativamente relacionada ao sucesso de uma postagem nas redes sociais (BRADY *et al.*, 2017).

Por fim, foram encontrados indícios de respaldo da H3. Esperava-se que as notícias falsas possuíssem mais conceitos sociopolíticos, palavras ligadas a lideranças, partidos e ideologias políticas, capazes de carregar afeto e incentivar a ativação de raciocínio motivado. De fato, nomes de atores políticos nacionais de relevância (“Sérgio Moro”, “Gilmar Mendes”, “Lula”, “Dilma”, “Aécio Neves”, “Michel Temer”, “Serra”, “Cunha”, “Randolfe”, “Bolsonaro”, “Dallagnol”, “Mantega”, “Hélio José”, “FHC”, “Herman Benjamin” e “Lindberg”), ideologias políticas de esquerda (“petista”, “comunista” e “socialista”) e

⁵⁵ O Iramuteq seleciona automaticamente apenas os termos com mais de 10 citações no banco de dados. Este trabalho escolheu por manter tal critério.

⁵⁶ Importante lembrar que as notícias falsas que estão presentes no banco Fake.Br condicionam tematicamente a maioria das notícias verdadeiras.

lideranças internacionais (“Kim Jong-un”, “Xi Jinping”, “Putin”, “Donald Trump”, [Angela] “Merkel”, “Shinzo” [Abe]) foram significativamente mais presentes entre as notícias falsas. Além disso, esse tipo de conteúdo ainda estava mais preenchido de termos ligados ao âmbito militar/bélico (“guerra”, “bomba”, “bombardeiro”, “invadir”, “arma”, “tanque”, “bombardear”, “exército”, “ISIS”⁵⁷, “general”, “tomahawk”⁵⁸, “soldado” e “balístico”).

Em contraposição as notícias verdadeiras citaram proporcionalmente mais o nome de Doria⁵⁹, Mercadante, Obama e Fernando Haddad, além dos nomes da procuradora geral da república Raquel Dodge. Vale destacar ainda uma presença maior de termos ligados às pessoas e empresas envolvidas no escândalo de corrupção do “Petrolão” entre as notícias verdadeiras (“OAS”, “Odebrecht”, “Emílio” [Odebrecht], “Glaucos da Costamarques”, “Delcídio” [Amaral], “Eliseu Padilha” e “Paulo Okamotto”) e siglas de alguns partidos (“PSDB”, “DEM”, “MDB”, “PSOL”, “PCdoB” e “PDT”). Além disso, há uma presença massiva de termos jurídicos no conteúdo verdadeiro (“processo”, “decisão”, “recurso”, “sanção”, “inquérito”, “condenação”, “julgamento”, “investigação”, “denúncia”, “instância”, “direito”, “liminar” e “desembargador”)⁶⁰.

Esses resultados coadunam com a hipótese de que as notícias falsas apelam ao uso do nome de lideranças e ideologias políticas, trazendo consigo afetos e tendendo a ativar a cognição quente (LOGE; TABER, 2005), além de se utilizar do medo bélico para chamar a atenção dos leitores e despertar o senso de urgência e ansiedade em torno de uma ameaça nuclear/militar. Por outro lado, as notícias verdadeiras focam mais em termos técnicos jurídicos e políticos, além de dar mais atenção a nomes periféricos da política institucional, mas que estavam diretamente ligados ao noticiário da época. Entretanto, diferentemente do que era esperado, as siglas partidárias – em sua maioria – foram mais citadas entre as notícias verdadeiras, apenas o PP, PT e PMDB (que mudou de nome em agosto de 2017 para MDB) foram mais mencionados entre as desinformações. Esse resultado pode estar ligado ao fato de que esses três partidos – à época – foram diretamente ligados a escândalos de corrupção do Petrolão e Carne Fraca, sendo fundamentais para a ativação de afetos ligados ao sentimento anticorrupção encontrado em notícias falsas, como no trecho a seguir.

⁵⁷ Referente ao grupo terrorista Estado Islâmico.

⁵⁸ Referente aos mísseis teleguiados norte-americanos de nome *Tomahawk*.

⁵⁹ Entretanto, a maior presença do nome do então prefeito de São Paulo entre as notícias verdadeiras pode ter se dado pelo uso incorreto do termo *Dória* nas desinformações.

⁶⁰ Os resultados completos, com todos os termos encontrados pelo Iramuteq, podem ser verificados e replicados a partir dos dados disponíveis no seguinte link: https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo.

(txt_1113) Os advogados de Antônio Palocci também declararam para Sérgio_Moro (em suas alegações finais) que Mantega autorizou pagamentos ilegais (através da Odebrecht) para o casal de marqueteiros João Santana e Mônica Moura? tais pagamentos foram feitos na Suíça. Os valores constantes da planilha italiano eram destinados ao PT e não ao acusado depois que Antonio Palocci deixou o cargo, o dinheiro passou a ser administrado por uma terceira pessoa, alegou a defesa. (MONTEIRO *et al.*, 2019).

O Iramuteq também realiza a classificação das classes gramaticais, categorizando cada termo presente no *corpus*. A partir da análise de especificidades é possível realizar a comparação entre o banco de dados das notícias falsas e verdadeiras e, portanto, testar as hipóteses 7 e 8 deste trabalho. A primeira hipótese que pode ser conferida com o uso dessas técnicas sobre as classes gramaticais é a H7, que afirma que as desinformações possuem mais pronomes pessoais, adjetivos e verbos que as informações verdadeiras. Ela é baseada em parte da literatura que encontrou essas características de forma reiterada em notícias falsas (BAPTISTA, 2020; CHOY; CHONG, 2018; PÉREZ-ROSAS *et al.*, 2020; MONTEIRO *et al.*, 2019). É possível afirmar que a hipótese se confirma parcialmente nesse primeiro teste, tanto a categoria genérica dos adjetivos⁶¹ (4,55), como os verbos (43,56) e os verbos suplementares (1,23) foram mais encontrados de forma significativa entre as notícias falsas.

No entanto, o Iramuteq subdivide os pronomes em diversas subcategorias, desses apenas os pronomes possessivos se concentraram de forma significativa entre as notícias falsas (3,07). Os pronomes interrogativos também eram mais presentes nesse grupo, mas por uma margem menor (0,72). Pronomes relativos, indefinidos, demonstrativos e pessoais foram mais utilizados no texto das notícias verdadeiras.

Já a hipótese 8 afirma que a desinformação possui em média menos substantivos e numerais que as informações verdadeiras. Essa afirmação se baseia em parte da literatura que encontrou tendências desse tipo em outros testes (OTT *et al.*, 2011; HORNE; ADALI, 2017; RASHKIN, 2017), acredita-se que a maior presença dessas formas gramaticais nesse grupo esteja ligada ao detalhamento dos fatos reportados, bem como o desenvolvimento de uma contextualização com informações complementares.

Os dados obtidos na análise corroboram parcialmente essa tese. Tanto os substantivos (2,86) como os substantivos suplementares (2,61) apresentaram uma presença maior no banco de notícias verdadeiras. Os numerais, contudo, tiveram uma concentração maior entre as

⁶¹ A análise do Iramuteq identificou duas categorias de adjetivos, levamos em conta a com maior número de termos identificados.

notícias falsas (0,77). Vale ressaltar, entretanto, que os adjetivos numéricos⁶² foram significativamente mais presentes no banco de dados de notícias verdadeiras (12,84), o que indica que os redatores desse tipo de conteúdo tendem a escrever mais numerais por extenso que os criadores de desinformação, mas o mesmo não vale para os algarismos numéricos.

Entretanto, é importante destacar outros resultados significativos que não foram previstos pelas hipóteses levantadas. Foi encontrada uma maior concentração de preposições (27,41), pronomes pessoais (14,49) e conjunções (4,18) entre as notícias verdadeiras, resultados que não foram reportados em outros estudos sobre o tema. As preposições e as conjunções, em específico, podem estar ligadas ao maior refinamento na escrita dos textos jornalísticos presentes nesse banco, o uso de termos como “em”, “de”, “a”, “por”, “sobre”, “entre”, “contra”, “sem” e “desde” indicam um cuidado na produção do conteúdo e no desenvolvimento da descrição do fato relatado. Esse resultado também pode ter relação com as descobertas da literatura sobre a lexicometria das mentiras, que encontraram uma maior presença de preposições e conjunções – chamadas de palavras excluientes pelos pesquisadores – entre pessoas que relatam histórias verdadeiras (PENNEBAKER *et al.*, 2003; HANCOCK *et al.*, 2007; NEWMAN *et al.*, 2002).

No entanto, os dados colhidos de pronomes pessoais foram de encontro à boa parte da literatura. Houve uma maior presença de termos como *nós*, *eu*, *ele* e *ela* entre as notícias verdadeiras, apenas o pronome pessoal *você* foi mais encontrado entre as informações falsas, sem muita significância, contudo. Esse resultado pode apontar para uma tentativa dos redatores de notícias falsas em emular alguma imparcialidade, evitando o uso dessa categoria gramatical.

Tabela 4 – Análise de especificidades das classes gramaticais presentes no banco de dados

Formas	Notícias Falsas	Notícias Verdadeiras	Nº de citações
Verbo	43,5692	-43,5692	54915
Não Reconhecido	11,2476	-11,2476	29541
Artigo Definido	9,9442	-9,9442	30269
Adjetivo	4,5507	-4,5507	26393
Advérbio Suplementar	3,1422	-3,1422	37764
Pronome Possessivo	3,0756	-3,0756	82
Verbo Suplementar	1,2384	-1,2384	3113
Adjetivo	0,9241	-0,9241	149

⁶² Essa categoria engloba os numerais escritos por extenso (um, dois, mil etc.), bem como numerais ordinais (primeiro, segundo, terceiro etc.).

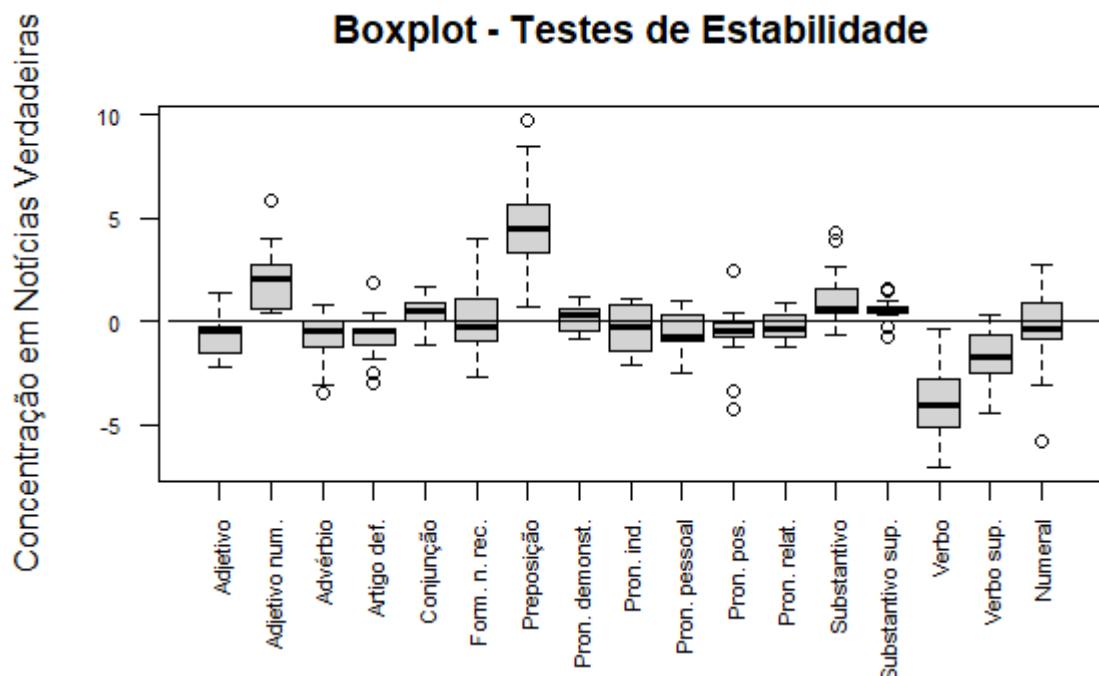
Formas	Notícias Falsas	Notícias Verdadeiras	Nº de citações
Numeral	0,7753	-0,7753	7254
Pronome Interrogativo	0,7295	-0,7295	299
Sigla	0,6212	-0,6212	13
Adjetivo Suplementar	-0,2449	0,2449	4
Advérbio	-0,6521	0,6521	25845
Pronome Relativo	-0,6747	0,6747	219
Onomatopeia	-0,7066	0,7066	15510
Pronome Indefinido	-1,5378	1,5378	2650
Pronome Demonstrativo	-2,3097	2,3097	6427
Substantivo Suplementar	-2,6148	2,6148	1635
Substantivo	-2,8657	2,8657	130693
Conjunção	-4,1887	4,1887	15524
Adjetivo Numérico	-12,8431	12,8431	26466
Pronome Pessoal	-14,499	14,499	4481
Preposição	-27,4142	27,4142	92004

Fonte: produção do autor, 2022.

Para testar se esses resultados não estão sendo influenciados por possíveis *outliers* – textos maiores com características léxicas muito particulares – foram criados bancos de dados a partir de segmentos textuais. O Iramuteq permite que o pesquisador crie um *subcorpus* a partir de termos específicos, ao selecionar esses termos o programa cria uma nova base de dados com os trechos em que a palavra foi originalmente citada. Buscou-se selecionar termos relacionados com a temática política da época, a seguir apresentamos as palavras escolhidas e a quantidade de vezes que elas foram citadas no banco de dados original (Apêndice H). A partir deste ponto foram realizadas análises de especificidades em cada *subcorpus*, os resultados das comparações das principais classes gramaticais foram transcritos em uma tabela e plotados.

É possível verificar na figura 1 que algumas categorias de palavras apresentaram uma maior concentração em um determinado tipo de notícia (verdadeira/falsa) em todos os 15 *subcorpus* analisados. Este foi o caso das preposições, adjetivos numéricos e verbos, enquanto os dois primeiros apresentaram uma concentração maior nos textos retirados de notícias verdadeiras em todos os 15 *subcorpus*, os verbos foram mais presentes nos textos enganosos.

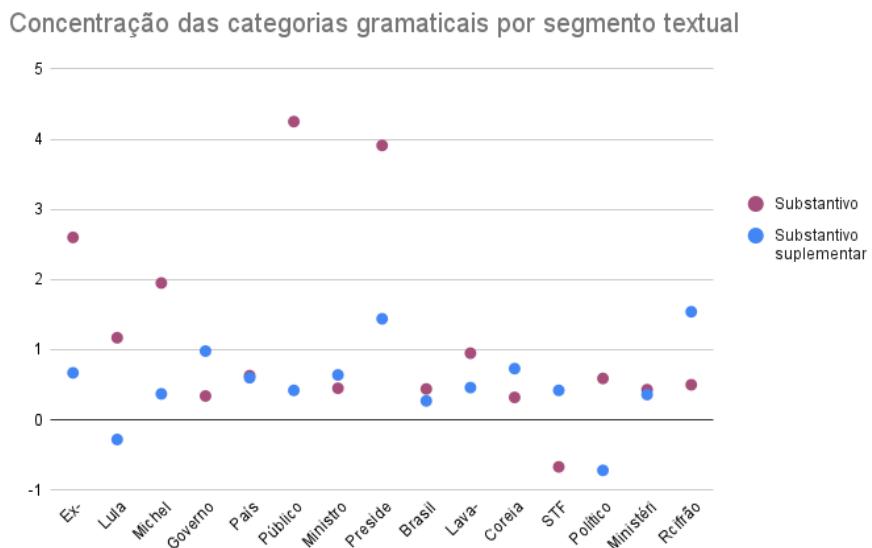
Gráfico 1 – Visualização produzida a partir das análises de especificidades das classes gramaticais encontradas nos 15 subcorpus. Valores negativos representam uma maior concentração da categoria em notícias falsas



Fonte: produção do autor, 2022.

Entretanto, outras formas também apresentaram resultados estáveis na maioria dos *subcorpus* estudados. Os gráficos das figuras 2 e 3 detalham melhor os dados obtidos a partir das análises de especificidades das classes gramaticais. A imagem da figura 2 indica como os substantivos, incluindo os suplementares, são encontrados de forma mais concentrada, com uma estabilidade razoável, nos segmentos extraídos de notícias verdadeiras. Essa categoria gramatical foi mais recorrente nas notícias verdadeiras em 14 dos 15 *subcorpus*, enquanto os suplementares em 13.

Gráfico 2 – Visualização produzida através da análise de especificidades dos *subcorpus* criados a partir de termos políticos. Valores negativos indicam uma maior presença da classe gramatical em textos falsos

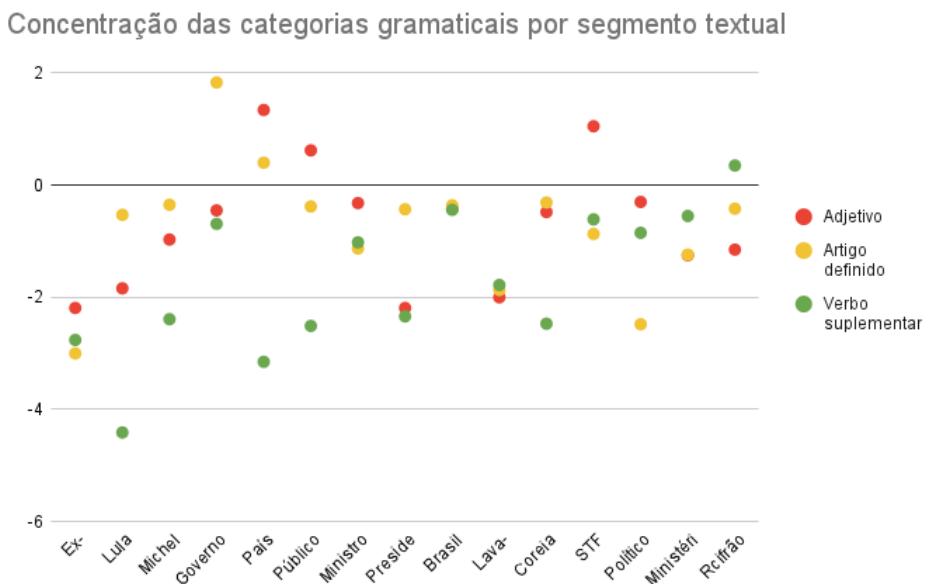


Fonte: produção do autor, 2022.

Já o gráfico da figura 3 detalha a maior presença de adjetivos (12), artigos definidos (13) e verbos suplementares (14) entre as notícias falsas. Os resultados iniciais obtidos indicam – assim como boa parte da literatura sobre o tema – que certas classes gramaticais são mais encontradas, de forma significativa e estável, em notícias falsas (verbos, adjetivos e artigos definidos) e o mesmo acontece com os textos jornalísticos verdadeiros (preposições, numerais escritos por extenso e substantivos).

Essas descobertas apontam para uma menor complexidade estrutural da escrita das notícias falsas com temática política. A maior presença de verbos e adjetivos podem estar ligados ao caráter mais descritivo da desinformação, limitada a transcrever um fato ou declaração e, sempre que possível, apresentar algumas adjetivações morais. As informações falsas geralmente não gastam muitas palavras com a explicação do contexto em que o fato narrado se insere, algo que a maioria das matérias jornalísticas fazem. Já que a intenção desse tipo de conteúdo é gerar uma reação rápida através do processamento heurístico da informação, se espera que ela seja curta e carregue apenas alguma história revoltante ou provocante que incentive seu compartilhamento.

Gráfico 3 – Visualização produzida através da análise de especificidades dos subcorpus criados a partir de termos políticos. Valores negativos indicam uma maior presença da classe gramatical em textos falsos.



Fonte: produção do autor, 2022.

Alguns indícios que corroboram ainda mais essa tese são: 1 - o tamanho médio das notícias falsas, elas são bem menores que as notícias verdadeiras; 2 - a maior presença de termos ligados à temática bélica/militar e; 3 - a maior presença de termos acusatórios e sensacionalistas entre as notícias falsas, características já ligadas ao sucesso de compartilhamento de informações políticas em redes sociais (HASELL; WEEKS, 2016).

Por fim, foi construído um banco de dados com notícias falsas e verdadeiras, advindas do banco Fake.Br, que não tinham a política como tema central. Foram reunidos 140 textos, sendo 70 notícias falsas e 70 verdadeiras que abordavam desde casos de violência, passando pelo dia a dia de celebridades e questões esportivas. O banco com notícias verdadeiras contém 80.065 palavras, enquanto o de notícias falsas 9.787. A média de palavras por texto verdadeiro foi de 1.492 e 195 nos falsos, ambas maiores que a média das publicações que envolviam política. Novamente, os resultados corroboram a hipótese 4, indicando um tamanho significativamente menor das notícias falsas, se comparadas com as verdadeiras.

Este banco em específico possibilita o teste da hipótese 6, que defende que a temática política interfere significativamente nos resultados ligados à presença de categorias gramaticais e à legibilidade dos textos de notícias falsas e verdadeiras. O Apêndice B permite a observação dos 20 termos ativos mais citados de cada categoria.

Tanto o Apêndice B, como os dados da tabela 6 permitem a observação da falta de termos referentes a temáticas mais específicas, como havia no banco de dados principal. Os resultados obtidos validam a seleção realizada durante o levantamento, já que não há a presença de termos claramente ligados ao noticiário político.

Tabela 5 – Termos mais citados no banco de dados sem notícias políticas

Notícias Falsas		Notícias Verdadeiras	
Termo	Concentração	Termo	Concentração
durante	13,7893	samba	4,8479
médico	10,8901	projeto	4,5214
publicar	10,8158	escola	3,8685
vaticano	10,04	região	3,2158
ator	9,4094	como	3,1383
igreja	9,1133	ponto	3,1225
informar	7,6198	gente	2,9184
câncer	6,9973	segundo	2,7036
hospital	6,8764	forma	2,684
papa	6,3707	violência	2,684

Fonte: produção do autor, 2022.

Novamente, observa-se a validação da seleção, nenhum termo em específico está diretamente ligado ao noticiário político. Apenas o nome *Vaticano* remete ao país europeu, mas também ao noticiário religioso. Nota-se ainda, a partir de uma observação inicial, a maior concentração de palavras ligadas a temas sensíveis entre as notícias falsas. Termos referentes a enfermidades (*Médico, Câncer, Hospital, Saudável, Doença*) e religião (*Vaticano, Igreja, Papa, Deus*) são mais presentes entre as desinformações selecionadas. Enquanto os dados das notícias verdadeiras apresentam uma variedade temática de difícil categorização. Nesse quesito, a hipótese 1, que prevê a presença de termos ligados a temas sensíveis e divisórios, se confirma mais nesse teste que no realizado com as notícias com temática política.

Por fim, comparamos os dados ligados à estrutura gramatical de ambas as categorias. Com isso busca-se verificar se o fator político das notícias importa de alguma forma para a estrutura das mesmas. A partir desses dados é possível afirmar que a maioria das formas gramaticais que apresentaram uma relação significativa com uma das categorias estudadas, voltaram a demonstrar essa tendência neste teste. A hipótese 7, que previa a maior presença de verbos, adjetivos e pronomes pessoais entre as notícias falsas, foi corroborada nesta análise de

especificidades. Novamente os verbos (13,07) e adjetivos (0,9) apresentaram uma maior concentração entre as desinformações, no entanto, os pronomes pessoais também foram mais encontrados nesse grupo (0,64), diferentemente do que havia acontecido no banco com temática política.

Por outro lado, a hipótese 8, que previa a maior presença de substantivos e numerais nas notícias verdadeiras, não repetiu os mesmos resultados. As preposições (2,35) e adjetivos numéricos (5,75) continuaram mais presentes em notícias verdadeiras, enquanto os numerais e substantivos apresentaram uma concentração maior nas notícias falsas.

Tabela 6 – Análise de especificidades das formas gramaticais presentes nas notícias falsas e verdadeiras sem temática política

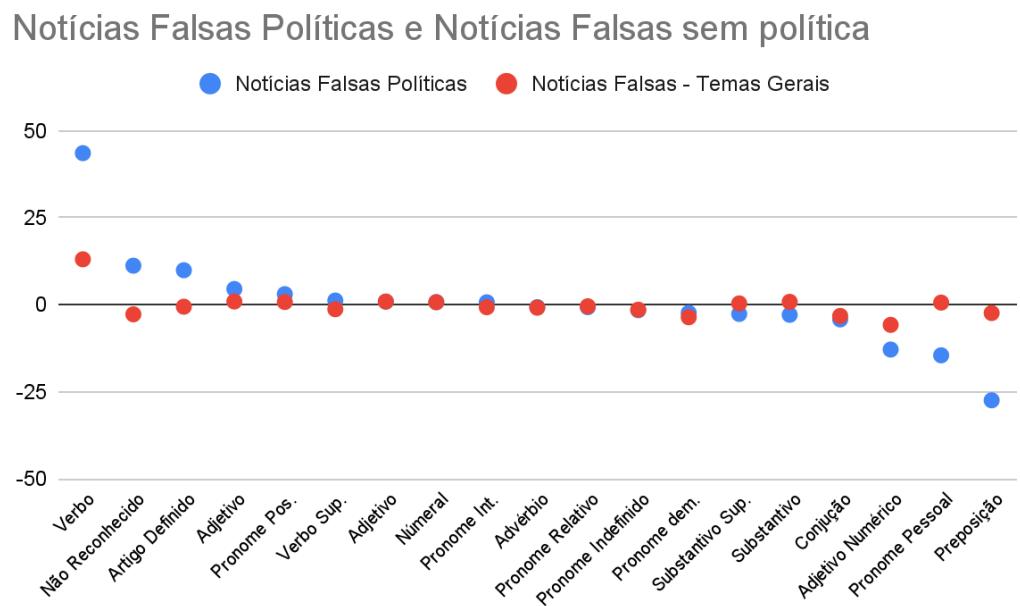
Análise de especificidades dos tipos gramaticais		
Tipo gramatical	Notícias Falsas	Notícias verdadeiras
Verbo	13,0793	-13,0793
Adjetivo	0,9567	-0,9567
Adjetivo	0,9389	-0,9389
Substantivo	0,8679	-0,8679
Pronome Possessivo	0,8133	-0,8133
Numeral	0,7565	-0,7565
Pronome Pessoal	0,6444	-0,6444
Substantivo Suplementar	0,4081	-0,4081
Pronome Relativo	-0,4112	0,4112
Artigo Definido	-0,5301	0,5301
Pronome Interrogativo	-0,6854	0,6854
Advérbio	-0,8256	0,8256
Verbo Suplementar	-1,2594	1,2594
Pronome Indefinido	-1,3969	1,3969
Preposição	-2,3587	2,3587
Forma Não Reconhecida	-2,7204	2,7204
Conjunção	-3,1846	3,1846
Pronome Demonstrativo	-3,5621	3,5621
Adjetivo Numérico	-5,752	5,752

Fonte: produção do autor, 2022.

Destaque-se ainda algumas classes gramaticais que apresentaram correlação no banco de dados político, mas que não demonstraram os mesmos resultados significativos agora, este foi caso dos pronomes possessivos. Algumas classes, contudo, apresentaram resultados diametralmente opostos, como os artigos definidos e verbos suplementares, que tendiam a ser encontrados mais entre notícias falsas, e dos pronomes pessoais, que apresentavam uma correlação forte com o banco de dados de notícias verdadeiras.

Esse resultado aponta para uma maior proximidade estrutural entre as notícias falsas e verdadeiras que não abordam a política como tema principal. Uma explicação possível para essa proximidade é a possível especialização dos redatores de desinformação política, que parece se adaptar mais ao modelo de “viralização” exigido pelas redes sociais, ou, ainda, uma maior preocupação com a mimetização do formato jornalístico entre os redatores de desinformações com outros temas. Os próximos testes podem oferecer mais respostas sobre essas suposições.

Gráfico 4 – Resultados comparados dos dados obtidos a partir das análises de especificidades realizadas em notícias falsas com temática política e sem temática política



Fonte: produção do autor, 2022.

5.2 LEGIBILIDADE

Uma das principais explicações correntes sobre o consumo e compartilhamento de desinformação *on-line* passa pela ideia de que os usuários das redes sociais usam

primordialmente o processamento heurístico de informações durante o tempo em que acessam essas mídias (PENNYCOOK *et al.*, 2021; PENNYCOOK; RAND, 2018, 2021). Uma das formas de incentivar a ativação desse modelo de processamento é garantindo uma leitura rápida (por muitas vezes dinâmica) e simples ao usuário, o que pode ser feito através do tamanho e legibilidade da informação (HORNE; ADALI, 2017; AGNHOTRI; BHATTACHARYA, 2016). Nesta seção serão realizados diversos testes de legibilidade com o intuito de verificar a validade da H5, que afirma que as desinformações são mais legíveis que as notícias verdadeiras, e H6, que declara que essas características são mais ressaltadas em notícias falsas com conteúdo político.

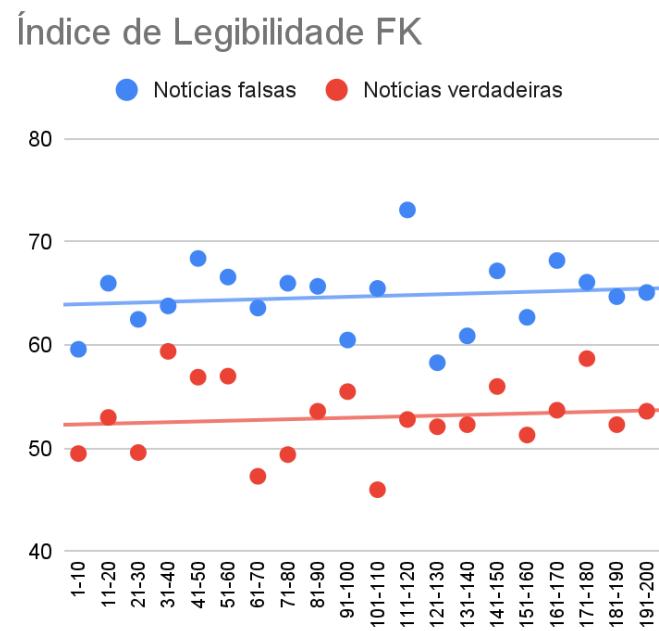
As análises de legibilidade foram realizadas a partir do *software* ALT (legibilidade.com), foram utilizadas como métricas de legibilidade: 1 - Teste Flesch-Kincaid, que vai de 0 a 100, onde 0 representa textos mais complexos e 100 mais legíveis; 2 - Nível de escolaridade Flesch-Kincaid, que vai de 0 a 20, onde 0 representa textos mais legíveis e 20 textos mais complexos; e 3 - Índice de Coleman-Liau, que possui classificações similares ao de escolaridade Flesch-Kincaid.

Todos essas métricas se baseiam no pressuposto de que o número de sílabas ou palavras por frase estão relacionadas com a complexidade de compreensão dos textos, elas já foram utilizadas anteriormente em outros estudos para verificar se os textos presentes em notícias falsas são mais acessíveis que os verdadeiros (HORNE; ADALI, 2017).

Para testar essa hipótese, selecionamos 400 notícias do banco Fake.Br (MONTEIRO *et al.*, 2019), 200 verdadeiras e 200 falsas. Subdividimos essas notícias em 40 grupos, cada um contendo 10 textos e criamos um novo banco de dados com os resultados obtidos a partir das análises de legibilidade disponibilizadas pelos *software* ALT.

Os gráficos da figura 5 apresentam os resultados do teste de legibilidade Flesch-Kincaid. É possível observar que os grupos com notícias falsas foram mais legíveis que os seus correspondentes verdadeiros, as linhas de tendência deixam claras as diferenças significativas entre as duas categorias. Esse teste inicial indica a disposição prevista pela H5, as notícias falsas possuem uma escrita mais simples e dinâmica que as notícias verdadeiras, um fator importante de incentivo ao processamento heurístico.

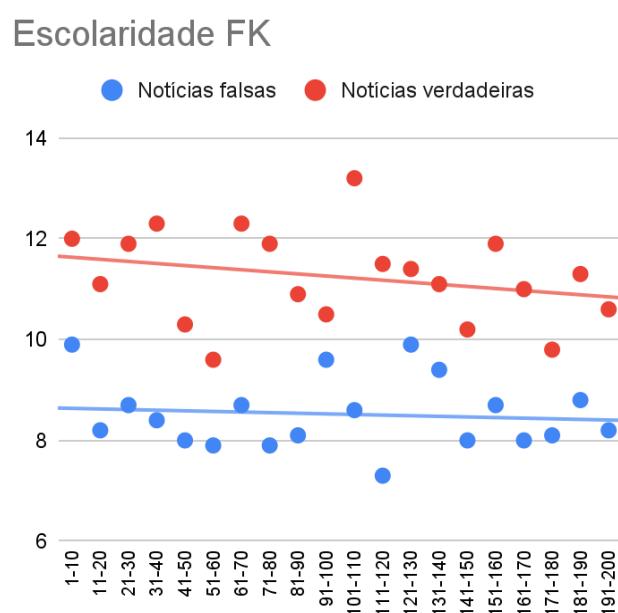
Gráfico 5 – Visualização produzida a partir das análises de legibilidade baseadas no índice Flesch-Kincaid



Fonte: produção do autor, 2022.

A figura 6 apresenta os resultados advindos da métrica de nível de escolaridade Flesch-Kincaid. Novamente, os grupos de textos contendo notícias falsas foram significativamente mais legíveis que os seus correspondentes de notícias verdadeiras.

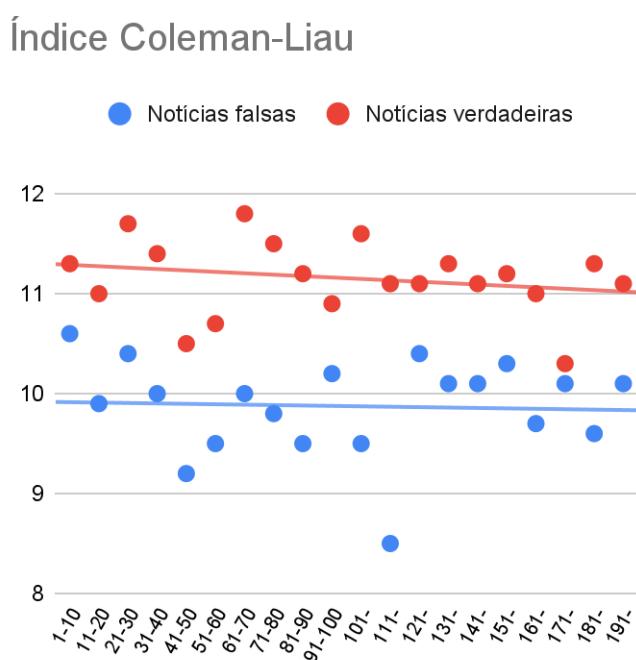
Gráfico 6 – Visualização produzida a partir das análises de legibilidade baseadas no nível de escolaridade Flesch-Kincaid



Fonte: produção do autor, 2022.

Por fim, foi realizado o teste de legibilidade a partir do índice de Coleman-Liau (1975), os resultados podem ser vistos na figura 7. Diferentemente das métricas baseadas nos trabalhos de Flesch (1979), que conta as palavras dos textos analisados, esse modelo contabiliza a quantidade de letras como base de seu cálculo. Entretanto, os seus resultados foram bastante próximos dos obtidos a partir do nível de escolaridade Flesch-Kincaid. Todos os grupos de textos que continham notícias falsas apresentaram uma maior simplicidade em sua escrita que os correspondentes da base de dados com as notícias verdadeiras.

Gráfico 7 – Visualização produzida a partir das análises de legibilidade baseadas no índice Coleman-Liau



Fonte: produção do autor, 2022.

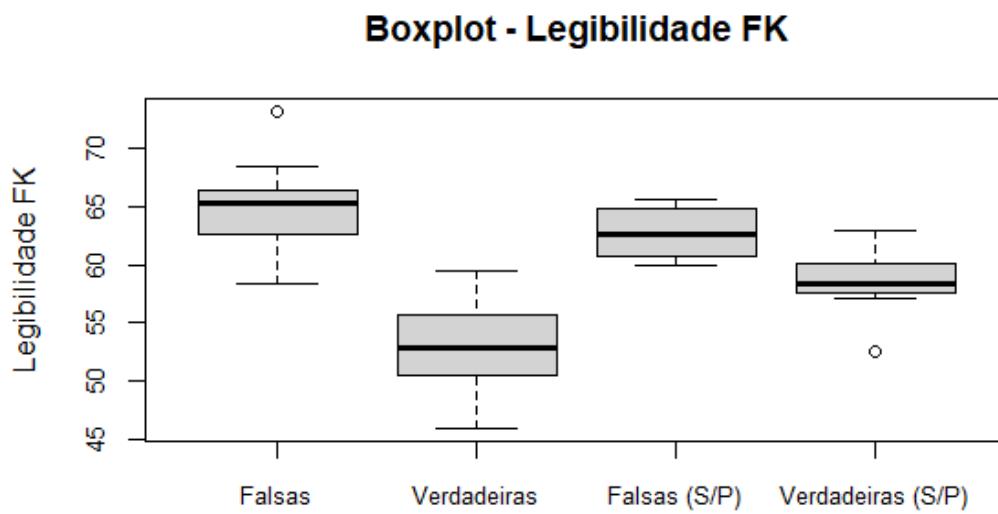
Também foram realizados testes de legibilidade no banco de dados com notícias sem temática política. O intuito deste teste é verificar se as diferenças de legibilidade entre as notícias falsas e verdadeiras se replicam em informações sem conteúdo político. Este experimento final complementa a validação da hipótese 6. Espera-se, portanto, que a temática política interfira diretamente nos resultados e que a diferença de legibilidade entre as notícias falsas e verdadeiras sem conteúdo político sejam menores que as que tratavam desse tema. Para possibilitar esse teste foram reunidas 140 notícias, 70 verdadeiras e 70 falsas, retiradas do banco Fake.Br (MONTEIRO *et al.*, 2019) que não possuíam conteúdo político⁶³. Para aplicação dos índices foi reproduzida a mesma técnica do teste principal: subdividimos o banco em *subcorpus*

63 Esta seleção pode ser verificada no banco de dados online desta dissertação. Ver em: https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo.

com 10 textos cada, dessa forma submetemos, no total, 14 grupos de textos à análise do *software* ALT.

Os primeiros dados obtidos a partir do índice de legibilidade Flesch-Kincaid (FLESCH, 1979) indicam que as notícias falsas sem conteúdo político também são – em média – mais legíveis que as verdadeiras, resultados similares aos encontrados no banco com publicações que tratavam do tema, voltando a corroborar a H5. No entanto, é possível observar algumas tendências a partir dessa análise inicial, como a maior proximidade entre os resultados. A termo de comparação, o intervalo médio entre os valores de legibilidade das notícias políticas falsas e verdadeiras no índice FK foi de 11,72 pontos, o mesmo intervalo nos bancos com textos sem foco político é de 4,32. Isso se dá por dois motivos: 1 - As notícias verdadeiras que não envolvem política têm uma escrita mais acessível; 2 - As notícias falsas que não envolvem política possuem uma escrita mais complexa.

Gráfico 8 – Visualização em boxplot produzida a partir da análise de legibilidade do índice Flesch-Kincaid.

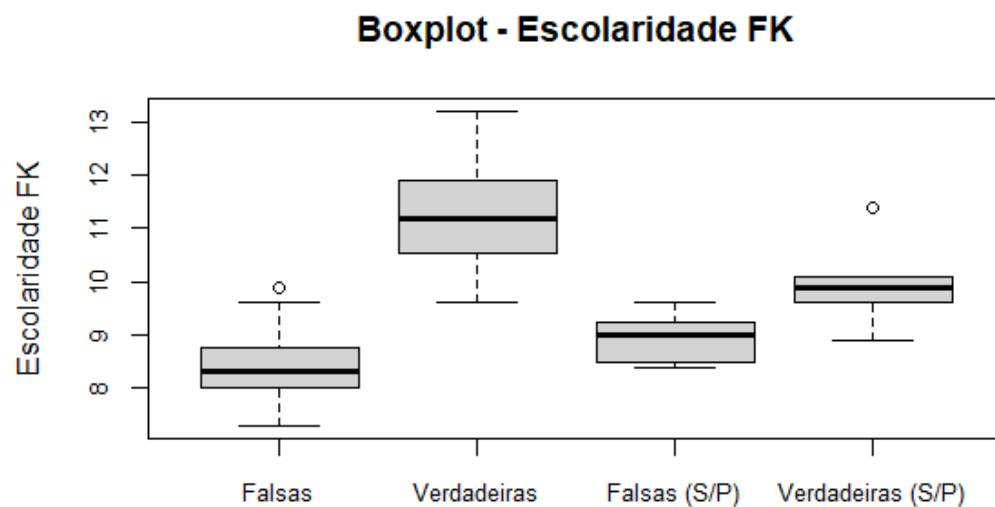


Fonte: produção do autor, 2022.

Essa afirmação pode ser verificada de forma mais clara a partir da média dos valores de cada categoria. Os *corpus* de notícias falsas com conteúdo político marcaram 64,72 pontos em média no índice FK, enquanto as que não abordavam o tema pontuaram 62,74. Da mesma forma, os grupos de textos com notícias políticas verdadeiras marcaram 53 pontos de média, enquanto que tratavam de outros temas pontuaram 58,42.

Os resultados do índice de escolaridade FK seguem apontando para a tendência das notícias falsas de serem escritas de maneira mais simples que as verdadeiras, mesmo quando a temática não é política. Entretanto, da mesma forma que no índice de legibilidade, a classificação de escolaridade também apontou uma menor diferença entre as notícias falsas e verdadeiras nesse banco ($2,72 \times 1,02$). Novamente tanto as notícias falsas sem conteúdo político se mostraram mais complexas, em média, que as que abordavam a temática ($8,52 \times 8,92$), como as verdadeiras foram escritas de forma mais simples ($11,24 \times 9,94$).

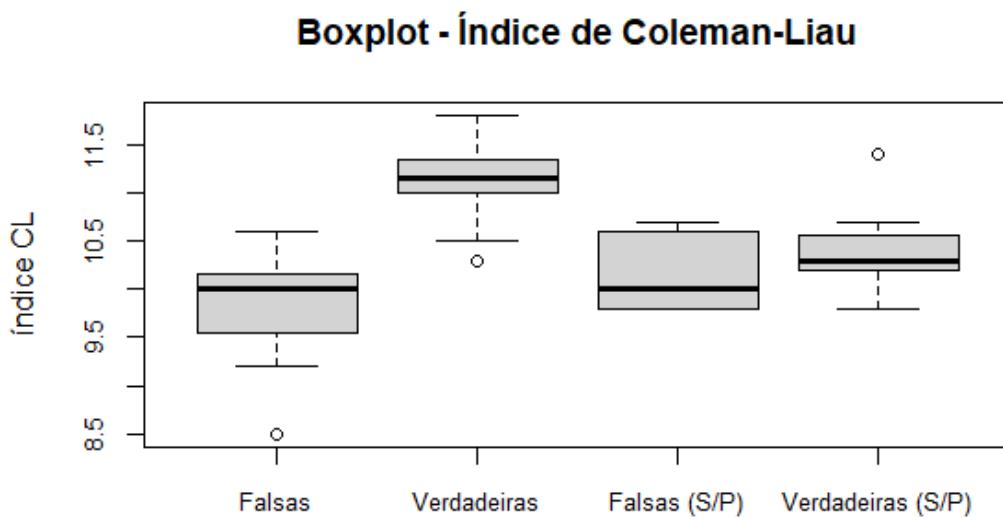
Gráfico 9 – Visualização em boxplot produzida a partir da análise de legibilidade do índice de escolaridade Flesch-Kincaid.



Fonte: produção do autor, 2022.

Os dados obtidos a partir dos testes de legibilidade via índice de Coleman-Liau seguem os verificados nos testes FK. Novamente as notícias falsas tendem a ser escritas de maneira mais simples que as verdadeiras, assim como no banco de dados político. Entretanto, os resultados continuam indicando uma maior complexidade das notícias falsas sem conteúdo político ($9,87 \times 10,18$) e uma escrita mais simples nas notícias verdadeiras ($11,15 \times 10,42$).

Gráfico 10 – Visualização em boxplot produzida a partir da análise de legibilidade do índice de Coleman-Liau.



Fonte: produção do autor, 2022.

Os resultados encontrados a partir da análise de legibilidade confirmam, portanto, as descobertas realizadas em outros trabalhos (HORNE; ADALI, 2017), e apontam que as notícias falsas possuem uma escrita mais simples que as notícias verdadeiras. Ou seja, elas possuem em média menos palavras e letras por sentença, indicativos de menor complexidade. Esses testes confirmam a hipótese 5, e são um indicativo do porquê do sucesso das notícias falsas nas redes sociais, sua linguagem simples, aliada ao modelo de processamento de informações baseado na heurística, são uma receita perfeita para o consumo em massa de desinformação *on-line*. Essa simplicidade aliada ao sensacionalismo e ao uso preponderante de conceitos sociopolíticos explicam em parte a causa de tais mensagens serem tão “viralizantes” nas redes sociais.

Contudo, os testes realizados com as notícias de conteúdos variados – que não tinham a pauta política como questão central – demonstram que as características aventadas na hipótese 5 são mais acentuadas no noticiário político, o que fortalece a hipótese 6 deste trabalho. Isso se dá porque os textos que abordam a temática política nos jornais possuem uma escrita mais complexa que as demais, enquanto as informações falsas que tratam da pauta são escritas de maneira mais simplificada. Uma possível explicação para essa particularidade pode ser encontrada na tese de investigações anteriores sobre a centralização e especialização do noticiário político desinformativo (CHAGAS *et al.*, 2019; PAVARIN, 2019). Essa expertise

dos redatores desse material pode estar interferindo diretamente na estrutura do conteúdo publicado, algo visto em menor escala entre os criadores de outros tipos de informações falsas. Outra possibilidade é que os autores de desinformações com outros temas possuem outra estratégia – como a mimetização do formato jornalístico – para obter sucesso nas redes.

5.3 TEMAS

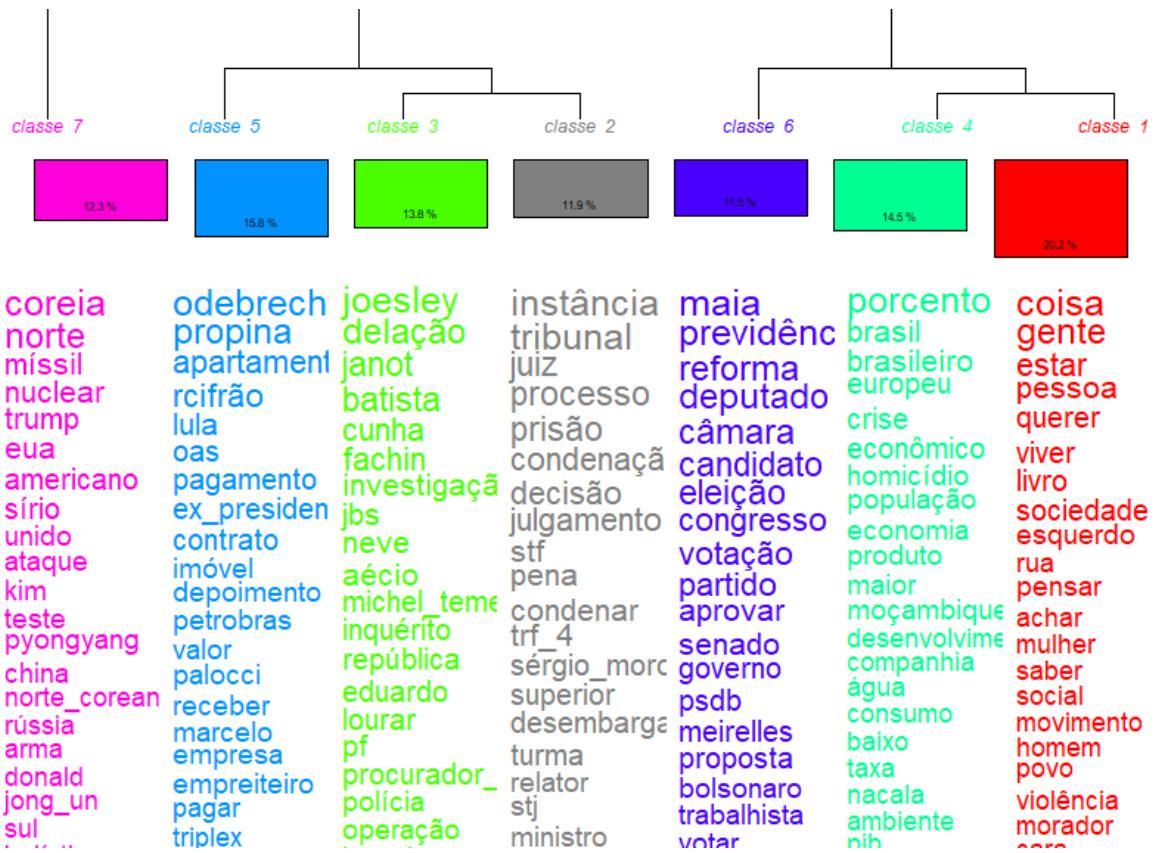
Também foram realizadas análises sobre os principais temas abordados nos dois bancos de dados. Utilizou-se da técnica da classificação hierárquica descendente (REINERT, 1990) para observar as principais pautas abordadas e verificar possíveis diferenças no tratamento de cada tema. Para realizar essa comparação foi necessário dividir o banco de dados geral em dois, já que a CHD oferecida pelo Iramuteq não subdivide o *corpus* pelas variáveis previamente classificadas, tal como a análise de especificidades faz.

Essa análise busca testar possíveis particularidades temáticas dos dois bancos de dados. Espera-se, novamente, a preponderância de temas divisórios (H1) e conceitos sociológicos ativadores de cognição quente entre as notícias falsas (H3), bem como uma abordagem mais simples dos temas tratados nas desinformações, o que também facilitaria o uso de heurísticas (H5). O uso de dendrogramas e das AFCs⁶⁴ possibilitará uma melhor visualização dessas abordagens.

A classificação hierárquica descendente, realizada pelo *software* Iramuteq na base de dados que continha informações verdadeiras, identificou 12.374 segmentos de texto, destes 12.101 foram classificados, cerca de 97,8% do *corpus* textual. Como é possível observar a partir da figura 11, foram estabelecidas 7 classes que preenchem 3 grandes eixos temáticos.

⁶⁴ As análises fatoriais por correspondência, bem como a exploração inicial dos dados, estão presentes nos apêndices C, D e E.

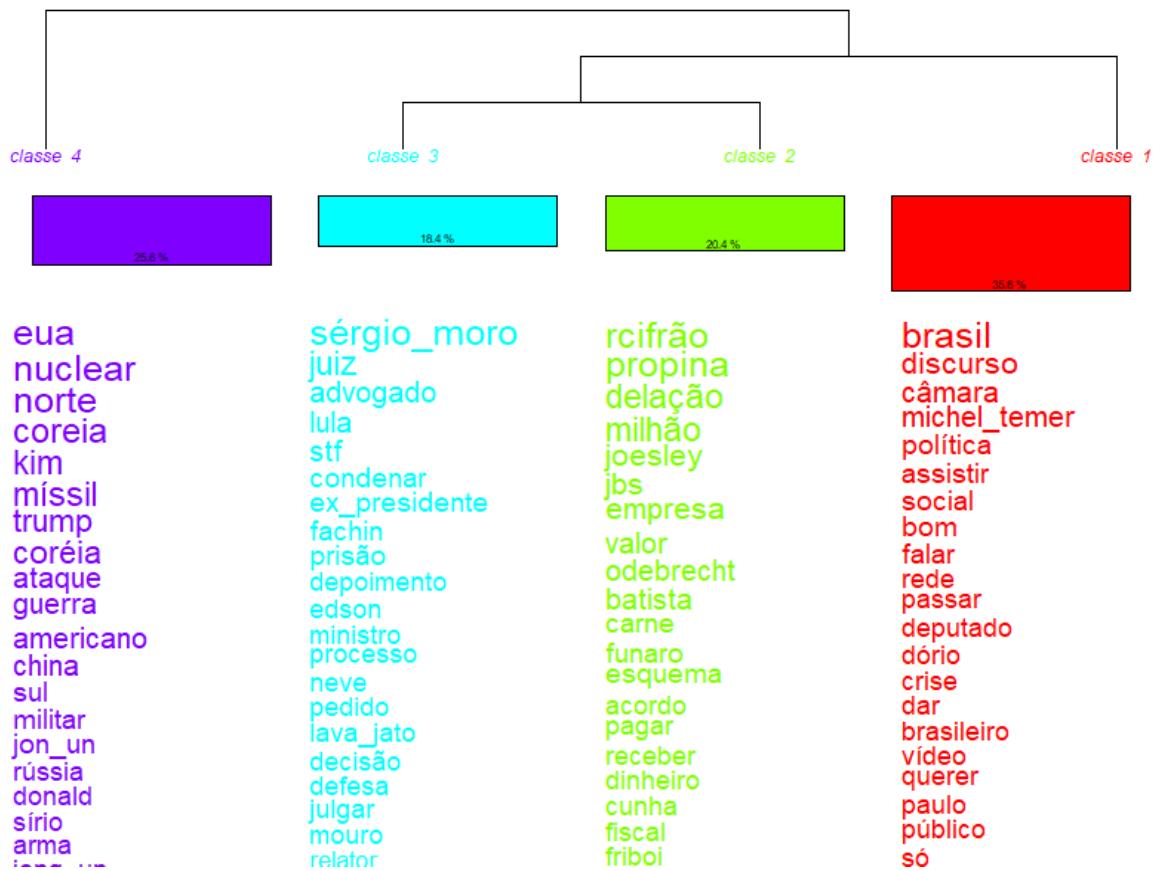
Figura 1 – Dendrograma obtido a partir da CHD das 400 notícias verdadeiras retiradas do Fake.Br.



Fonte: produção do autor, 2022.

O banco de dados com notícias falsas também passou pelo mesmo processo de análise via CHD. Foram 1.906 segmentos textuais identificados nesse *corpus*, destes 1.747 (91,66%) foram classificados em uma das classes apresentadas no dendrograma da figura 11. Pode-se observar, novamente, a presença de três grandes eixos temáticos no banco com textos desinformativos, entretanto, desta vez a CHD delimitou apenas 4 classes. Diferentemente das notícias verdadeiras, que interagem bastante entre si, as classes delimitadas a partir das notícias falsas são bem isoladas (exceto pela 2 e 3), o que pode ser verificado na AFC presente no Apêndice G.

Figura 2 – Dendrograma obtido a partir da CHD das 400 notícias falsas retiradas do Fake.Br



Fonte: produção do autor, 2022.

A análise das CHDs via dendrogramas e AFCs permite a observação de diferenças significativas na abordagem das pautas tratadas nos dois bancos de dados⁶⁵. Antes de mais nada, é importante relembrar que o banco Fake.Br foi levantado com o intuito de garantir, na medida do possível, similaridade temática entre as notícias falsas e verdadeiras. De início é possível notar que esse objetivo foi parcialmente alcançado, afinal de contas, os assuntos abordados nos dois bancos por muitas vezes se encontram. Ambos tratam da crise no mar da Coreia, dos escândalos de corrupção da época, do andamento judicial dos processos envolvendo estes escândalos e os encaminhamentos da política institucional.

Entretanto, enquanto o banco com notícias verdadeiras apresenta linhas bem detalhadas dos temas abordados, deixando claras – a partir da CHD – as diferenças entre os escândalos ligados ao MDB, como a operação Carne Fraca, e ao PT, no caso o Petrolão e tratando de pautas econômicas em seu corpo, as notícias falsas simplificaram sua abordagem, tratando das acusações a partir de nomes e termos centrais. Esses resultados indicam a

⁶⁵ Para mais detalhes sobre a exploração dos dados recomenda-se a leitura dos apêndices D e F.

simplicidade descritiva presente nas desinformações, tanto que os grandes eixos temáticos são representados de forma isolada na AFC correspondente (Apêndice E). Ou seja, diferentemente das notícias verdadeiras, que promovem a interação de diversas classes temáticas, provavelmente com o intuito de apresentar os contextos em que os fatos estão inseridos (seja econômicos, institucionais ou investigativos), as notícias falsas isolam seus temas (corrupção, Coreia e política diária), evitando relacionar o fato abordado com outros assuntos, diminuindo assim os custos informacionais para o leitor.

Esses resultados são indicadores que corroboram a H3 e H5. O modelo de abordagem temática das notícias falsas é simplificado e tende a ser monotemático, evitando correlações com outros temas. Esse tratamento aponta uma certa preocupação por parte dos redatores da desinformação com a simplicidade e sensacionalismo. A notícia falsa é, além de curta, estruturada em torno de um fato (geralmente inventado ou deturpado), se utiliza de nomes importantes da política como base de suas pautas e não traz outras questões ao texto, focando unicamente na desinformação relatada.

6. DISCUSSÃO

Os dados obtidos a partir dos testes realizados neste trabalho apontam para algumas diferenças substanciais entre a estrutura textual das notícias falsas e verdadeiras. Os resultados advindos do banco Fake.Br (MONTEIRO *et al.*, 2019) demonstram uma estratégia de escrita voltada para exposição de descrições simplificadas e a descontextualização das informações. Além disso, diferentemente do que se apontava inicialmente em parte da literatura (LAZER *et al.*, 2018), as notícias falsas não reproduzem com perfeição o formato jornalístico em seu corpo textual, longe disso.

Essa diferença ocorre principalmente porque os dois formatos de conteúdo possuem finalidades e pressupostos completamente diferentes. Enquanto o jornalismo tradicional, fonte de grande parte do banco de dados de notícias verdadeiras, preza pela técnica de redação e pela apuração da informação, os autores das desinformações estudadas almejam – prioritariamente – a “viralização” do seu conteúdo nas mídias sociais, o engajamento emotivo e a escalada da polarização política.

Os textos de caráter jornalístico tradicionalmente buscam responder às seguintes perguntas ao tentar relatar um fato: *O quê* aconteceu? *Quando* se deu? *Onde* ocorreu? *Como* ou *de que forma* veio a se dar o fato? *Por que* o fato relatado aconteceu. Os manuais de redação jornalística recomendam que essas perguntas sejam respondidas sob uma estrutura específica que envolve uma manchete (ou título), a lide (ou *lead*), que complementa o título, fornece as principais informações da notícia e tem como função despertar a atenção do leitor para o texto, e o corpo, que contém o desenvolvimento mais amplo e detalhado dos fatos (SARMENTO; TUFANO, 2004). A matéria a seguir, retirada do banco Fake.BR, apresenta essa estrutura, a lide está grifada.

(txt_1020) PPL anuncia pré-candidatura de João Vicente Goulart à Presidência da República. Filho do ex-presidente João Goulart, o pré-candidato fundou instituto em homenagem ao pai e disputará o cargo pela primeira vez. Jango foi deposto pelo golpe militar de 1964. O Partido Pátria Livre (PPL) anunciou nesta terça-feira (13) o nome de João Vicente Goulart como pré-candidato do partido à Presidência da República. A legenda também anunciou o professor da Universidade Católica de Brasília (UCB) Léo Alves como pré-candidato a vice-presidente. Pelo calendário eleitoral deste ano, os candidatos só serão confirmados entre 20 de julho e 5 de agosto, período em que poderão ser realizadas as convenções partidárias. Filho do ex-presidente João Goulart, o pré-candidato do PPL fundou um instituto em homenagem ao pai e disputará o cargo pela primeira vez. Segundo a revista "Época", João Vicente Goulart decidiu deixar o PDT no ano passado, em razão de estar insatisfeito com o atual presidente da sigla, o ex-ministro do Trabalho Carlos Lupi. João Vicente Goulart é autor do livro "Jango e Eu: Memórias de um exílio sem volta". Na nota em que anunciou a pré-candidatura, o PPL afirma que o país vive um momento de "miséria, desindustrialização, devastação dos serviços públicos, terrível insegurança pública e o mais lastimável espetáculo de decadência moral". A legenda afirmou, ainda, que a pré-candidatura "é

a que melhor condensa a luta histórica do povo brasileiro, algo premente na situação atual de nosso país, exausto devido a uma política insana de destruição nacional, que conduziu ao desemprego de milhões de brasileiros; ao fechamento de milhares de empresas nacionais; ao malbaratamento dos bens públicos – e à corrupção absolutamente alucinada e sem peias, exposta vergonhosamente à nação nos últimos anos". Outros partidos também já definiram seus pré-candidatos ao Palácio do Planalto, entre os quais o PT, que anunciou o nome do ex-presidente Luiz Inácio Lula da Silva; o PDT, que lançou o ex-governador do Ceará Ciro Gomes; o DEM, que tentará eleger o atual presidente da Câmara, Rodrigo Maia; e o PSL, que lançou o nome do deputado Jair Bolsonaro. Jango era o presidente do Brasil em 1964, quando foi deposto pelo golpe militar com apoio de setores da sociedade, que temiam que ele desse um golpe de esquerda, o que os partidários dele sempre negaram. O ex-presidente morreu em Mercedes (Argentina) em 1976 e, segundo o site do Instituto João Goulart, "supõe-se que tenha sido envenenado" (MONTEIRO *et al.*, 2019, grifo do autor).

No entanto, após uma análise detalhada dos resultados obtidos, é possível concluir que, diferentemente das notícias verdadeiras, as informações falsas não possuem corpo. O redator do material desinformativo se limita a responder *o quê, quando e onde* se deu o fato relatado, a parte do *como* muitas vezes é simplificada ou inexistente no relato, já o *porquê* é ignorado. Os exemplos a seguir demonstram esse tipo de estrutura:

(txt 1005) Boechat sobre Lula: "Sua ALMA pode ser honesta, mas no seu governo fez-se o diabo". "Lula tem todo o direito de dizer que sua ALMA pode ser honesta, porém isso não significa nada." disse o apresentador. " No seu governo fez-se o diabo [...] no seu partido fez-se o diabo [...] o tesoureiro do PT Vaccari, que o sr. tentou isentar, está em cana [...] o Zé Dirceu, figura mais importante do seu partido, está em cana [...]" disse Boechat. confira: (MONTEIRO *et al.*, 2019).

(txt 1006) PT diz que Lula não tem dinheiro para sobreviver e chama Moro de "mentiroso, mesquinho e conspirador". O Partido dos Trabalhadores soltou uma nota de repúdio à decisão do juiz Sérgio Moro. Sobre o bloqueio dos patrimônios do ex-presidente (R\$ 606 mil), o partido alega que Lula está pobre e não tem mais dinheiro para o "básico". abre aspas Depois de condenar o ex-presidente Lula sem provas, propagar mentiras e contradizer sua própria sentença, o juiz Sergio Moro decidiu agora vingar-se de um inocente. Ao bloquear os bens de Lula, Moro decretou uma pena de asfixia econômica que priva o ex-presidente de sua casa, dos meios para subsistir e até para se defender das falsas acusações. Foi uma decisão mesquinha, tramada em segredo ao longo de 9 meses com a Força Tarefa de Curitiba, e concluída após a forte reação da sociedade e do mundo jurídico à sentença injusta no caso do tríplex. É um caso típico de retaliação, de quem se vale da cumplicidade com a Rede Globo para cometer todo tipo de arbitrariedades contra o maior líder popular do País. fecha aspas (MONTEIRO *et al.*, 2019).

(txt 1008) Coordenador da Lava-Jato se revolta: "Aumento de impostos? Os corruptos desviam 200 bilhões por ano...". O procurador da República Deltan Dallagnol , coordenador da Força Tarefa do Ministério Público Federal na Lava-Jato, é uma das várias pessoas que lutam para tirar o país das mãos dos "ladrões de gravata" Em uma mensagem publicada nas redes sociais, Dallagnol desabafou sobre o aumento dos impostos anunciado pelo governo Temer: Precisamos recuperar nossa dignidade "É claro que os brasileiros vão compreender o aumento de impostos [...] desviam 200 bilhões por ano praticando corrupção [...] gastam mais do que devem com emendas milionárias a fim de livrar Temer das acusações de corrupção... "E agora querem colocar a conta no nosso bolso, aumentando impostos. Toda vez que eu for abastecer o carro, vou pensar na saúde e na educação dos pobres, nos buracos das estradas e na infraestrutura precária que prejudica os investimentos [...] vou lembrar disso tudo" (MONTEIRO *et al.*, 2019)

Uma das consequências dessa diferença de estrutura é a presença massiva de verbos e adjetivos e a carência de preposições, conjunções, substantivos e adjetivos numéricos nas *fake news*. De fato, as notícias falsas seguem o modelo jornalístico tradicional até certo ponto, há uma manchete sensacionalista, seguida por uma lide tão chamativa quanto, mas elas não desenvolvem a contextualização em que o fato relatado se insere, pelo contrário, qualquer tipo de conclusão da matéria se dá através da inserção de opiniões do autor, que geralmente se utiliza de adjetivações morais para se referir às figuras políticas citadas no texto.

Outra característica presente nesse tipo de conteúdo é o modelo exacerbadamente declaratório de notícias (os exemplos anteriormente citados deixam isso claro). O chamado jornalismo declaratório é uma prática vista em diversos canais de comunicação e que vem sendo criticada por jornalistas e comunicadores, ela consiste em uma série de estratégias de publicação que priorizam as declarações de atores políticos e o debate entre opiniões distintas, dando pouca importância a apuração do objeto discutido (OLIVEIRA, 2020). O modelo declaratório permite que grandes conglomerados jornalísticos se tornem reprodutores de discursos oficiais e assessorias, conferindo crédito a narrativas que são, na maioria das vezes, planejadas⁶⁶.

As notícias falsas, contudo, levam esse modelo a outro patamar. As desinformações estudadas se utilizam recorrentemente de citações polêmicas de alguma figura pública, seja ela falsa ou descontextualizada, de maneira a atacar um alvo político ou, até mesmo, o próprio citado. As declarações são seguidas por comentários do redator, geralmente recheados de adjetivações e detrações ao político, partido ou mesmo instituição difamada. Essas características básicas do formato das notícias falsas indicam novamente o porquê da presença preponderante de verbos nesse modelo de comunicação, já que essas formas gramaticais geralmente são utilizadas para descrever ações, enquanto a maior concentração de adjetivos pode ser explicada pela simplificação do contexto relatado e pela inserção rápida da opinião do autor da desinformação.

A falta de conjunções e, principalmente, preposições é outro indicativo desse formato simplificado e excessivamente declaratório. Preposições como *contra*, *sem* e *exceto* fazem parte das chamadas *exclusive words*, palavras consideradas indicativos de complexidade da formulação de um texto ou história. Além de serem mais utilizadas quando as pessoas falam a verdade (PENNEBAKER *et al.*, 2003; HANCOCK *et al.*, 2007; NEWMAN *et al.*, 2002), essas

⁶⁶ Um dos casos recentes mais emblemáticos se deu na cobertura da operação Lava-Jato, segundo relatos de jornalistas e da própria ex-assessora do ex-juiz Sérgio Moro “a imprensa comprava tudo que era vazado” pelo magistrado. Ver mais em: <https://theintercept.com/2018/10/29/lava-jato-imprensa-entrevista-assessora/>.

categorias gramaticais também indicam uma redação mais cuidadosa e voltada para o desenvolvimento de descrições detalhadas.

Outro resultado que indica essa estrutura simplificada são as “ilhas” temáticas encontradas na AFC das notícias falsas. Essa visualização gráfica da classificação hierárquica descendente demonstra como certos termos interagem apenas entre si nas desinformações. Isso representa a simplicidade temática desse conteúdo, que se limita a citar os mesmos atores políticos dentro dos mesmos cenários (Lula x Moro; Trump x Kim Jong-un; Corrupção do MDB...), evitando a interação entre as diferentes pautas da época.

É possível observar também a falta de cuidado com detalhes de ortografia (vide a presença de termos como “Coreia”, “Jon-um” e “Dória”, que figuraram entre as palavras com maior presença entre as notícias falsas), característica já apontada no trabalho de Monteiro (*et al.*, 2019). Esse resultado pode estar relacionado a dois atributos ligados à produção de notícias falsas: a agilidade, visto que alguns trabalhos já apontavam a imensa produção que pequenos centros disseminadores de desinformação eram capazes de fazer em pouco tempo (LAZER *et al.*, 2018; ALLCOTT; GENTZKOW, 2017), e a inexistência de uma editoria – com a presença de revisores – que possam verificar possíveis erros de ortografia no texto. A predisposição para a maior utilização de numerais no formato extenso (adjetivos numéricos) nas notícias verdadeiras também pode estar ligada à técnica jornalística e ao detalhamento de informações.

Também se verificou uma maior legibilidade dos textos desinformativos. Os testes realizados ratificam a ideia de que o conteúdo advindo do banco com informações verdadeiras possui uma escrita significativamente diferente das informações falsas. As descobertas vão, mais uma vez, de encontro à ideia inicial de que as notícias falsas almejam mimetizar o formato jornalístico tradicional (GOMES; DOURADO, 2019), o que põe em xeque a classificação inicial de *fake news* dada por Lazer (*et al.*, 2018) indicando que as desinformações políticas não tentam reproduzir o formato tradicional de escrita jornalística, pelo contrário, possuem um modelo específico de escrita que pouco se parece com os textos presentes nos jornais. Seja na forma como trata o objeto, relata os fatos, ou mesmo na complexidade da escrita, as notícias falsas possuem uma estrutura própria, capaz de gerar suspeitas ao olhar atento de um leitor.

Entretanto, é justamente essa particularidade estrutural que guarda parte do segredo do sucesso das desinformações. Textos de escrita simples, monotemáticos, preenchidos com conceitos sociopolíticos, com alguma informação bombástica/sensacionalista são ativadores de raciocínio motivado e do processamento heurístico. Esse formato de comunicação, ao mesmo

tempo que reativa afetos a partir da citação e elogio/detração de figuras políticas, traz consigo um texto curto de linguagem simplificada, perfeitamente adaptável para o compartilhamento nas mais diversas redes sociais.

Por fim, os resultados indicam como essas variáveis são ressaltadas no noticiário político. A diferença estrutural das notícias falsas e verdadeiras que têm essa pauta como central é mais acentuada que as de textos com outros temas. Esses dados indicam algumas particularidades da escrita, tanto das desinformações como dos textos jornalísticos que tratam da política nacional e internacional. Isso pode estar ligado ao editorial político dos jornais, geralmente um setor mais técnico e que exige um maior cuidado de apuração e redação. Mas também demonstra a especificidade desse tipo de desinformação, seja pela especialização dos seus redatores ou pela finalidade diferente das outras notícias falsas.

Essa descoberta possibilita um questionamento importante: é a desinformação política um fenômeno comunicacional específico? Estruturalmente falando, restou claro que o texto desinformativo político não busca reproduzir o formato de escrita jornalística, mas isso não parece se repetir de forma tão clara entre as notícias que não abordam esse tema. Aparentemente as notícias falsas sobre celebridades, esportes e outros temas tentam se aproximar mais do formato de escrita jornalística, agindo de forma semelhante ao que acreditava LAZER (*et al.*, 2018). Esse resultado demonstra como os princípios de produção e a finalidade dos dois tipos de desinformação podem ser diferentes, enquanto as notícias falsas com conteúdo político buscam ativar emoções a partir de um modelo hiperpartidário e extremamente simplificado de escrita, as suas similares que tratam de outras questões seguem a estratégia de mimetização da escrita jornalística.

Fica claro, portanto, que esse novo formato de comunicação entre as elites e a militância política possui características importantes e que devem ser analisadas com cuidado. A diferença significativa entre os textos falsos e verdadeiros pode responder a questionamentos que vão além de modelos computacionais capazes de reconhecer automaticamente desinformações, ela está diretamente relacionada com o sucesso desse formato de propaganda política nos últimos anos, principalmente nos ambientes de mídias sociais. Sucesso esse que pode ser explicado pela relação processamento-estrutura da informação, cada vez mais saliente nas desinformações políticas.

Seja por adotar uma estratégia de redação mais parecida com a escrita publicitária do que jornalística, ou por seguir tendências há muito estabelecidas pelo jornalismo sensacionalista

e pela imprensa marrom, as desinformações apresentam características únicas capazes de explicar – parcialmente – o seu sucesso de compartilhamento e, até mesmo, a predisposição de crença de indivíduos em histórias consideradas absurdas. Além disso, trata-se de uma estratégia barata e eficiente de engajamento político e definição de agenda.

Estudos futuros, no entanto, podem verificar se essas diferenças estruturais tornam o texto mais atrativo para todos os tipos de leitores, ou se as características individuais dos consumidores de notícias podem estar relacionadas com esse formato específico. A produção de pesquisas com design experimental é capaz de fornecer algumas das respostas às perguntas deixadas em aberto, verificando o potencial de “viralização” de um conteúdo a partir do controle das variáveis relacionadas à veracidade, afetividade e simplicidade de um texto. Recomenda-se também que sejam levadas em conta variáveis de controle ligadas ao entrevistado como o partidarismo, a idade (ou analfabetismo informacional) e a sofisticação política.

6 LIMITAÇÕES E AGENDA DE PESQUISA

É importante, antes de adentrar nas conclusões deste trabalho, delimitar algumas das limitações que ele possui. Com isso busca-se também apontar uma agenda de pesquisa que possa preencher os espaços vazios que essa análise deixa. A primeira restrição elencada se refere à profundidade estatística. Apesar dos diversos testes aplicados ao longo da pesquisa, é possível comparar *corpus* textuais através de análises de texto automatizadas com maior validade estatística, como testes randomizados, por exemplo, tal como realizado por MONTEIRO (*et al.*, 2019).

Da mesma forma, os *softwares* que foram utilizados nesta empreitada acadêmica possuem grande potencial como ferramenta de exploração e, principalmente, testes de hipóteses, entretanto, também carregam consigo algumas limitações. O Iramuteq é um programa sólido, já testado em diversas linhas de pesquisa – inclusive na ciência política – e foi adaptado ao léxico português há algum tempo, sendo utilizado desta forma em diversos estudos (DOURADO, 2021; CERVI, 2018; BAPTISTA, 2020). No entanto, ele não identifica os símbolos de pontuação, que seriam de grande utilidade para esta pesquisa e apresenta algumas falhas de categorização, como a que ocorreu com os adjetivos e na preposição *em* nos testes do banco Fake.Br.

Já o programa ALT (legibilidade.com), apesar de ter sido uma ferramenta extremamente útil para as análises aqui aplicadas, é um projeto recente que ainda não foi testado em outras produções acadêmicas. Da mesma forma, o medidor de legibilidade, diferentemente do Iramuteq, não disponibiliza muitas ferramentas que possibilitem a replicação dos dados testados. No entanto, este trabalho disponibiliza os testes realizados no programa em arquivos *html*⁶⁷.

Além disso, um problema vivido por quase todo pesquisador da área também se fez aqui: encontrar uma base de dados ampla e replicável. Uma das poucas bases com notícias falsas e verdadeiras disponibilizada publicamente é, justamente, a do projeto Fake.Br (MONTEIRO *et al.*, 2019). A grande maioria dos outros estudos feitos no Brasil ou não disponibilizam as bases levantadas, ou não trabalham com informações falsas e verdadeiras ao mesmo tempo (o que impede a comparação). Até hoje o foco da maioria das pesquisas sobre desinformação no Brasil continua sendo o processo de distribuição e consumo, e não o conteúdo espalhado. Além disso, o período de coleta dos dados já está um pouco defasado (as notícias

⁶⁷ Ver em: https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo.

mais recentes são de janeiro de 2018), o que impediu que o trabalho obtivesse conclusões mais pertinentes com o noticiário político contemporâneo.

Outra questão que atinge esse trabalho se refere a visualização de fenômenos que não são especificamente a *falsidade* da informação. É possível que as diferenças estruturais visualizadas nesta pesquisa sejam sintomas de outras características dos textos observados, como o hiperpartidarismo midiático (de onde muitas dessas matérias vieram) ou mesmo a escrita profissional jornalística. Trabalhos futuros podem utilizar uma base de dados com textos precisos advindos de mídias hiperpartidárias como variável de controle, possibilitando assim uma maior certeza de que as características visualizadas estão relacionadas, de fato, à falsidade do texto, e não ao seu hiperpartidarismo.

Também é necessário abordar os problemas ligados à literatura existente sobre a estrutura da informação falsa. Apesar dos diversos trabalhos que almejam criar modelos preditivos capazes de selecionar desinformações automaticamente, e do relativo sucesso destes (FREIRE; GOLDSCHMIDT, 2019), o porquê da existência de diferenças estruturais entre as notícias falsas e verdadeiras ainda é pouco abordado pela literatura, tanto nacional como a internacional. Estudos como o de Pennycook (*et al.*, 2021), contudo, apontam a importância do contexto e do formato do conteúdo consumido no processamento de informações.

Por fim, sabe-se que há uma diferença significativa entre o modelo de redação presente nas desinformações e nas informações precisas, resta a pergunta – já levantada neste trabalho – se essas diferenças interferem de forma significativa na forma como consumimos material político e se ele é um dos quesitos que fazem as *fake news* se disseminarem de forma mais rápida e eficiente que as notícias verdadeiras nas redes sociais *on-line*. Uma conclusão segura sobre essa pergunta só será alcançada a partir de experimentos que testem esses formatos de escrita com diversos públicos.

7 CONCLUSÃO

As notícias falsas são mais curtas, monotemáticas e legíveis que as verdadeiras. Além disso, esse tipo de conteúdo se utiliza mais de nomes de lideranças políticas, termos referentes a ideologias de esquerda e ameaças bélicas/militares em seu corpo. Também são caracterizadas por recorrerem mais a verbos e adjetivos relacionados à moral, além de conter, proporcionalmente, menos preposições, substantivos e adjetivos numéricos em seu texto, o que está ligado ao seu formato excessivamente declaratório e sensacionalista que rejeita a apresentação de dados e contextualização da informação.

De fato, as notícias falsas são redigidas sob modelos muito mais atraentes para o usuário das mídias sociais, principalmente se este indivíduo for partidário ou apoiador de algum movimento ou figura política. Elas ativam os modelos mentais de processamento de informações que todos nós possuímos, capazes de interferir diretamente no modo como vemos o mundo e interagimos com outras pessoas. As desinformações, portanto, seguem a cartilha da “viralização”, sendo eficazes ao citar termos ligados à cognição quente, promovendo a ativação de afetos ligados ao viés de confirmação e raciocínio motivado (LODGE; TABER, 2005). Além disso, elas possuem uma abordagem sensacionalista inserida em um texto simples e curto, algo perfeito para um leitor apressado que, diante de tantas informações presentes em suas *timelines*, se utiliza reiteradamente da heurística para consumir e julgar o que é passível de compartilhamento (PENNYCOOK *et al.*, 2021).

No entanto, talvez a descoberta mais promissora desta pesquisa seja a relação da temática política com essas características. Notícias falsas que tratam desse tema são, em média, mais curtas, simples, monotemáticas e declaratórias que suas similares, isto é, aquelas que abordam outras questões. Essa revelação é um indicativo de como essas desinformações são os principais exemplos desse formato de comunicação e como elas buscam – e alcançam –, através de ferramentas textuais, suas maiores finalidades: a ampla divulgação e o engajamento.

Esta pesquisa demonstrou, portanto, que a ideia inicial de que as notícias falsas buscam mimetizar o formato jornalístico para confundir o leitor não se sustenta a partir de uma observação comparativa dos dados textuais. Mais especificamente, as desinformações políticas parecem desconsiderar o modelo de escrita jornalística em sua produção. Sabendo disso, talvez seja necessário começar a tratar esse formato de comunicação como um fenômeno específico, mais ligado à propaganda política, com suas características particulares, que ao modelo clássico da boataria espalhada por tablóides de fofoca.

Diante dessas descobertas um cenário preocupante se apresenta. As notícias falsas políticas, sob esse modelo de escrita simples, se apresentam como uma ferramenta barata e eficiente de publicidade. A disseminação em massa de desinformação é capaz de interferir diretamente no comportamento político de indivíduos partidários, principalmente quando reproduzidas em conjunto com setores da elite política. Sendo assim, porque não as utilizar?

A criação de um sistema organizado de desinformação se tornou uma estratégia política de sucesso e lucrativa para políticos e produtores. Até o momento não existem muitas desvantagens que tornem essa escolha pouco atrativa. As instituições ainda são bastante tímidas no combate à desinformação, principalmente no que se refere à punição de atores políticos que reproduzem material difamatório. As mídias sociais muito menos, não há interesse em reduzir uma das principais fontes de engajamento virtual. Quase não existem, no momento da produção deste trabalho, barreiras formais e informais capazes de impedir ou gerar desinteresse pelo uso desse formato de comunicação.

Esta análise, portanto, acaba por apontar que esse novo formato de comunicação dificilmente vai ser abandonado, a ascensão das mídias hiperpartidárias são a prova disso. A estabilidade democrática passa, atualmente, pelas poucas linhas de textos emocionados e legíveis feitos para minar o diálogo e as pontes entre os diversos setores político-ideológicos do país. As descobertas aqui realizadas indicam que o sucesso das *fake news* pode ser um fenômeno mais complexo do que parece.

Espera-se, contudo, que essas descobertas apresentadas mudem um pouco o panorama da pesquisa sobre desinformação, voltando os olhos da academia para os detalhes presentes no conteúdo deste material e seus possíveis efeitos no comportamento político da população. É fundamental saber que a observação e análise das *fake news*, sob os mais diversos ângulos, pode apresentar novos caminhos que viabilizam a melhor compreensão do impacto que essas mensagens provocam na forma como interagimos e escolhemos nossos líderes e representantes nos tempos atuais.

REFERÊNCIAS

- AGNIHOTRI, Arpita; BHATTACHARYA, Saurabh. *On-line review helpfulness: role of qualitative factors*. *Psychology & Marketing*. [S. l.], v. 33, n. 11, p. 1006–1017, 2016.
- ALARCON, Gene; GAMBLE, Rose; JESSUP, Sarah; WALTER, Charles; RYAN, Tyler; WOOD, David; CALHOUN, Chris. Application of the heuristic-systematic model to computer code trustworthiness: The influence of reputation and transparency. *Cogent Psychology*. [S. l.], v. 4, n. 1, 2017.
- ALLCOTT, Hunt; GENTZKOW, Matthew. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*. [S. l.], v. 31, n. 2, p. 211-236, 2017.
- BAPTISTA, João Pedro. Ethos, pathos e logos: análise comparativa do processo persuasivo das (fake) news. *Eikon*. [S. l.], n. 7, 2020.
- BAPTISTA, João Pedro; GRADIM, Anabela. Understanding fake news consumption: a review. *Social Sciences*. [S. l.], v. 9, n. 10, 2020.
- BARDIN, Laurence. *Análise de Conteúdo*. Lisboa: Edições 70, 1977.
- BERELSON, Bernard. *Content analysis in communication research*. Glencoe: The Free Press, 1954.
- BERGER Jonah; MILKMAN Katherine. What Makes *On-line* Content Viral?. *Journal of Marketing Research*. [S. l.], v. 49, n. 2, p. 192-205, 2012.
- BOLSEN, Toby; DRUCKMAN, James; COOK, Fay Lomax. The influence of partisan motivated reasoning on public opinion. *Polit Behav*. [S. l.], v. 36, p. 235–262, 2014.
- BRADY, William; CROCKETT, Molly; VAN BAEL, Jay. The MAD model of moral contagion: the role of motivation, attention, and design in the spread of moralized content *on-line*. *Perspectives on Psychological Science*. [S. l.], v. 15, n. 4, p. 978-1010, 2020.
- BRADY, William; WILLS, Julian; JOST, John; TUCKER, Joshua; VAN BAEL, Jay. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*. [S. l.], v. 114, n. 28, p. 7313-7318, 2017.
- BURSZTYN, Victor S.; BIRNBAUM, Larry. Thousands of small, constant rallies: a large-scale analysis of partisan *Whatsapp* groups. Vancouver, ASONAM'19, 27-30 de Agosto de 2019.
- CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*. [S. l.], v. 21, n. 2, p. 513-518, 2013.
- CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. Tutorial para o uso do software Iramuteq. *Iramuteq*. Disponível em: <http://www.iramuteq.org>. Acesso em: 19 abr. 2018.
- CARPENTER, R. H. Stylistic analysis for law enforcement purposes: A case study of a language variable as an index of a suspect's caution in phrasing answers. *Communication Quarterly*. [S. l.], v. 29, p. 32–39, 1981.

CARVALHO, Priscila Ramos; CASTRO, Paulo César; SCHNEIDER, Marco André. Desinformação na pandemia de Covid-19: similitudes informacionais entre Trump e Bolsonaro. *Em Questão*. Porto Alegre, v. 27, n. 3, p. 15-41, 2021.

CAVIQUE, Luis. Legibilidade de artigos científicos: análise de dados da RCC. *Revista de Ciências da Computação*. [S. l.], v. 3, n. 3, p. 59-65, 2008.

CERVI, Emerson. Análise de conteúdo automatizada para conversações em redes sociais *on-line*: uma proposta metodológica. 42º encontro anual da Anpoc, GT17 (Mídias, política e eleições), Caxambu, 22-26 de outubro de 2018.

CHAGAS, Viktor; CARREIRO, Rodrigo. Macarthismo no zap: Como se comporta a rede anticomunista de apoiadores de Jair Bolsonaro. In: CERVI, Emerson Urizi; WEBER, Maria Helena (org); *impactos político-comunicacionais nas eleições brasileiras de 2018*. Curitiba: CPOP, 2021. p. 201-228.

CHAGAS, Viktor; MODESTO, Michele; MAGALHÃES, Dandara. O Brasil vai virar Venezuela: medo, memes e enquadramentos emocionais no Whatsapp pró-Bolsonaro. *Esferas*. [S. l.], n. 14, 2019. Disponível em: <https://cutt.ly/BARSfdw>. Acesso em: 13 set. 2021.

CHOY, Murphy; CHONG, Mark. Seeing through misinformation: A framework for identifying fake *on-line* news. *Cornell University*, [S. l.], 2018.

COLEMAN, M.; LIAU, T. L. *A computer readability formula designed for machine scoring*. *Journal of Applied Psychology*, v. 60, n. 2, p. 283-284, 1975.

DAMASCENO, E. A. Lexicometria, geração de descritores, construção de ontologias e ensino de línguas: implicações e perspectivas. In: MAGALHÃES, José Sueli de; TRAVAGLIA, Luis Carlos (org.). *Múltiplas Perspectivas em Linguísticas*. 1. ed. Uberlândia: EDUFU, 2008. p. 01-3037.

DECHÊNE, Alice; STAHL, Christoph; HANSEN, Jochim; WANKE, Michaela. The Truth about the truth: a meta-analytic review of the truth effect. *Personality and Social Psychology Review*. [S. l.], v. 14, n. 2, p. 238-257, 2010.

DEPAULO, Bella; LINDSAY, James; MALONE, Brian; MUHLENBRUCK, Laura; CHARLTON, Kelly; COOPER, Harris. Cues to deception. *Psychological Bulletin*. [S. l.], v. 129, n. 1, p. 74–118, 2003.

DOURADO, Tatiana Maria Silva Galvão. *Fake News na eleição presidencial de 2018 no Brasil*. 308 f, 2020. Tese (Doutorado em Comunicação e Cultura Contemporâneas) – Universidade Federal da Bahia, Salvador, 2020.

DOURADO, Tatiana. A natureza política das fake news sobre a eleição presidencial de 2018 no Brasil. In: CERVI, Emerson Urizi; WEBER, Maria Helena (org.). *Impactos Político-Comunicacionais nas Eleições Brasileiras de 2018*. Curitiba: CPOP, 2021. p. 255-284.

DRUCKMAN, James. The politics of motivation. *Critical Review: A Journal of Politics and Society*. [S. l.], v. 24, n. 2, p. 199-216, 2012.

- DRUCKMAN, James; PETERSON, Erick; SLOTHUUS, Rune. How Elite Partisan polarization affects public opinion formation. *The American Political Science Review*. [S. l.], v. 107, n. 1, p. 57-79, 2013.
- EMPOLI, Giuliano da. *Os engenheiros do caos*. Tradução de Arnaldo Bloch. 1. ed. São Paulo: Vestígio, 2020.
- FLESCH, Rudolf. *How to write plain English*. New York: Harper and Row, 1979.
- FLYNN, D. J.; NYHAN, Brendan; REIFLER, Jason. The Nature and origins of misperceptions: understanding false and unsupported beliefs about politics. *Advances in Political Psychology*. [S. l.], v. 38, n. 51, p. 127-150, 2017.
- FREIRE, Paulo; GOLDSCHMIDT, Ronaldo. Uma introdução ao combate automático às fake news em redes sociais virtuais. *Tópicos em gerenciamento de dados e informações*. [S. l.], 2019.
- FUKS, Mário; MARQUES, Pedro. *Afeto ou ideologia: medindo polarização política no Brasil?*. In: 12º ENCONTRO DA ABCP, 2020, Universidade Federal da Paraíba, João Pessoa (PB). Área Temática: Comportamento Político e Opinião Pública [...]. [S. l.: s. n.], 2020.
- GARRETT, Kelly; SUDE, Daniel; RIVA, Paolo. Toeing the party lie: ostracism promotes endorsement of partisan election falsehoods. *Political Communication*. [S. l.], v. 37, 2020.
- GOMES, Wilson; DOURADO, Tatiana. Fake news, um fenômeno de comunicação política entre jornalismo, política e democracia. *Estudos em Jornalismo e Mídia*. [S. l.], v. 6, n. 2, 2019.
- GURREA, Terrádez M. Palabras clave y léxico frecuente como indicadores de mentira. *Behavior & Law Journal*. [S. l.], v. 1, n. 1, p. 63-69, 2015.
- HANCOCK, J. T., CURRY, L. E., GOORHA, S., & WOODWORTH, M.. On Lying and being lied to: a linguistic analysis of deception in computer-mediated communication. *Discourse Processes*. [S. l.], v. 45, n. 1, p. 1-23, 2007.
- HASELL, A.; WEEKS, Brian E. Partisan Provocation: the role of partisan news use and emotional responses in political information sharing in social media. *Human Communication Research*. [S. l.], v. 42, p. 641-661, 2016.
- HORNE, Benjamin; ADALI, Silbel. This is just in: Fake News packs a lot in title, uses simpler, repetitive content intext body, more similar to satire than real news. Proceedings of the International AAAI Conference on Web and Social Media, 11(1). Retirado de: <https://ojs.aaai.org/index.php/ICWSM/article/view/14976>, 2017.
- HOYLER, S. O boato: comunicação patológica. *Revista de Administração de Empresas*. [S. l.], v. 6, n. 21, p. 59-73, 1966.
- IYENGAR, Shanto *et al*. The Origins and consequences of affective polarization in the united states. *Rev. Political Sci.* [S. l.], v. 22, p. 129-146, 2019.
- IZUMI, Maurício; MOREIRA, Davi. O texto como dado: desafios e oportunidades para as ciências sociais. *BIB*. São Paulo, n. 869, p. 138-174, 2018.

JERIT, Jennifer; ZHAO, Yangzi. Political Misinformation. *Annual Review of Political Science* v. 23, n. 1, p. 77-94. 2020.

JONES-JANG, Mo; MORTENSEN, Tara; LIU, Jingjing. *Does media literacy help identification of fake news? information literacy helps, but other literacies don't*. American Behavioral Scientist. [S. l.], v. 65, n. 2, p. 371-388, 2021.

KNAPP Mark; HART Roderick; DENNIS, Harry. An exploration of deception as a communication construct. *Hum. Commun. Res.* [S. l.], v. 1, p. 15–29, 1974.

KUKLINSKI, James; QUIRK, Paul; JERIT, Jennifer; SCHWIEDER, David; RICH, Robert; Misinformation and the currency of democratic citizenship. *Journal of Politics*. [S. l.], v. 62, p. 790-816, 2000.

KUMAR, Srijan; SHAH, Neil; False information on web and social media: a survey. *Arxiv*. [S. l.], 2018. Disponível em: <https://arxiv.org/pdf/1804.08559.pdf>. Acesso em: 07 ago. 2019.

KUNDA, Ziva. The Case for motivated reasoning. *Psychological Bulletin*. [S. l.], v. 108, n. 3, p. 480–98, 1990.

LAZER, David *et al.*. The Science of fake news. *Science*. [S. l.], v. 359, p. 1094-1096, 2018.

LISBOA, Lucas; FERRO, João Victor.; BRITO, José Rubens; LOPES, Roberta; *A Disseminação da desinformação promovida por líderes estatais na pandemia da COVID-19*. In: WORKSHOP SOBRE AS IMPLICAÇÕES DA COMPUTAÇÃO NA SOCIEDADE (WICS), 1. , 2020, Cuiabá. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, p. 114-121, 2020. Disponível em: <https://doi.org/10.5753/wics.2020.11042>. Acesso em: 21 abr. 2021

LODGE, Milton; TABER, Charles; The Automaticity of affect for political leaders, groups, and issues: an experimental test of the hot cognition hypothesis. *Political Psychology*. [S. l.], v. 26, n. 3, 2005.

LYRA, Danilo; AMARAL, Claudio; Apreensibilidade e legibilidade de artigos científicos de um periódico nacional. *Tekhne e Logos*. Botucatu, SP, v. 3, n. 3, 2012.

MARTEL, Cameron; PENNYCOOK, Gordon; RAND, David. Reliance on emotion promotes belief in fake news. *Cogn. Research*. [S. l.], v. 5, n. 4, 2020.

MARTINO, Luís Mauro Sá. *Teoria das Mídias Digitais*: linguagens, ambientes, redes. 2. ed. Vozes: Petrópolis, 2015.

McCOMBS, Maxwell. *Setting the Agenda*: the mass media and the public opinion. 2. ed. Polity Press: Cambridge, 2014.

MELO, Phillip; MESSIAS, Johnnatan; RESENDE, Gustavo; GARIMELLA, Kiran; ALMEIDA, Jussara; BENEVENUTO, Fabrício. *Whatsapp Monitor*: a fact-checking system for *Whatsapp*. Proceedings of the International AAAI Conference on Web and Social Media, v. 13, n. 1, p. 676-677, 2019.

MIHALCEA, Rada; STRAPPARAVA, Carlo. *The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language*. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 309–312, Suntec, Singapore, 4 August 2009.

MONDAK, Jeffery. Public opinion and heuristic processing of source cues. *Political Behavior*, [S. l.], v. 15, p. 167–192, 1993.

MONTEIRO, Rafael; SANTOS, Roney; PARDO, Thiago; ALMEIDA, Tiago; RUIZ, Evandro, OTO, Vale. *Contributions to the study of fake news in portuguese: new corpus and automatic detection results. Lecture Notes in Computer Science*. [S. l.], p. 324-334, 2019.

NEWMAN Mathew; PENNEBAKER James; BERRY Dianne; RICHARDS Jane. *Lying Words*: predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*. [S. l.], v. 29, n. 5, p. 665-675, 2002.

NEWMAN, Nic *et al.* *Digital News Report 2019*. Oxford: Reuters Institute, 2019.

NYHAN, Brendan. Facts and Myths about Misperceptions. *Journal of Economic Perspectives*. [S. l.], v. 34, n. 3, p. 220-36, 2020.

O’NEIL, Cathy. *Weapons of Math Destruction*. New York: Crown, 2016.

OLIVEIRA, Israel. *Jornalismo declaratório*. 1. ed. São Paulo: Editora Casa Flutuante, 2020.

OLIVEIRA, Lucy. *Análise de texto automatizada e análise de conteúdo: abordagens combinadas e apontamentos sobre a produção latino-americana*. Trabalho apresentado no X Congresso Latino-americano de Ciência Política, ALACIP, 2019

OTT, Myle; CHOI, Yejin; CARDIE, Claire; HANCOCK, Jeffrey. *Finding deceptive opinion spam by any stretch of the imagination*. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 309–319, Portland, Oregon, June 19-24, 2011

PARISER, Eli. *O Filtro Invisível*: o que a internet está escondendo de você. Rio de Janeiro, Zahar, 2012.

PAVARIN, Guilherme. Como a milícia digital bolsonarista resgatou sua máquina de fake news para atacar universitários. *The Intercept Brasil*. [S. l.], 2019. Disponível em: <https://cutt.ly/OAYx36Y>. Acesso em: 17 set. 2019.

PENNEBAKER, James; MEHL, Matthias; NIEDERHOFFER, Kate. Psychological aspects of natural language use: our words, our selves. *Annu. Rev. Psychol.* [S. l.], v. 54, p. 547–77, 2003.

PENNYCOOK, Gordon; CANNON, Tyrone; RAND, David. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology*. [S. l.], v. 147, n. 12, p. 1865–1880, 2018.

PENNYCOOK, Gordon; EPSTEIN, Ziv, MOSLEH, Mohsen *et al.* Shifting attention to accuracy can reduce misinformation *on-line*. *Nature*. [S. l.], n. 592, p. 590–595, 2021. Disponível em: <https://doi.org/10.1038/s41586-021-03344-2>. Acesso em: 22 abr. 2019.

PENNYCOOK, Gordon; RAND, David. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. [S. l.], 2018.

PENNYCOOK, Gordon; RAND, David. *The Psychology of fake news. Trends in Cognitive sciences.* [S. l.], v. 25, p. 388-402, 2021.

PINHEIRO-MACHADO, Rosana. *Amanhã vai ser maior: o que aconteceu com o Brasil e possíveis rotas de fuga para a crise atual.* São Paulo: Editora Planeta do Brasil, 2019.

PUGLISI, Maria Laura; FRANCO, Barbosa. *Análise de Conteúdo.* 2. ed. Rio de Janeiro: Liber Livro Editora, 2005.

RASHKIN, Hannah; CHOI, Eunsol; JANG, Jin Yea; VOLKOVA, Svitlana; CHOI, Yejin. *Truth of varying shades: analyzing language in fake news and political fact-checking.* Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2931–2937, Copenhague, Dinamarca, 2017.

RATINAUD, Pierre. IRAMUTEQ: Interface de R pour les analyses Multidimensionnelles de textes et de questionnaires [Computer Software]. *Iramuteq.* [S. l.], 2009. Disponível em: <http://www.iramuteq.org>. Acesso em: 08 abr. 2019.

RECUERO, Raquel; GRUZD, Anatoliy. Cascatas de fake news políticas: um estudo de caso no *Twitter*. *Galaxia*. São Paulo, n. 41, p. 31-47, 2019.

RECUERO, Raquel; SOARES, Felipe; ZAGO, Gabriela. Polarização, hiperpartidarismo e câmaras de eco: como circula a desinformação sobre COVID-19 no *Twitter*. *Contracampo*. Niterói, v. 40, n. 1, 2021.

REINERT, Max. Alceste, une méthodologie d'analyse des données textuelles et une application: aurelia de gerard de nerval. *Bulletin de Methodologie Sociologique.* [S. l.], v. 26, p. 24-54, 1990.

RESENDE, Gustavo; MELO, Philipe; REIS, Julio; VASCONCELOS, Marisa; ALMEIDA, Jussara; BENEVENUTO, Fabrício.. *Analyzing Textual (Mis)Information Shared in Whatsapp Groups.* WebSci '19: Proceedings of the 10th ACM Conference on Web Science June 2019 P. 225–234, 2019.

RESENDE, Gustavo; MESSIAS, Johnnatan; SILVA, Mário; ALMEIDA, Jussara; VASCONCELOS, Marisa; BENEVENUTO, Fabrício. A System for monitoring public political groups in *Whatsapp*. *WebMedia*. Salvador, v. 18, p. 16-19, 2018.

RICHARDSON, Roberto Jarry; PERES, José Augusto; WANDERLEY, José Carlos; CORREIA, Lindoya; PERES, Maria Holanda. *Pesquisa Social: métodos e técnicas.* 3. ed. São Paulo: Atlas, 2015.

SALVIATI, Maria Elisabeth. *Manual do aplicativo IRAMUTEQ.* versão 0.7, Alpha 2 e R versão 3.2.3). Planaltina: [S. l.], 2017.

SARMENTO, Leila Lauar; TUFANO, Douglas. *Português: literatura, gramática, produção de texto.* São Paulo: Moderna, 2004.

SOARES, Felipe Bonow; VIEGAS, Paula; BONOTO, Carolina; RECUERO, Raquel. Covid-19, desinformação e *Facebook*: circulação de URLs sobre a hidroxicloroquina em páginas e grupos públicos. *Galáxia*. São Paulo, n. 46, 2021.

- TABER, Charles; LODGE, Milton. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*. [S. l.], v. 50, p. 755–769, 2006.
- TANDOC Jr., Edson; LIM, Zheng Wei; LING, Richard. Defining “Fake News”. *Digital Journalism*. [S. l.], v. 6, n. 2, p. 137-153, 2018.
- UNDEUTSCH, Udo. *Forensische psychologie*. Gottingen: Verlag fur Psychologie, 1967.
- VAN BAVEL, Jay; PEREIRA, Andrea. *The Partisan brain: an identity-based model of political belief*. *Trends in Cognitive Sciences*. [S. l.], v. 22, p. 213-224, 2018.
- VOUSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news *on-line*. *Science*. [S. l.], n. 359, p. 1146-1151, 2018.
- VRIJ, Aldert. *Detecting lies and deceit: the psychology of lying and the implications for professional practice*. 2. ed., Chichester, Reino Unido: Wiley, 2008.
- WARDLE, Claire. The Need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*. [S. l.], v. 6, n. 8, p. 951–963, 2018.
- ZHAO, Zilong; ZHAO, Jichang; SANO, Yukie *et. al*. Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*. [S. l.], 2020. Disponível em: <https://cutt.ly/2AYmhSQ>. Acesso em: 7 jun. 2019.

APÊNDICE A – Análise estatística exploratória (notícias com temática política)

Uma das vantagens do *software* Iramuteq é a seleção dos chamados *termos ativos*⁶⁸. Palavras que carregam mais conteúdo que os *termos suplementares*, é sobre eles que a maioria das análises automatizadas são feitas. A partir da contagem de palavras do banco de dados é possível ter uma noção de quais os temas estavam sendo abordados durante o período em que as notícias foram colhidas e quais atores e instituições políticas eram figuras centrais do noticiário. A seguir observa-se uma tabela com os vinte termos ativos mais citados no banco de dados, o que permite uma visualização – ainda que superficial – dos temas mais tratados no período em que as notícias foram colhidas⁶⁹.

Tabela 7 – Termos ativos mais citados entre as notícias falsas e verdadeiras do banco de dados com política

Notícias Falsas			Notícias verdadeiras		
Termo	Nº de presenças	Categoria gram.	Termo	Nº de presenças	Categoria gram.
dizer	433	ver	dizer	2338	ver
lula	283	nom	estar	2277	ver
presidente	229	nom	presidente	1638	nom
michel_temer	203	nr ⁷⁰	lula	1376	nom
país	196	nom	afirmar	1108	ver
eua	176	nr	ano	1050	nom
ministro	162	nom	governo	1020	nom
nuclear	138	adj	michel_temer	1000	nr
norte	136	nom	país	948	nom
brasil	128	nom	público	816	nom
governo	127	nom	federal	794	adj
trump	121	nr	ministro	752	nom
coreia	118	nr	ex_presidente	725	nr
sérgio_moro	113	nr	brasil	692	nom

⁶⁸ PENNEBAKER *et al.* (2003) utilizam o termo *palavras de conteúdo e partículas* para se referir a essa divisão. O Iramuteq divide os termos entre *ativos* e *suplementares*. O *software* sugestiona ao pesquisador as categorias nomes (substantivos), pronomes, verbos, adjetivos, palavras não reconhecidas e advérbios como termos ativos, no entanto seguimos a recomendação de Camargo e Justo (2016) e retiramos os advérbios dessa categoria.

⁶⁹ A sigla *nom*, presente nas colunas referentes às categorias gramaticais, se refere aos substantivos.

⁷⁰ A sigla *nr* se refere aos termos não reconhecidos, cujo software não conseguiu classificar em uma das categorias gramaticais.

Notícias Falsas			Notícias verdadeiras		
Termo	Nº de presenças	Categoria gram.	Termo	Nº de presenças	Categoria gram.
ano	113	nom	lava_jato	674	nr
rcifrão ⁷¹	112	nr	dia	669	nom
míssil	110	nom	caso	652	nom
dilma	109	nr	pessoa	643	nom
kim	107	nom	stf	625	nr
público	106	adj	rcifrão	616	nr

Fonte: produção do autor, 2022.

Observa-se, portanto, a presença massiva de termos referentes a atores (Lula e Michel Temer), instituições (STF) e cargos políticos (Presidente, Ministro), bem como o nome de países (Brasil e Coreia) e operações policiais (Lava-Jato). Lembrando que a maioria das notícias colhidas são do segundo semestre de 2017, momento em que o julgamento de recursos judiciais do ex-presidente Luiz Inácio Lula da Silva gerava grandes debates, bem como os avanços da operação Lava-Jato, os projetos políticos e escândalos de corrupção ligados ao governo Temer. A crise diplomática entre a Coreia do Norte e Estados Unidos foi o principal tema de destaque relacionado à política internacional no período, o nome da península asiática foi citado mais de 700 vezes.

⁷¹ O termo *rcifrão* é referente à sigla R\$, que foi modificada durante o processo de tratamento dos dados.

APÊNDICE B – Análise estatística exploratória (notícias sem temática política)

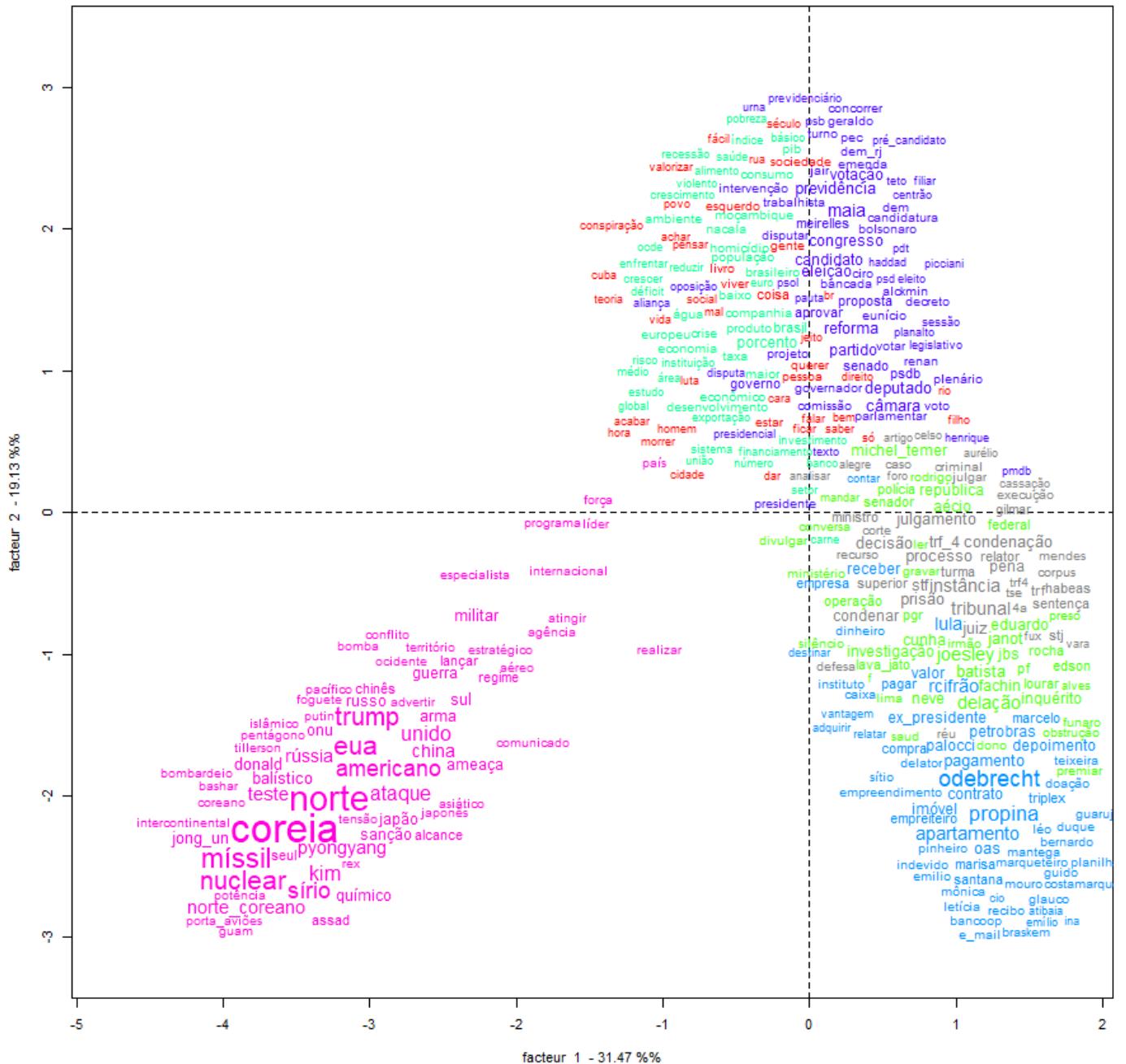
A tabela a seguir demonstra a falta de termos referentes a temáticas mais específicas, como havia no banco de dados principal. Os resultados obtidos validam a seleção realizada durante o levantamento, já que não há a presença de termos claramente ligados ao noticiário político.

Tabela 8 – Termos ativos mais citados nas notícias falsas e verdadeiras

Notícias Falsas			Notícias verdadeiras		
Termo	Nº de presenças	Categoria gram.	Termo	Nº de presenças	Categoria gram.
ano	46	nom	ano	275	nom
dizer	43	ver	dizer	237	ver
pessoa	35	nom	novo	142	adj
ficar	22	ver	brasil	142	nom
médico	21	nom	mesmo	131	adj
homem	20	nom	dia	122	nom
brasil	20	nom	família	116	nom
social	19	adj	dar	110	ver
dar	19	ver	ficar	105	ver
rede	18	nom	samba	98	nom
filho	18	nom	país	96	nom
vida	17	nom	vida	94	nom
passar	17	ver	passar	94	ver
local	17	nom	querer	91	ver
cidade	17	nom	vez	89	nom
publicar	16	ver	grande	88	adj
mulher	16	nom	empresa	88	nom
jornal	16	nom	contar	85	ver
conseguir	16	ver	filho	83	nom
vídeo	15	nom	polícia	82	nom

Fonte: produção do autor, 2022.

APÊNDICE C – AFC do banco de dados geral



Classe 1: Vermelho

Classe 2: Cinza

Classe 3: Verde

Classe 4: Azul Claro

Classe 5: Azul Escuro

Classe 6: Rosa

APÊNDICE D - Análise exploratória das notícias verdadeiras via CHD

O noticiário enquadrado como verdadeiro pelo banco de dados Fake.Br (MONTEIRO *et al.*, 2019) colhido nessa pesquisa possui três grandes eixos temáticos enquadrados em sete classes. O primeiro eixo aborda a crise diplomática entre a Coreia do Norte e os Estados Unidos, representado pela classe 7, que reúne 12,3% dos segmentos textuais que foram classificados. Os termos centrais da categoria foram palavras ligadas à tensão que ocorreu entre os diversos países da região e os EUA (*Coreia, Norte, Míssil, Trump, EUA, Nuclear, Americano, Ataque, Kim, Teste, Pyongyang, China e Rússia*). Mas nota-se que nessa categoria também se enquadram outros temas ligados ao noticiário internacional como a Guerra da Síria (*Sírio, ONU e Assad*), que provavelmente eram tratados nas mesmas matérias ou contextos que o evento do mar da Coreia. Os trechos a seguir foram retirados de duas notícias cujo conteúdo está completamente inserido nesta classe pela CHD e demonstram a forma como a mídia tratava as tensões entre Pyongyang e Washington:

(txt_1325) As 4 opções não militares para enfrentar o desafio da Coreia do Norte. Recorrer à força contra Kim Jong_un é uma decisão perigosa, mas as opções não_militares também apresentam inconvénientes. O último teste nuclear anunciado pela Coreia do Norte no domingo passado voltou a alarmar não apenas Washington mas também países vizinhos na Ásia. Depois de os norte_coreanos terem anunciado que acionaram uma bomba que, segundo a mídia oficial de Pyongyang, seria de hidrogênio e poderia ser transportada por um míssil de longo alcance, o presidente dos EUA, Donald Trump, convocou assessores militares para discutir as opções disponíveis. (MONTEIRO *et al.*, 2019)

(txt_1376) Seul treina time para decapitar ditador vizinho. Com tropa de elite treinada para matar Kim Jong_un, Seul quer pelo menos assustar ditador. Na última vez que se soube que a Coreia do Sul havia conspirado para assassinar a liderança norte_coreana, nada saiu como planejado. (MONTEIRO *et al.*, 2019)

Outro eixo temático envolve o dia a dia da política, seja a institucional (classe 6) ou a mundana (classes 1 e 4). A classe 6 reúne 11,6% dos segmentos classificados e tem como principais termos nomes ligados a lideranças políticas de vários espectros da direita brasileira (*Maia, Meirelles, Bolsonaro, Michel Temer*), instituições (*Congresso, Eleição, Partido, Câmara, Governo, Senado*) e processos legislativos (*Reforma, Previdência, Eleição, Votação, Aprovar e Proposta*). Os textos a seguir representam bem a temática abordada nesta classe e demonstram como as tramitações de projetos de reformas e as negociações em torno do parlamento são peça central das matérias:

(txt_1064) Michel_Temer acha que é Itamar, mas é Sarney, diz cientista político. Para Marcos Nobre, mais do que o arquivamento da denúncia contra Michel_Temer, votação na Câmara significou cisão incontornável do PSDB e ascensão de Aécio como o novo Cunha.. O resultado da votação na Câmara nesta quarta_feira tem efeitos muito

mais amplos do que o arquivamento da denúncia por corrupção passiva contra o presidente Michel_Temer. A afirmação é do cientista político da Unicamp Marcos Nobre, um especialista em PMDB. (MONTEIRO *et al.*, 2019)

(txt_1129) Congresso corre contra o tempo para aprovar reforma política; veja propostas. Parlamentares têm até 7 de outubro para aprovar mudanças a fim de que novas regras entrem em vigor já em 2018. Entre os pontos, está a criação de fundo público de cerca de R\$ 3,6 bilhões.. Em meio à instabilidade e à divisão do Congresso em relação ao governo Michel_Temer, um tema une a classe política: a necessidade de uma reforma nas regras para as próximas eleições, principalmente, as de 2018. (MONTEIRO *et al.*, 2019)

Enquanto isso, a classe 1 agrupa 20,2% dos segmentos, e tem como palavras base termos ligados a pautas sociais da política, remetendo recorrentemente a palavras ligadas à ideia de coletividade (*Sociedade, Gente, Brasil, Brasileiro*), a questões da vida diária (*Violência, Rua, Mulher, Homem, Vida, Viver*) e mesmos ideologias políticas (*Movimento, Social, Esquerdo(a) e Direito(a)*). Os exemplos abaixo são explicativos, o primeiro tem como assunto principal as críticas políticas do músico Marcelo Yuka, enquanto o segundo reflete sobre intolerância presente na sociedade brasileira:

(txt_1033) Marcelo Yuka acredita que ser utópico é um direito seu. Então pinta um mundo que considera melhor que o que temos, sem Donald Trump e Jair Bolsonaro. Falando em um momento sem qualquer traço ideológico, ele cita os dois para explicar por que seu primeiro disco solo, lançado na sexta (6), se chama Canções para depois do ódio. (MONTEIRO *et al.*, 2019)

(txt_1224) A intolerância aumentou no Brasil ou sempre fomos um país assim? Nem lá nem cá. O Brasil também tem uma forte tradição pacifista. Vamos falar do fenômeno do Uber. O Uber é uma inovação tecnológica excelente para a sociedade: diminui o CO₂, aumenta o convívio com as pessoas, é uma economia. Só que essa inovação desfaz um território estável, que era o do taxista. Então, essa profissão antiga se vê, de uma hora para outra, fadada à destruição. (MONTEIRO *et al.*, 2019)

Já a classe 4, que representa 14,52% do *corpus* e interage fortemente com a 1, aborda fundamentalmente questões econômicas da política diária, tratando desde problemas relacionados ao orçamento e crescimento do país (*Crise, Econômico, Economia, Produto, Desenvolvimento, PIB, Exportação e Orçamentário*) a questões ligadas ao consumidor e preço de produtos da cesta básica (*Consumo, Água, Taxa, Salário, Carne, Financiamento e Emprego*). Os trechos a seguir exemplificam a abordagem dessa classe:

(txt_1413) Estamos mais vulneráveis do que em 2008, diz ex_presidente do BCE. Para Jean_Claude Trichet, é inquietante ver o endividamento da economia global seguir no mesmo ritmo pré_crise. PARIS _ Em 2017, a economia da zona do euro, o grupo de 19 países da União Europeia que adota a moeda única, registrou seu melhor desempenho em 10 anos, com um crescimento de 2,5 porcento de seu Produto Interno Bruto (PIB). (MONTEIRO *et al.*, 2019)

(txt_1456) Crise hídrica e racionamento alertam para fim da cultura da abundância, dizem especialistas; leia entrevista. Rodízio no fornecimento de água completa um ano na próxima terça; desafio é ampliar oferta e seguir reduzindo demanda, dizem professores. (MONTEIRO *et al.*, 2019)

Por fim, o último eixo temático girou em torno dos escândalos de corrupção da época, as denúncias realizadas e os julgamentos dos tribunais, reunindo três classes (2, 3 e 5) da CHD. Duas dessas classes tiveram como termos centrais palavras ligadas a denúncias de desvios de dinheiro público e pagamento de propina, no entanto, a diferença das classes 3 e 5 foram os escândalos relatados. A classe 3, que reuniu 13,8% dos segmentos textuais, abordou as denúncias de corrupção envolvendo empresas e empresários do ramo alimentício (*Joesley, Batista, JBS*), membros do PMDB (*Cunha, Eduardo, Michel Temer*), PSDB (*Aécio, Neve*) e de órgãos de investigação (*PF, Polícia, Inquérito*) e acusação (*Janot, PGR*). Os trechos selecionados abaixo exemplificam essa abordagem, principalmente no que se refere aos escândalos ligados à JBS e os irmãos Batista:

(txt_1092) PF faz busca e apreensão nas casas de Joesley e Saud em SP. Eles foram presos neste domingo na capital paulista e devem ser transferidos para Brasília nesta segunda. PF também está na casa do procurador Marcello Miller, no Rio. . A Polícia Federal cumpre na manhã desta segunda-feira (11), na capital paulista, quatro mandados de busca e apreensão em imóveis relacionados a Joesley Batista, um dos donos da J&F, e ao executivo da empresa Ricardo Saud, presos neste domingo (10). (MONTEIRO *et al.*, 2019)

(txt_1462) Controladoria_Geral da União (CGU), o ministro Osmar Serraglio (PMDB_PR) foi citado e flagrado em grampo da Operação Carne Fraca , que apura um esquema de pagamento de propinas envolvendo frigoríficos e fiscais do Ministério da Agricultura. A frente da CGU, uma das principais atribuições de Serraglio será o combate à corrupção na administração pública federal. A pasta é responsável no governo federal por ações de prevenção e combate à corrupção, auditoria pública e ampliação da transparência da gestão pública. Osmar Serraglio deve deixar a chefia do Ministério da Justiça para assumir a cadeira de ministro da Transparência. (MONTEIRO *et al.*, 2019)

Enquanto a classe 5, representante de 15,78% dos segmentos, estava voltada para os escândalos envolvendo empresas e empresários do setor da construção civil (*Odebrecht, OAS, Marcelo, Empreiteiro*) e de combustíveis (*Petrobras*), e lideranças do Partido dos Trabalhadores (*Lula, ex-presidente, Palocci*). É possível verificar nos fragmentos abaixo a centralidade das acusações contra o ex-presidente Lula nesta classe:

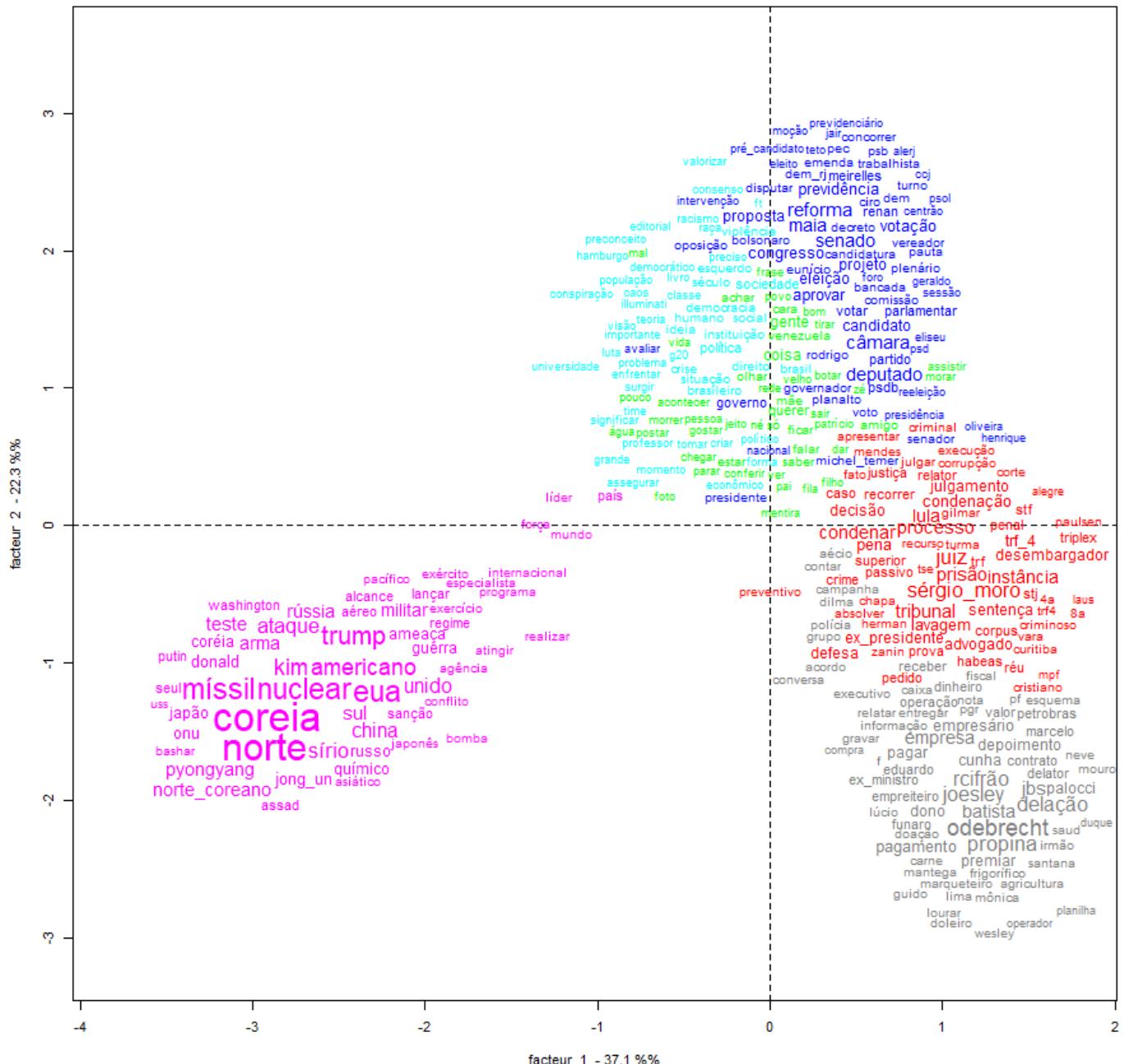
(txt_1079) Lula é condenado na lava_jato a 9 anos e 6 meses de prisão no caso do triplex. Na sentença, juiz Sérgio_Moro cita documentos e depoimentos que comprovam que apartamento no litoral de SP era destinado ao ex_presidente, diz que há provas documentais e que Lula faltou com a verdade. Leia trechos. É a primeira vez, na história, que um ex_presidente é condenado por um crime comum no Brasil. A decisão permite que Lula recorra em liberdade. (MONTEIRO *et al.*, 2019)

(txt_1333) Delator diz que ajudou advogado de Lula a ocultar que Odebrecht executou reforma de sítio em Atibaia. Delator diz que comprou cofre para guardar dinheiro da obra na propriedade frequentada pela família de Lula; Instituto Lula afirmou que ex_presidente não é dono do sítio e que não pediu nem autorizou ninguém a pedir que fossem feitas reformas no imóvel. (MONTEIRO *et al.*, 2019)

Esse eixo também reúne a classe 2, representante de 14% dos segmentos textuais, mais voltada para questões envolvendo o judiciário (*Instância, Tribunal, Juiz, STF, TRF-4, Desembargador e Sérgio Moro*) e atos processuais (*Processo, Prisão, Condenação, Decisão, Julgamento, Pena e Condenar*). Um ponto interessante é a presença de *Lula*, de forma significativa, também nessa classe, o que demonstra a centralidade do ex-presidente no noticiário da época. A análise fatorial por correspondência (Apêndice D) permite a visualização de como essa classe interliga os eixos que envolvem a política nacional. Os dois trechos de notícias selecionados apontam para a importância das instituições do judiciário nela:

(txt_1103) STJ nega habeas corpus preventivo por unanimidade e decide que Lula pode ser preso após segunda instância. Pedido da defesa tinha objetivo de evitar execução – depois de esgotados os recursos – da sentença do TRF da 4a Região, que condenou ex_presidente a 12 anos de prisão. **(txt_1256)** TRF_4 nega recurso da defesa de Lula sobre condenação em segunda instância no processo do triplex. Por decisão do próprio TRF_4, Lula pode ser preso para começar a cumprir a pena quando acabarem os recursos no tribunal. Porém, uma decisão provisória do STF impede a prisão até que o plenário da Corte julgue pedido de habeas corpus. (MONTEIRO *et al.*, 2019)

APÊNDICE E – Análise Fatorial por Correspondência (Notícias Verdadeiras)



Classe 1: Vermelho

Classe 2: Cinza

Classe 3: Verde

Classe 4: Azul Claro

Classe 5: Azul

Classe 6: Azul Escuro

Classe 7: Rosa

APÊNDICE F – Análise exploratória das notícias falsas via CHD

O primeiro eixo (classe 4) aborda a crise diplomática entre Coreia do Norte e Estados Unidos, da mesma forma que nas notícias verdadeiras. Há, contudo, uma presença maior desse assunto nesse *corpus*, como já indicava a análise de especificidades, a CHD indicou que 25,59% dos segmentos classificados se enquadram neste grupo. Os termos centrais desse eixo também foram bastante similares aos encontrados nas notícias verdadeiras (*Norte, Nuclear, EUA, Coreia, Trump, China...*), todavia, vale a pena destacar a presença das palavras escritas incorretamente *Coréia* e *Jon-un*⁷², que, não apenas tiveram grande concentração entre as notícias falsas, como também foram termos centrais nessa categoria. As passagens selecionadas são exemplos de como os produtores de desinformação trataram o tema, levantando as pautas de catástrofe biológica e nuclear:

(txt_1342) Ataque biológico? Coreia do Norte descobre pandemia de vírus mortal e recusa ajuda. País de Kim Jon_un luta contra um surto viral com o potencial de devastar a humanidade? e não permite ajuda externa. De acordo com o britânico Daily Star, o país norte_coreano estaria lutando contra uma terrível gripe aviária. Ainda não foi descartada a tese de um suposto ataque biológico. (MONTEIRO *et al.*, 2019)

(txt_1316) Míssil ICBM da Coreia pode atingir o Havaí e colocar mais de 1 milhão de americanos em risco. Especialistas em armamentos nucleares contratados pela Coreia do Norte relataram ao ditador Kim Jong_un que os mísseis balísticos fabricados no país do ditador já possuem autonomia suficiente para atingir o Havaí. Os oficiais acreditam que se o estado norte_americano for atingido por um ICBM (míssil nuclear balístico para longas distâncias), o armamento teria a mesma força destrutiva que as bombas atômicas que atingiram Hiroshima e Nagasaki no fim da segunda guerra mundial. (MONTEIRO *et al.*, 2019)

O segundo eixo temático, da mesma forma que na CHD das notícias verdadeiras, reúne palavras ligadas à política brasileira institucional (classe 1). A classe 1 representa 35,6% dos segmentos textuais e tem nomes de políticos que eram titulares de cargos importantes (*Michel Temer, [Rodrigo] Maia, [Renan] Calheiros, Dória*), de instituições políticas (*Senado, Câmara, Governo, Prefeito*), questões econômicas (*Crise, Caro, Emprego*) e de pautas legislativas da época (*Previdência, Votação, Votar, Parlamentar*). Nota-se, contudo, uma abordagem mais sensacionalista do tema, diferentemente do que é visto nas notícias verdadeiras.

(txt_1169) Sítio de Atibaia poderá servir de abrigo para ex-moradores da Cracolândia diz site. De acordo com o portal Juntos pelo Brasil, uma ideia inovadora foi ventilada nos bastidores da política paulistana. Após a remoção de centenas de pessoas da Cracolândia, os dependentes químicos ficaram sem local para se abrigar. O Prefeito João Dória, em uma decisão que dividiu opiniões entre especialistas, tomou uma atitude energética (considerada arbitrária para alguns e correta para outros) e removeu

⁷² O termo *Coréia* foi citado 92 vezes nas notícias falsas analisadas.

os usuários de crack do centro de São Paulo. A Secretaria de Direitos Humanos de SP chegou a declarar que a atitude do prefeito foi desumana e colocou o cargo à disposição. Será que o problema deve ser simplesmente varrido para debaixo do tapete? Basta remover as pessoas das ruas ou esses dependentes precisam de tratamento e acompanhamento psicológico? O famoso sítio de Atibaia, que até agora não tem pai nem mãe, foi cogitado para acolher e tratar essas pessoas que foram removidas da Cracolândia, informou o portal Juntos pelo Brasil. (MONTEIRO *et al.*, 2019)

(txt_1136) Malafaia sobre deputada que pediu derramamento de sangue: Petralhada! Mesmo discurso do Estado Islâmico. O polêmico Silas Malafaia não perdoou a fala infeliz e carregada de ódio da deputada Benedita da Silva (PT_RJ). Dessa vez temos que tirar o chapéu para o pastor. Durante um seminário do PT, Benedita pediu derramamento de sangue contra o golpe. O discurso dessa senhora, que citou a Bíblia para justificar um futuro derramamento de sangue e redenção no Brasil, deixou muitos perplexos, inclusive o senador petista peemedebista Roberto Requião, que estava ao lado da parlamentar. (MONTEIRO *et al.*, 2019)

Por fim, o terceiro eixo temático reserva as maiores disparidades entre o banco de dados de notícias falsas e verdadeiras, abordando assuntos ligados ao poder judiciário (classe 3) e a escândalos de corrupção (classe 2). As palavras mais significativas da classe 3 – que reúne 18,4% dos segmentos textuais – se relacionam de alguma forma ao poder judiciário (*Sérgio Moro, Juiz, STF, Fachin, Advogado, Edson, Ministro, Justiça*), também sendo dada importância ao nome do ex-presidente *Lula* nessa classe. As passagens a seguir abordam as investigações de escândalos de corrupção ligados aos ex-presidentes Michel Temer e Lula.

(txt_1068) Família diz que homem da mala está sendo torturado para entregar Michel_Temer: Ele morre mas não delata. . Atendendo um pedido da defesa de Rodrigo Rocha Loures, o ministro Edson Fachin determinou (no último dia 07) que o ex_deputado fosse transferido da penitenciária da Papuda para uma sala nas dependências da Polícia Federal em Brasília. Os advogados alegaram que Rocha Loures corria sério risco de morte. (MONTEIRO *et al.*, 2019)

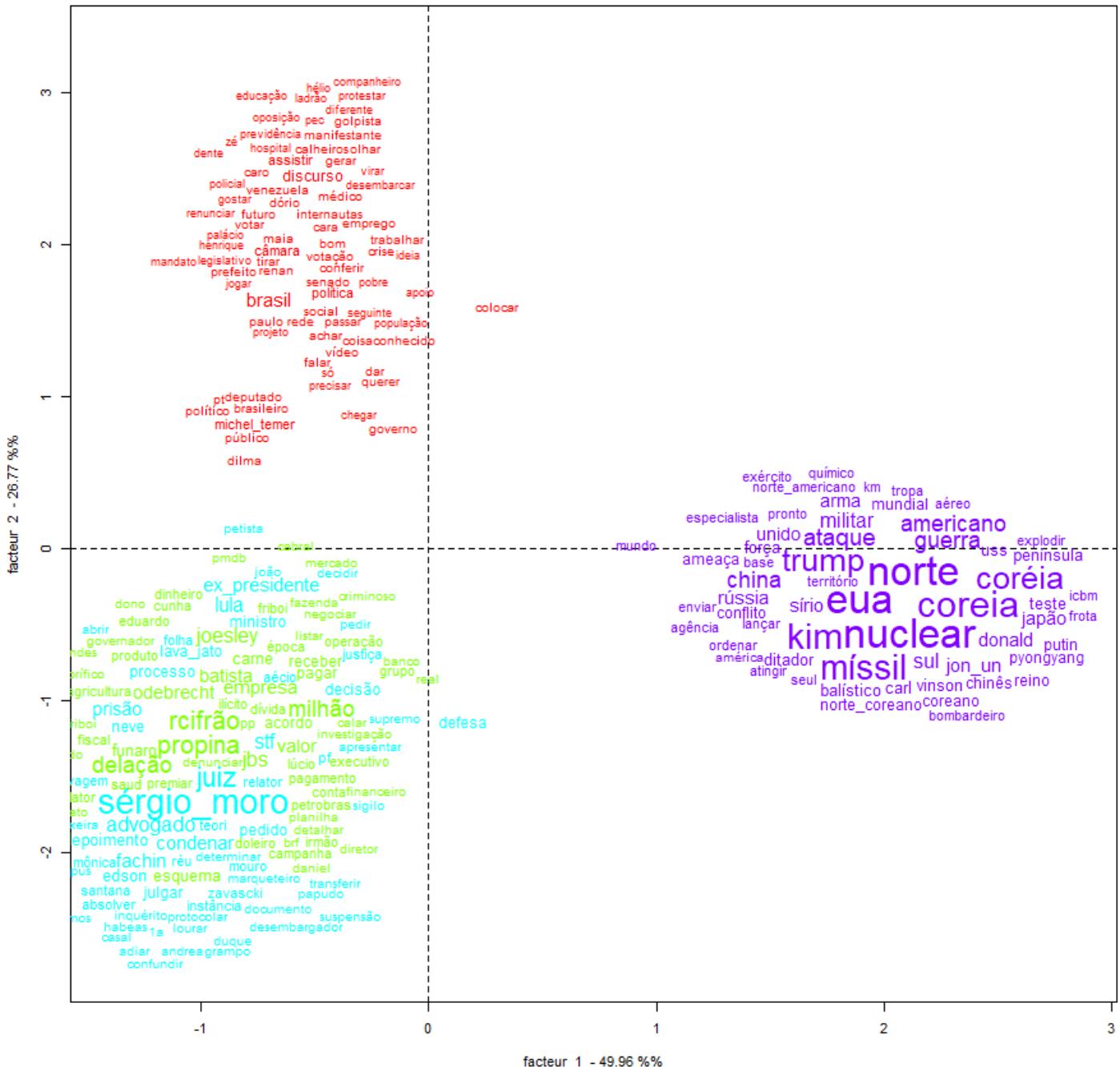
(txt_1091) Lula perde mais uma: Desembargador nega suspensão do processo do Triplex. O TRF da 4a Região (através da 8a turma) julgou o mérito de mais um recurso impetrado pelos advogados do ex_presidente Lula. A decisão, que já havia sido julgada anteriormente pelo mesmo tribunal, foi mantida e o processo seguirá normalmente. O primeiro pedido de habeas corpus foi protocolado em maio (um dia antes da audiência de Lula com o juiz Sérgio_Moro) e pedia a suspensão imediata da ação relacionada ao triplex do Guarujá. (MONTEIRO *et al.*, 2019)

Já a classe 2 representa 20,36% do *corpus* e tem como termos centrais os nomes de empresas e empresários ligados a escândalos de corrupção da época (*Odebrecht, JBS, Empresa, Joesley, Carne, Batista*), bem como palavras referentes aos crimes (*Rcifrão [R\$], Propina, Milhão, Dinheiro, Esquema*) e ao processo investigativo (*Delação e Operação*). Os excertos selecionados indicam esse formato de denúncia de parte das desinformações.

(txt_1442) Procuradoria_Geral da Fazenda indica que 2016 amargou mais de 1 trilhão de reais em sonegação fiscal. O Sindicato Nacional dos Procuradores da Fazenda (Sinprofaz) estima que as perdas com a sonegação somente em 2016 chegaram a R\$ 500 bilhões. De acordo com uma lista divulgada pelo Ministério da Fazenda, a soma das dívidas dos 500 maiores devedores ultrapassava os R\$ 400 bilhões no ano passado. Isso somente com os 500 maiores devedores! A maior devedora da Fazenda hoje é a mineradora Vale, privatizada em 1997, durante o governo FHC... (MONTEIRO *et al.*, 2019)

(txt_1223) Diretor da JBS confessa que pagou R\$ 600 milhões em propina. Em sua delação premiada, o diretor da JBS Ricardo Saud, detalhou a trajetória dos crimes da JBS em números. Como se fosse a coisa mais natural do mundo, Saud disse aos procuradores que pagou propinas (que totalizam 600 milhões de reais) para 1.829 candidatos de 28 partidos... (MONTEIRO *et al.*, 2019)

APÊNDICE G – Análise Fatorial por Correspondência (Notícia Falsa)



Classe 1: Vermelho

Classe 2: Verde

Classe 3: Azul Claro

Classe 4: Roxo

APÊNDICE H – Fases da coleta e tratamento dos dados no Iramuteq

1. Os dados foram coletados da base Fake.Br, que é pública e está disponibilizada na plataforma GitHub (<https://github.com/roneysco/Fake.br-Corpus>). Todos os textos são classificados numericamente pelos mantenedores do banco. Foram escolhidas as notícias a partir do número 1000, o que facilitou a coleta dos dados.
2. O conteúdo dos textos foi analisado a partir de sua temática geral. Caso houvesse alguma abordagem a pautas ou figuras ligadas à política nacional e internacional, o texto era selecionado para o banco de dados da pesquisa.
3. O banco de dados da pesquisa foi construído no *software* LibreOffice. Esse *software* é mais adaptável ao Iramuteq (que não reconhece textos editados no Microsoft Office, ou em formato doc., demandando a utilização de arquivos em txt.).
4. A numeração dada pelo banco de dados original foi mantida para termos de identificação, revisão e replicação. As bases disponibilizadas no GitHub (https://github.com/ulissesgdm/Dados_Dissertacao_Ulisses_Melo) identificam cada texto com a numeração original dada pelo Fake.Br.
5. A partir do momento em que as 800 notícias sobre política (400 verdadeiras e 400 falsas) foram colhidas e transferidas para o arquivo txt. o banco da pesquisa foi fechado e passou a ser editado de acordo com as exigências de leitura do Iramuteq (CAMARGO; JUSTO, 2016).
6. Foram retirados os seguintes símbolos do banco: (...), ([...]), (‘) e (‘‘).
7. Os seguintes símbolos foram substituídos: (-) por (_) e os espaços duplos por espaços simples.
8. Os seguintes símbolos foram substituídos por termos: (%) por (porcento) e (\$) por (cifrão).
9. Posteriormente foi realizada a modificação de alguns nomes importantes para o banco de dados que não estavam sendo reconhecidos.
10. Os nomes “Sérgio Moro” – quando próximos por um espaçamento – foram unificados no termo “Sérgio Moro”, já que o software estava reconhecendo a palavra “Moro” como termo ligado ao verbo “morar”. O termo isolado “Moro” também foi substituído por “Sérgio Moro”, já que se observou que quase a totalidade das citações eram referentes à figura do então juiz da Lava Jato.
11. Da mesma forma, os nomes “Michel Temer” e “Temer” foram substituídos por “Michel Temer”, já que a palavra *temer* estava sendo entendida como verbo.

12. O nome “Luiz Inácio Lula da Silva” foi substituído por “Lula”, já que a citação do nome completo do ex-presidente poderia afetar na CHD do termo “Lula”.
13. As notícias que não entraram no banco sobre política foram registradas são adicionadas em outro banco. A partir do momento em que o banco não-político atingiu a marca de 140 textos (70 verdadeiros e 70 falsos) ele foi fechado e passou pelo mesmo tratamento dos dados que envolviam política (exceto as alterações dos nomes de lideranças políticas).

APÊNDICE I – Tratamento dos dados e aplicações de testes no ALT (legibilidade.com)

1. Os dados colhidos para os testes de legibilidade partem do mesmo ponto dos que foram levantados para as análises no Iramuteq, partindo do texto de número 1000 do Fake.Br;
2. Foram selecionados 400 textos com conteúdo político e 140 que não abordavam esse tema;
3. Não houve nenhum outro tipo de tratamento que alterasse o conteúdo, já que a plataforma ALT reconhece todos os tipos de caracteres;
4. Os testes foram realizados a partir da subdivisão do banco. Eram selecionados 10 textos de cada vez, os resultados eram colhidos e inseridos em uma tabela e posteriormente plotados. Essa subdivisão do banco ocorreu devido à problemas ligados ao ALT, que não conseguia processar muitos textos de uma vez (principalmente do banco de notícias verdadeiras, que era bem maior que o de informações falsas);
5. Esse tratamento, contudo, permitiu uma observação mais detalhada, diminuindo a interferência de possíveis outliers;

APÊNDICE J – Termos ligados ao noticiário político selecionados para criação de *subcorpus*

Tabela 9 – Termos selecionados para a criação dos subcorpus de teste, e o respectivo número de citações no banco de dados

Termo	Nº de citações
Presidente	1868
Lula	1662
Michel Temer	1203
Governo	1147
País	1145
Público	922
Ministro	918
Ex-presidente	822
Brasil	821
Lava-Jato	765
R\$ (Rcifrão)	728
Coreia	711
STF	709
Político	626
Ministério	599

Fonte: produção do autor, 2022.