

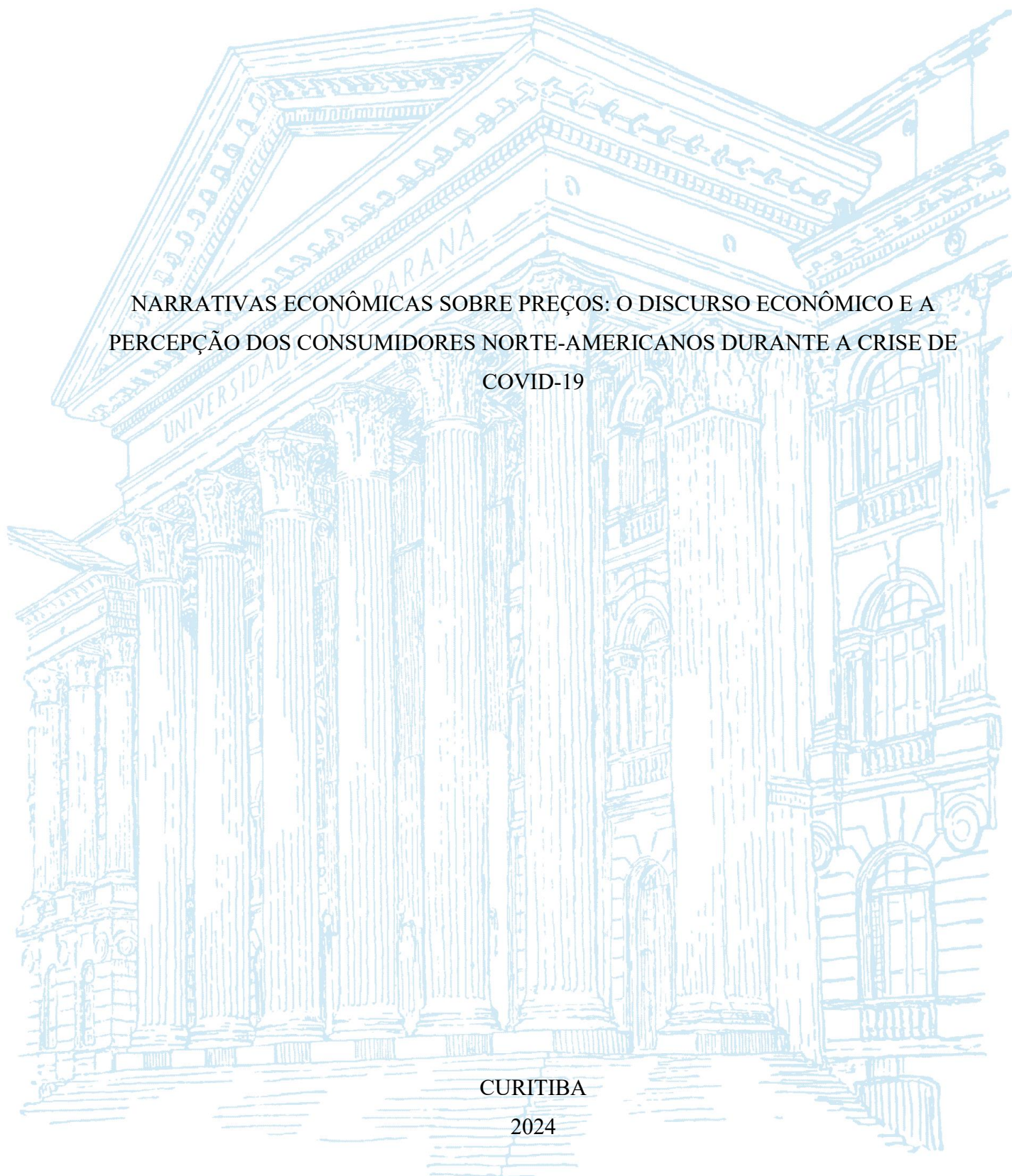
UNIVERSIDADE FEDERAL DO PARANÁ

GLEDSON GUSTAVO VIAN DA SILVA

NARRATIVAS ECONÔMICAS SOBRE PREÇOS: O DISCURSO ECONÔMICO E A
PERCEPÇÃO DOS CONSUMIDORES NORTE-AMERICANOS DURANTE A CRISE DE
COVID-19

CURITIBA

2024



GLEDSOON GUSTAVO VIAN DA SILVA

NARRATIVAS ECONÔMICAS SOBRE PREÇOS: O DISCURSO ECONÔMICO E A
PERCEPÇÃO DOS CONSUMIDORES NORTE-AMERICANOS DURANTE A CRISE DE
COVID-19

Trabalho de Conclusão de Curso apresentado ao curso de Graduação em Ciências Econômicas, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas.

Orientador(a): Prof(a). Dr(a). Adriana Sbicca Fernandes

CURITIBA

2024

TERMO DE APROVAÇÃO

GLEDSON GUSTAVO VIAN DA SILVA

NARRATIVAS ECONÔMICAS SOBRE PREÇOS: O DISCURSO ECONÔMICO E A PERCEPÇÃO DOS CONSUMIDORES NORTE-AMERICANOS DURANTE A CRISE DE COVID-19

Trabalho de Conclusão de Curso apresentado ao curso de Graduação em Ciências Econômicas, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas.

Prof(a). Dr(a) Adriana Sbicca Fernandes

Orientadora – Departamento de Economia, UNIVERSIDADE FEDERAL DO
PARANÁ (UFPR)

Prof. Dr. Leonardo Matsuno da Frota

Departamento de Economia, UNIVERSIDADE FEDERAL DO PARANÁ (UFPR)

Prof. Dr. José Guilherme Silva Vieira

Departamento de Economia, UNIVERSIDADE FEDERAL DO PARANÁ (UFPR)

Cidade, 03 de dezembro de 2024.

Para minha mãe, cuja dedicação incondicional ao longo da minha vida foi fundamental para que esse trabalho pudesse se tornar realidade.

“As estórias que as pessoas contam, também são estórias sobre como a economia se comporta.” (Akerlof; Shiller, 2010, p. X, tradução nossa).

RESUMO

As narrativas econômicas são histórias que moldam o comportamento e as percepções das pessoas, influenciando as expectativas dos consumidores e direcionando as discussões econômicas, especialmente em momentos de crise, como durante a COVID-19. O objetivo de estudar narrativas relacionadas a preços durante períodos de crise é compreender se os discursos dominantes refletem as preocupações dos consumidores, pois a literatura sustenta que os preços de alimentos têm um maior impacto sobre as expectativas dos consumidores. Nesse contexto, é fundamental avaliar se as principais notícias sobre a variação de preços de determinados bens de consumo estão alinhadas com o que a teoria empírica econômica tradicional relata, uma vez que essas informações influenciam seu comportamento e geram consequências reais na economia. Seguindo esta premissa, este trabalho realizou uma análise textual de artigos de jornais norte-americanos publicados entre dezembro de 2019 e julho de 2020, com foco nas narrativas econômicas relacionadas a preços. Entre os temas mais recorrentes identificados, o preço do petróleo foi o que mais se destacou nos artigos analisados. A ênfase dessa narrativa econômica ocorreu devido à queda dos preços do petróleo, que foi majoritariamente retratada de forma negativa, como foi observado na análise de sentimento textual desenvolvida. Em contrapartida, a narrativa sobre os preços do petróleo não refletia a preocupação mais urgente dos consumidores, pois negligencia o efeito do aumento do poder de compra do consumidor, consequência da queda de preços. Constatou-se que a ênfase nos choques dos preços do petróleo ofuscou a discussão sobre o aumento significativo nos preços dos alimentos. Estudos do Centro de Pesquisa do Consumidor de Michigan, nos Estados Unidos, corroboram que os preços de alimentos influenciam mais as percepções dos consumidores do que os preços do petróleo. Dessa forma, é importante que tanto a mídia quanto os formuladores de políticas estejam atentos às narrativas econômicas dominantes, pois estas influenciam decisões que afetam o bem-estar social e o desenvolvimento econômico, assegurando que as necessidades dos consumidores sejam refletidas nas políticas econômicas.

Palavras-chave: Narrativas econômicas, preços de petróleo, preços de alimentos, análise textual, expectativas, comportamento.

ABSTRACT

Economic narratives are stories that shape people's behavior and perceptions, influencing consumer expectations and directing economic discussions, especially in times of crisis, such as during COVID-19. The purpose of studying narratives related to prices during crisis periods is to understand whether dominant discourses reflect consumer concerns, as literature holds that food prices have a greater impact on consumer expectations. In this context, it is essential to assess whether the main news regarding price variations of certain consumer goods aligns with what traditional empirical economic theory reports, as this information influences consumer behavior and generates real consequences for the economy. Following this premise, this study conducted a textual analysis of articles from American newspapers published between December 2019 and July 2020, focusing on economic narratives related to prices. Among the most recurring topics identified, oil prices were the most prominent in the analyzed articles. The emphasis on this economic narrative occurred due to the drop in oil prices, which was predominantly portrayed negatively, as observed in the textual sentiment analysis. Conversely, the oil price narrative did not reflect consumers' most pressing concern, as it overlooked the effect of increased consumer purchasing power, a consequence of the price drop. It was found that the emphasis on oil price shocks overshadowed the discussion on the significant increase in food prices. Studies by the Michigan Consumer Research Center in the United States confirm that food prices influence consumer perceptions more than oil prices. Thus, it is important for both the media and policymakers to be attentive to dominant economic narratives, as these influence decisions that affect social welfare and economic development, ensuring that consumer needs are reflected in economic policies.

Keywords: Economic narratives, oil prices, food prices, textual analysis, expectations, behavior.

LISTA DE FIGURAS

FIGURA 1 – FLUXOGRAMA: MARCAÇÃO DE CLASSES GRAMATICAS.....	20
FIGURA 2 – FORMAÇÃO DE BIGRAMAS (N = 2).....	22
FIGURA 3 – FLUXOGRAMA MODELO VADER.....	23
FIGURA 4 – ENTRADA E NORMALIZAÇÃO DE TEXTO.....	23
FIGURA 5 – PONTUAÇÃO DE SENTIMENTO.....	24
FIGURA 6 – SOMA DE PONTUAÇÕES.....	24
FIGURA 7 – NORMALIZAÇÃO ALGÉBRICA.....	25
FIGURA 8 – FLUXOGRAMA: ANÁLISE PRELIMINAR.....	34
FIGURA 9 – FLUXOGRAMA: ANÁLISE TEXTUAL.....	35
FIGURA 10 – FREQUÊNCIA DE CO-OCORRÊNCIA DE BIGRAMAS E TERMOS.....	47
FIGURA 11 – EXTRAÇÃO DE FRASES DO CORPO DA NOTÍCIA.....	48
FIGURA 12 – RESULTADOS DE ANÁLISE DE SENTIMENTO POR EXPRESSÃO.....	52
FIGURA 13 – ÍNDICE DE PREÇOS AO CONSUMIDOR (2016 – 2020).....	56
FIGURA 14 – COMPARAÇÃO ENTRE “IPC-NÚCLEO” E “IPC-ALIMENTOS”.....	58
FIGURA 15 – EVOLUÇÃO ENTRE “IPC-NÚCLEO” E “IPC-GASOLINA”.....	58
FIGURA 16 – EXPECTATIVA E INFLAÇÃO OBSERVADA (2017 – 2023).....	59
FIGURA 17 – MENÇÕES: GASOLINA E ALIMENTOS (2021 –2023).....	61

LISTA DE TABELAS

TABELA 1 – CONTAGEM DE NOTÍCIAS POR PAÍS DE ORIGEM.....	36
TABELA 2 – AS 20 FONTES DE NOTÍCIAS MAIS COMUNS.....	37
TABELA 3 – OS 10 SUBSTANTIVOS MAIS COMUNS.....	38
TABELA 4 – INFLAÇÃO: N-GRAMAS (CORPO DAS NOTÍCIAS).....	40
TABELA 5 – INFLAÇÃO: N-GRAMAS (TÍTULO DAS NOTÍCIAS).....	41
TABELA 6 – TERMOS RECORRENTES: “PREÇOS”.....	42
TABELA 7 – PREÇOS: N-GRAMAS (TÍTULO DAS NOTÍCIAS).....	43
TABELA 8 – TERMOS RECORRENTES: “CONSUMIDOR”.....	44
TABELA 9 – CONSUMIDOR: N-GRAMAS (TÍTULO DAS NOTÍCIAS).....	45
TABELA 10 – RESULTADOS DE ESTATÍSTICA DESCRITIVA.....	49
TABELA 11 – RESULTADOS: TESTES DE NORMALIDADE.....	50
TABELA 12 – RESULTADOS: CORRELAÇÃO DE PEARSON.....	50
TABELA 13 – RESULTADOS: TESTE DE SPEARMAN.....	51
TABELA 14 – RESULTADOS: TESTE F E DE LEVENE.....	51
TABELA 15 – INFLAÇÃO E EXPECTATIVAS DO CONSUMIDOR.....	60
TABELA 16 – VARIAÇÃO MENSAL DE PREÇOS DA GASOLINA (2020-2021).....	61

SUMÁRIO

1	INTRODUÇÃO	12
2	QUAL É O DIÁLOGO ENTRE A ANÁLISE TEXTUAL E AS NARRATIVAS ECONÔMICAS?	14
3	METODOLOGIA	19
3.1	FONTE DE DADOS: AYLIEN COVID NEWS DATASET	19
3.2	TÉCNICAS DE ANÁLISE TEXTUAL.....	20
3.2.1	Marcação de Classes Gramaticais	20
3.2.2	Análise Textual: N-Grama	21
3.2.3	Análise de sentimentos: Modelo VADER.....	22
3.3	ESTATÍSTICA DESCRITIVA, TESTES DE HIPÓTESES E CORRELAÇÃO.....	26
3.3.1	Estatística Descritiva	26
3.3.1.1	Desvio Padrão.....	26
3.3.1.2	Assimetria.....	26
3.3.1.3	Curtose.....	27
3.3.2	Teste de Hipótese: Normalidade Amostral	28
3.3.2.1	Teste de Jarque-Bera (JB):	28
3.3.2.2	Teste de Hipótese: Teste de Shapiro-Wilk:	29
3.3.3	Teste de Hipótese: Testes de Variância.....	29
3.3.3.1	Teste F de Fisher	29
3.3.3.2	Teste de Levene.....	30
3.3.4	Testes de Correlação	31
3.3.4.1	Teste de Correlação de Pearson.....	31
3.3.4.2	Teste de Correlação de Spearman	32
4	ANÁLISE PRELIMINAR E ANÁLISE TEXTUAL DE NOTÍCIAS.....	34
4.1	ANÁLISE PRELIMINAR DA BASE DE DADOS	36
4.2	ANÁLISE TEXTUAL: BUSCA POR TERMOS MAIS RECORRENTES	39
4.2.1	Resultados de Co-ocorrência de Termos em Notícias.....	45
4.3	ANÁLISE TEXTUAL: ANÁLISE DE SENTIMENTOS	48
4.3.1	Resultados de análise de sentimento entre os trechos de texto	52
5	AS NOTÍCIAS ECONÔMICAS REFLETIRAM AS PRINCIPAIS PREOCUPAÇÕES DOS CONSUMIDORES?	54

5.1	QUAL O EFEITO DE UMA QUEDA NOS PREÇOS DE PETRÓLEO SOBRE O CONSUMO?	54
5.2	QUAIS BENS DE CONSUMO FORAM MAIS AFETADOS PELA INFLAÇÃO?	55
5.2.1	Qual é a relação entre os preços de gasolina e de alimentos?	57
5.3	DISCUSSÃO SOBRE RESULTADOS:.....	62
6	CONCLUSÃO	64
	REFERÊNCIAS	67

1 INTRODUÇÃO

A linguagem escrita é uma forma primordial de comunicação entre os agentes econômicos e, por meio das atuais tecnologias de programação, tornou-se um objeto de estudo quantificável. Pode parecer incomum para os acadêmicos mais conservadores, mas, por meio da análise de frequência de palavras e expressões em um grupo de textos, é possível obter dados quantificáveis, o que permite uma análise estatística e a avaliação de sua relevância. Segundo Loughran e McDonald (2020), a análise textual oferece ferramentas para examinar variáveis econômicas que são difíceis de capturar com dados tradicionais, por representar associações de ideias e conceitos não previstos pela análise convencional.

Seguindo essa linha de raciocínio, Shiller (2019) utilizou dados provenientes de técnicas de análise textual, para explicar como algumas histórias ganham popularidade, se espalham ao longo do tempo, influenciam os comportamentos das pessoas e se transformam em narrativas econômicas. Conforme Shiller (2019, p. 287, tradução nossa) alega que “à medida que os métodos de pesquisa avançam e que mais dados de mídias sociais se acumulam, a análise textual se tornará uma força mais poderosa na economia.” O autor também afirma que essa metodologia “nos ajudará a compreender melhor as manipulações e enganos deliberados que experienciamos, e a formular políticas econômicas que levem em conta as narrativas” (Shiller, 2019, p. 287, tradução nossa).

Tendo em vista a sugestão do autor, será feita a análise textual de notícias econômicas de dezembro de 2019 a julho de 2020 — período marcado pelo início da pandemia de COVID-19 e pela instabilidade econômica mundial, que fomentou a discussão de diversas narrativas econômicas. Com o objetivo de analisar empiricamente, este trabalho buscará identificar qual foi a principal narrativa econômica relacionada a temas de interesse do consumidor, com foco específico em “preços”, produzida pelo jornalismo americano durante o período mencionado. Além disso, analisa como essa narrativa se relaciona com o comportamento do consumidor.

Para identificar qual a narrativa econômica foi a mais proeminente para “preços”, foi realizada uma busca dos bens de consumo cujos preços foram mais citados, e analisada a ocorrência simultânea desses bens juntamente com o termo “consumidor” no corpo das notícias. Portanto, o primeiro objetivo dessa busca, é discutir a relação entre o principal tema tratado por jornalistas e, entender como ele afeta o consumo e as expectativas dos consumidores em relação à economia. A partir dos dados encontrados, faremos uma comparação com os bens de consumo e serviços que foram mais inflacionados durante o ano

de 2020. Em seguida, será realizada uma análise dos itens que possivelmente mais influenciaram as expectativas inflacionárias dos consumidores para o ano de 2021. Portanto, o segundo objetivo deste trabalho, é discutir a relevância da narrativa econômica mais proeminente envolvendo preços, ao considerar o contexto econômico para o consumidor no período das notícias.

Com base no princípio de que as pessoas utilizam atalhos mentais derivados principalmente de experiências anteriores e, de informações disponíveis para tomar decisões, serão utilizados conceitos da Economia Comportamental, uma vez que "narrativas são construções humanas que são misturas de fatos, emoções, interesses humanos e outros detalhes irrelevantes que formam uma impressão na mente humana" (Shiller, 2019, p. 65, tradução nossa). Portanto, a forma como uma informação é disposta dentro de uma notícia econômica, pode facilmente conduzir as percepções do consumidor e afetar as expectativas dele, de tal modo que interfira no próprio percurso da economia. Para identificar as narrativas presentes nas notícias analisadas, foi utilizada a Análise de Sentimento, que é uma técnica de Processamento de Linguagem Natural. Segundo Wang, Aduruki e Dasari (2023), a técnica é frequentemente utilizada para capturar os sentimentos transmitidos por notícias econômicas e, as impressões públicas através de comentários em redes sociais como o Reddit e X (Twitter).

O terceiro objetivo deste trabalho, será a condução de uma análise de sentimentos para as notícias que envolvem “preços”, para a narrativa que apareceu com o maior destaque. O intuito é entender qual foi o enquadramento que os jornalistas atribuíram às informações, e verificar se ele é compatível com as reais consequências econômicas previstas pela análise econômica tradicional.

Este trabalho tem o objetivo de evidenciar a relevância do estudo de narrativas econômicas que não estão alinhadas as expectativas dos agentes econômicos. Essa característica é relevante para a formulação de políticas econômicas que procurem atender às necessidades mais urgentes em momentos críticos. Contudo, a análise do contágio dessa narrativa ao longo do tempo, bem como seus impactos diretos, não será abordada neste estudo. O propósito principal concentra-se na identificação da narrativa econômica mais discutida nas notícias analisadas. Porém, o trabalho não analisará como a principal narrativa econômica encontrada influenciou a economia durante o período estudado. Por meio dos objetivos mencionados, este trabalho busca contribuir para o entendimento das narrativas econômicas sobre os consumidores durante a pandemia, ampliando o diálogo entre a Análise Textual e a Economia Comportamental.

2 QUAL É O DIÁLOGO ENTRE A ANÁLISE TEXTUAL E AS NARRATIVAS ECONÔMICAS?

O livro "Economia Narrativa: Como as Estórias se Tornam Virais e Impulsionam Grandes Acontecimentos Econômicos" de Robert Shiller, explica a importância das narrativas econômicas dentro da economia. Por narrativa, deve ser entendido que é uma história ou um conjunto de histórias que podem ser verdadeiras ou fictícias. Segundo Shiller (2019) muitos economistas não dão a devida atenção para o contexto histórico e social no qual as narrativas são formadas, porque consideram que isso seria objeto de estudo de outras ciências, tal como a Linguística e a Sociologia. No entanto, o autor destaca que os acadêmicos dessas áreas enfrentam dificuldades ao discutir a teoria econômica, o que gera uma lacuna em relação aos efeitos das narrativas econômicas sobre a economia.

É necessário ressaltar que o estudo sobre narrativas dentro das ciências econômicas não foi iniciado por Shiller. A importância das narrativas na economia já havia sido abordada anteriormente por McCloskey (1990). Segundo Roos e Reccius (2022), a principal diferença entre o tratamento desenvolvido por McCloskey e Shiller, é que McCloskey destaca a importância das narrativas na maneira como a prática econômica ocorre, enquanto Shiller argumenta que as narrativas são essenciais para a análise de eventos econômicos, como bolhas ou crises financeiras.

Embora o objetivo do trabalho não seja analisar o “contágio” das narrativas, Shiller (2019) argumenta que as narrativas têm comportamento similar ao de epidemias, no que diz respeito à propagação e crescimento de uma narrativa, assim como a subsequente queda. A forma como uma narrativa econômica se espalha então, pode ser através da mídia ou de pessoa para pessoa, portanto ambos os canais, refletem a importância da análise textual sobre as informações que são escritas por portais de notícias. Segundo Shiller (2019), as narrativas seriam espalhadas por pessoas que estariam "infectadas", e que a espalhariam por algum tempo, e então, haveria um período de esquecimento, no qual tal narrativa perderia força e as pessoas parariam de discuti-la. O autor entende que o “contágio” de uma narrativa pode acontecer quando os indivíduos conversam entre si, independentemente de ser pessoalmente ou dentro das redes sociais.

O autor afirma que economia está atrasada em relação a outros campos de estudo, no que diz respeito à importância das narrativas, pois o problema de muitos economistas é que são intelectuais que estão preocupados apenas com os principais indicadores econômicos, tais como o PIB, inflação, dentre outros. Além disso, as demais áreas de conhecimento também

não se arriscam a debater sobre temas econômicos de forma aprofundada e, consequentemente, existe um espaço que acaba não sendo preenchido pelos acadêmicos de economia. Um ponto importante sobre a teoria de Narrativas Econômicas de Shiller (2019), é sobre a consiliência, que é a união entre diferentes campos de conhecimento, o que o autor julga como fundamental na análise da economia narrativa, conforme a passagem abaixo:

A palavra consiliência, [...], significa a união do conhecimento entre diferentes disciplinas acadêmicas, especialmente entre as ciências e as humanidades. Quando se reflete que a economia é composta por pessoas vivas e conscientes, que veem suas ações à luz de histórias carregadas de emoções e ideias, percebe-se a necessidade de múltiplas perspectivas. Portanto, a economia narrativa requer conceitos de quase todos os departamentos universitários (Shiller, 2019, p. 12, tradução nossa).

Ao considerarmos que este trabalho fez amplo uso de técnicas de análise textual, a primeira forma de consiliência, será entre a Ciência Econômica e um subcampo da Ciência da Computação, o Processamento de Linguagem Natural (PLN). Outra forma de consiliência que pode ser ressaltada é entre o Jornalismo e a Ciência Econômica, em que ambas as profissões utilizam uma metodologia de pesquisa, observação e análise de dados. A profissão do jornalista pode requerer que ele escreva notícias sobre diversas áreas como economia, saúde, educação, segurança, dentre diversos outros temas, sem que este indivíduo seja um especialista sobre essas áreas. Ao levar em conta o jornalismo como uma atividade que visa analisar e transmitir informações ao grande público, temos que os jornalistas econômicos podem influenciar as percepções e conduzir as expectativas dos consumidores em relação a economia. Não há garantias de que os jornalistas que produzem notícias de teor econômico tenham o mesmo conhecimento acadêmico que um economista, e uma possível consequência, é que a informação escrita não corresponda as previsões econômicas factuais.

Como um dos objetivos do trabalho é rastrear qual foi o principal tema discutido sobre “preços”, a análise textual permitirá uma visualização sobre qual narrativa relacionada aos preços apareceu com maior destaque durante o primeiro semestre de 2020. Conforme Shiller (2019, p. 21, tradução nossa) “também há contágio de uma fonte de notícias e programas de entrevistas para outro, à medida que eles assistem e leem as histórias uns dos outros.”. O autor aponta que os jornalistas prestam atenção nas notícias em países estrangeiros, uma vez que o que viraliza em um pode viralizar em outro. Portanto, o conteúdo produzido por um canal de notícias, pode não ser consequência do que é mais relevante para os leitores, mas sim do que é o assunto mais popular no momento. Além disso, Shiller (2019)

fala que as histórias de maior importância são as sobre a confiança do consumidor no sistema bancário, assim como em negócios.

Ao levarmos em conta que o jornalismo tem o objetivo de informar um público-alvo sobre determinado acontecimento, entende-se que este possui grande poder de influência sobre o sentimento do mercado financeiro em relação às empresas ou ações, sendo uma origem de discussão sobre especulações financeiras. A análise de Bask, Forsberg e Östling (2024) afirma que vários estudos mostraram como o tom de notícias afeta os retornos das ações. Além disso, segundo Heston e Sinha (2017, apud Bask, Forsberg e Östling, 2024), informações divulgadas por meio de notícias podem prever os retornos das ações com um a dois dias de antecedência, com notícias positivas causando um aumento rápido nos preços das ações e notícias negativas causando uma reação mais lenta. Um segundo estudo que reforça essa relação, ainda segundo Garcia (2013, apud Bask, Forsberg e Östling, 2024), revela uma correlação entre o tom das palavras e os resultados das ações no futuro, principalmente durante recessões econômicas. Portanto, o jornalismo desempenha um papel econômico crucial, uma vez que pode moldar como investidores podem reagir a determinadas informações acerca de eventos financeiros, o que pode influenciar na compra e venda de ativos.

Quanto aos efeitos psicológicos causados pelas informações, Kahneman (2012) tratou do efeito de enquadramento – também conhecido como “*framing*” – no qual a forma como uma informação é apresentada, tende a influenciar as percepções e decisões das pessoas. Outro aspecto importante da teoria comportamental de Kahneman (2012), relevante para a análise desenvolvida, são os atalhos mentais que as pessoas utilizam para tomar decisões, chamados de heurísticas. Um exemplo é a heurística de disponibilidade, na qual, quanto maior for a exposição de um indivíduo a um evento, mais ele tenderá a superestimar a probabilidade de o evento acontecer com ele. Por exemplo, se uma pessoa vê constantemente nos noticiários casos de roubo de carros, ela poderá estar mais propensa a contratar um seguro.

Portanto, conforme Kahneman (2012), a maneira como uma informação é apresentada pode influenciar as expectativas dos consumidores em relação à economia. Além disso, os estudos sobre os efeitos de enquadramento revelam que uma informação pode ser provocar emoções distintas dependendo da forma como é apresentada. Este trabalho buscou identificar o tipo de enquadramento utilizado nas notícias econômicas referentes aos consumidores, com o objetivo de analisar se esse enquadramento foi predominantemente

positivo ou negativo. A disposição das informações é um fator crucial para moldar as percepções dos consumidores sobre a economia.

A análise textual dentro do campo de Processamento de Linguagem Natural é realizada através de métodos empíricos para coleta de dados linguísticos, como a técnica “N-Grama”, que é um método de análise de dados que calcula quais as combinações mais frequentes com uma palavra ou frase específica, Shiller (2019) faz amplo uso da análise textual, de forma a descrever como algumas narrativas – como a narrativa do bimetalismo e do Bitcoin –, se tornam virais ao longo do tempo. A justificativa do autor para utilizar essa técnica de análise textual é que “as narrativas assumem a forma de sequência de palavras, o que torna os princípios da linguística relevantes” (Shiller, 2019, p. 16, tradução nossa).

Como segunda forma de consiliência entre a Ciência Econômica e o Processamento da Linguagem Natural, será utilizada a Análise de Sentimento. A Análise de sentimento, é técnica de análise textual, cujo objetivo é encontrar a impressão sentimental que ele causa no leitor. Segundo Taherdoost e Madanchian (2023), a análise de sentimento é uma técnica para avaliar o caráter subjetivo de uma informação, ou seja, analisar o quanto uma informação é positiva ou negativa, em um intervalo de -1 a +1. Além disso, os autores relatam que existem estudos que aplicam esta técnica desde o final da década de 1990. No contexto econômico, a análise de sentimento pode ser utilizada para várias finalidades, como entender qual a opinião dos eleitores sobre um candidato à eleição, a opinião de consumidores sobre um determinado produto, dentre outros.

Existem diversas formas de realizar a análise de sentimento, mas será explicada a forma geral como este processo ocorre. Vamos analisar o caso da Frase A, abaixo:

- Frase A: "A taxa de inflação aumentou cerca de 2% nos últimos 3 meses".

O leitor que se deparar com esta frase, muito provavelmente vai atribuir a ela um sentido negativo, que é causado pela presença da palavra "inflação", e como não há nenhum outro termo que altere o sentido da frase, o sentimento final transmitido por ela é negativo. A partir disso, poderíamos atribuir para a palavra “inflação” uma pontuação de "-1". Em seguida, temos o caso da frase B:

- Frase B: "Política fiscal adotada pelo governo, induziu ao crescimento do PIB em 0,6%, mas causa aumento da taxa de inflação".

Ao analisarmos a frase B, enxergamos dois aspectos nela: um positivo – causado pela informação de que houve um aumento no crescimento do PIB – e um negativo, que é consequência do aumento da taxa de inflação. Para este caso, não podemos concluir se houve um maior impacto positivo ou negativo na economia, decorrente da falta de contexto sobre o

que pode ter acontecido. Podemos atribuir uma pontuação positiva para a palavra "crescimento", no valor de "1" e sabemos que o termo inflação tem um valor de "-1". O sentimento geral transmitido pela frase é ambíguo, ou neutro, e isso é representado pela soma das duas pontuações, que resulta em zero.

Em notícias de contexto econômico, a utilização desta técnica se justifica pela possibilidade de processar grandes volumes de dados de maneira mais rápida e eficiente, do que de uma forma manual. Segundo Loughran e McDonald (2020), dentro do contexto econômico, esta técnica é utilizada no mercado financeiro como auxílio na previsão de tendências de mercado e, posterior tomada de decisão sobre investimentos. A análise de sentimentos pode ser utilizada para tentar entender o comportamento do investidor, assim como o sentimento do mercado com determinadas empresas ou ações.

O motivo que inspirou a escolha do termo “preços” como foco para as análises linguísticas, é pela importância central na mentalidade dos consumidores e sua influência sobre o Índice de Sentimento do Consumidor. Shiller (2019) ressalta o papel da economia comportamental sobre o desenvolvimento da ciência econômica e enfatiza a importância de George Katona, psicólogo econômico pioneiro em estudos de como percepções, expectativas e sentimentos dos consumidores moldam suas decisões econômicas e impactam variáveis como consumo e poupança. Katona (1975) defendia que políticas econômicas deveriam considerar fatores psicológicos, já que o comportamento humano é crucial para prever e influenciar a atividade econômica. Conforme Curtin (2016), Katona acreditava que os consumidores focavam em variáveis econômicas essenciais, como emprego, renda, inflação e taxas de juros, sem necessariamente fazer previsões baseadas em dados passados. Warneryd (1982) também observa que Katona identificava uma lacuna no entendimento popular sobre fatos econômicos, como a relação entre produção, salários e preços. Uma das contribuições mais significativas de Katona foi o desenvolvimento do Índice de Sentimento do Consumidor, criado para medir as expectativas e ações dos consumidores. Hoje, o índice é calculado com base em cinco perguntas que avaliam a percepção sobre as condições econômicas presentes e futuras, bem como as expectativas em relação à inflação e ao consumo de bens duráveis, reforçando a importância dos preços no comportamento e expectativas dos consumidores.

3 METODOLOGIA

A seguir, serão detalhadas as etapas essenciais para a construção deste trabalho. Primeiramente, será apresentada uma ampla base de dados, que contém cerca de 1,6 milhões de notícias e já foi utilizada em outros estudos. Posteriormente, será explicada a técnica de marcação de palavras por classes gramaticais, bem como a técnica de análise de N-gramas, e por fim, o detalhamento sobre a análise de sentimentos VADER.

3.1 FONTE DE DADOS: AYLIEN COVID NEWS DATASET

A base de dados utilizada para fazer este trabalho, é a Aylien (2020), que é um arquivo JSONL de 25gb, que apresenta mais de 1,6 Milhões de notícias a nível internacional, de novembro de 2019 a julho de 2020. A legitimidade do uso da base de dados Aylien (2020), é reforçada pelo uso dela em diversas pesquisas acadêmicas, como em: Pitale (2020), Al-Obeidat et al. (2020), Min et al. (2021), Gupta e Katarya (2021), Wen et al. (2022), Bayram (2022), Zhan, Gui e Procter (2023), Dorn et al. (2023) e Son (2023), que aplicaram a base para a exploração de diversos temas.

No estudo de Dorn *et al.* (2023), foi feito um estudo de modo a rastrear o gênero dos especialistas que foram utilizados como argumentos de autoridade. O argumento que o estudo apresenta, é que geralmente, os jornalistas recorrem a especialistas de alto nível de instituições acadêmicas, ao invés de procurar indivíduos que tenham mais experiência de trabalho, mas que por estarem afiliados a instituições de menor prestígio, não são procurados com tanta frequência. Na pesquisa de Wen *et al.* (2022), a base de dados foi utilizada com o objetivo de analisar as medidas não farmacológicas de intervenção que foram utilizadas na época.

O processamento dos dados foi feito através da linguagem de programação Python, que é uma ferramenta de código aberto, popularmente utilizada por empresas como Google, Facebook e Amazon, para processar e analisar dados. De forma mais específica, foi utilizado o Apache Spark que, segundo a Amazon (2024), é um sistema que pode oferecer consultas analíticas rápidas de dados de qualquer tamanho, o que permitiu uma análise mais precisa dos dados.

3.2 TÉCNICAS DE ANÁLISE TEXTUAL

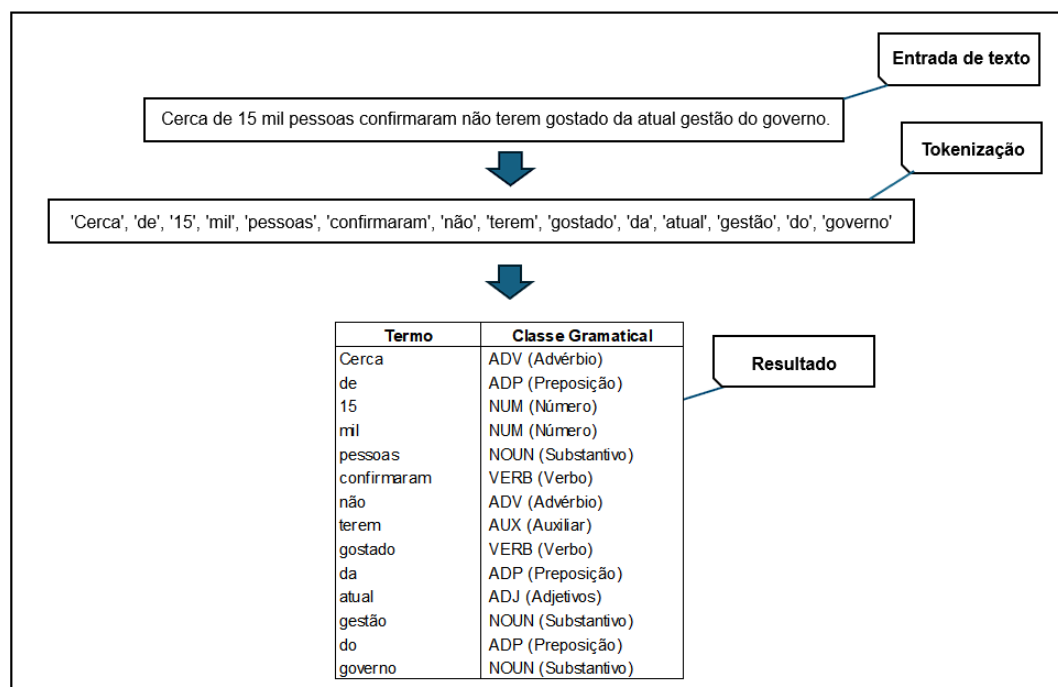
Nesta sessão, serão mostradas as três metodologias de análise textual utilizadas durante este trabalho: sendo a primeira a marcação de classes gramaticais - também conhecida como *POS tagging* (*Part of Speech – Tagging*), - a segunda será a técnica de pesquisa de N-gramas, e por último, a Análise de Sentimentos.

3.2.1 Marcação de Classes Gramaticais

Conhecida como *POS tagging* – sigla em inglês para Part-of-Speech tagging, que em tradução livre significa “Marcação de Classe Gramatical” –, refere-se à classificação das palavras de acordo com a sua função gramatical, como adjetivos, substantivos, verbos, entre outros (Bird; Klein; Loper, 2009). Para realizar essa tarefa, foi utilizada a biblioteca *spaCy* do Python, amplamente empregada em processamento de linguagem natural, conforme Vasiliev (2020).

Abaixo, na Figura 1, está um fluxograma que detalha as etapas do processo de análise das classes gramaticais.

FIGURA 1 – FLUXOGRAMA: MARCAÇÃO DE CLASSES GRAMATICAIS



FONTE: elaborado a partir das ideias de Bird, Klein e Loper (2009).

Inicialmente, temos a entrada de uma frase ou um texto dentro de um algoritmo. Em seguida, temos a separação das palavras que compõem esta frase para a formação de uma lista de palavras, que é uma técnica mais popularmente conhecida como “Tokenização”. Por último, temos a comparação de cada uma das palavras com as palavras do dicionário do spaCy, que então faz a marcação gramatical.

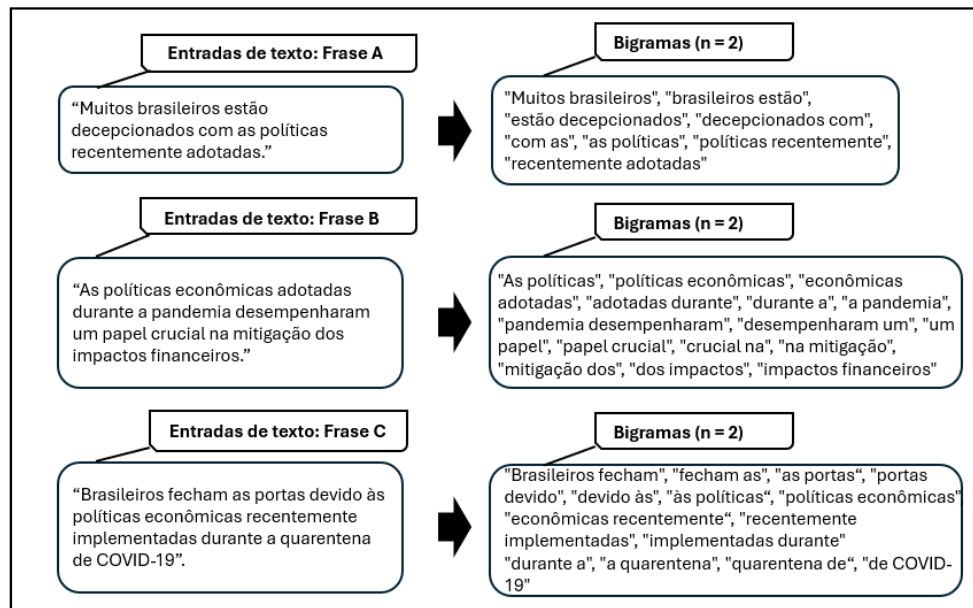
3.2.2 Análise Textual: N-Grama

De acordo com Jurafsky e Martin (2018, p. 32, tradução nossa), o “N-grama é uma sequência n de palavras: o 2-grama (o qual chamaremos de bigrama), é uma sequência de duas palavras [...], e o 3-grama (um trigrama) é uma sequência de três palavras”, assim como também temos o 4-grama (um quadrigrama), e assim sucessivamente. Ou seja, o “ n ” seria o número de palavras dentro de um conjunto. Além disso, a análise de n-gramas se justifica para “escolher palavras contextualmente melhores” (Jurafsky, Martin, 2018, p. 32, tradução nossa). Por exemplo, temos as seguintes frases:

- Frase A: "Muitos brasileiros estão decepcionados com as políticas econômicas recentemente adotadas.”
- Frase B: "As políticas econômicas adotadas durante a pandemia desempenharam um papel crucial na mitigação dos impactos financeiros.”
- Frase C: "Brasileiros fecham as portas devido às políticas econômicas implementadas durante a quarentena do COVID-19."

Ao inserirmos estas frases no algoritmo, ele irá gerar bigramas utilizando todas as palavras de cada frase, conforme no fluxograma da Figura 2 a seguir:

FIGURA 2 – FORMAÇÃO DE BIGRAMAS (N = 2)



FONTE: elaborado pelo autor, a partir das ideias de Jurafsky e Martin (2018).

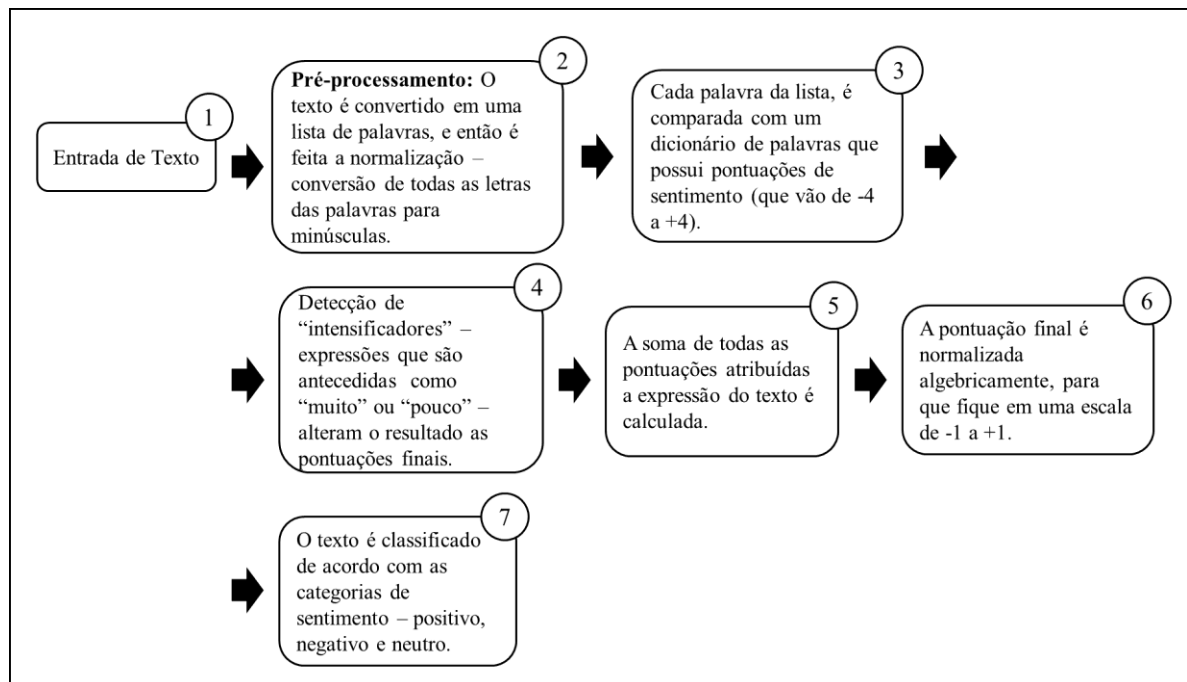
Uma das vantagens de usar modelos N-Grama, é que conforme o valor N aumenta, podemos capturar mais palavras e ter uma maior precisão sobre o contexto nos quais as palavras estão sendo utilizadas, mas isso levanta a possibilidade de aparecerem resultados menos frequentes. Esta técnica permite capturar as expressões mais utilizadas dentro de um corpo de texto, o que é importante para mensurar a recorrência de um determinado tema dentro de um grande volume de dados, como a base de notícias Aylien (2020).

3.2.3 Análise de sentimentos: Modelo VADER

Dentro da área de análise textual existem diversas metodologias para realizar a análise de sentimentos, uma das mais populares é o modelo VADER (Valence Aware Dictionary and Sentiment Reasoner), ou “Dicionário Consciente de Valência Analisador de Sentimento” em tradução livre, é um dicionário desenvolvido para a análise de sentimentos.

O VADER foi desenvolvido por uma equipe de pesquisa do Instituto de Tecnologia da Georgia (Georgia Tech), que foi liderado por Eric Gilbert e C.J. Hutto. Essa metodologia atribui pontuações de sentimentos às palavras do texto analisado, com base em um dicionário cujas palavras foram classificadas quantitativamente como neutras, positivas ou negativas (Hutto; Gilbert, 2014). Com base nisso, foi elaborado o fluxograma a seguir, com o intuito de representar as principais etapas do modelo de análise textual.

FIGURA 3 – FLUXOGRAMA MODELO VADER



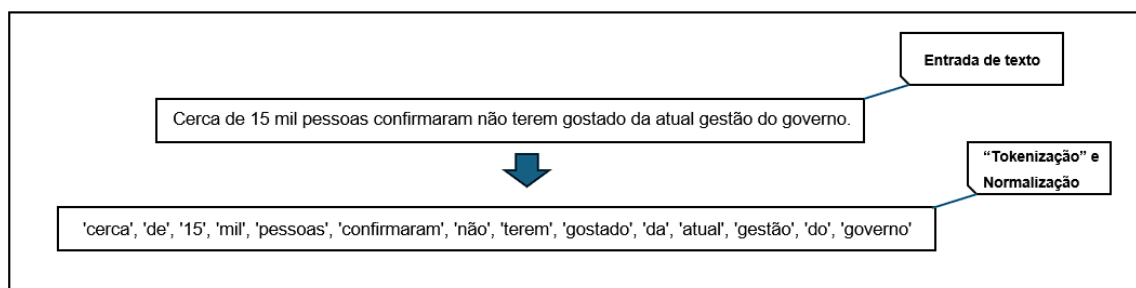
FONTE: Elaborado com base nas ideias de Hutto e Gilbert (2014).

Conforme indicado no fluxograma, após a entrada de um texto dentro do algoritmo, este texto será processado e convertido em uma lista de caracteres – processo que é chamado de “Tokenização”. Por exemplo, temos a frase a seguir:

- Frase A: “Cerca de 15 mil pessoas confirmaram não terem gostado da atual gestão do governo”.

A frase é convertida em uma lista de caracteres, e então é feita a normalização – que é a conversão dos caracteres que estão em letra maiúscula para letra minúscula –, em seguida, temos a informação resultante conforme é demonstrada na Figura 4:

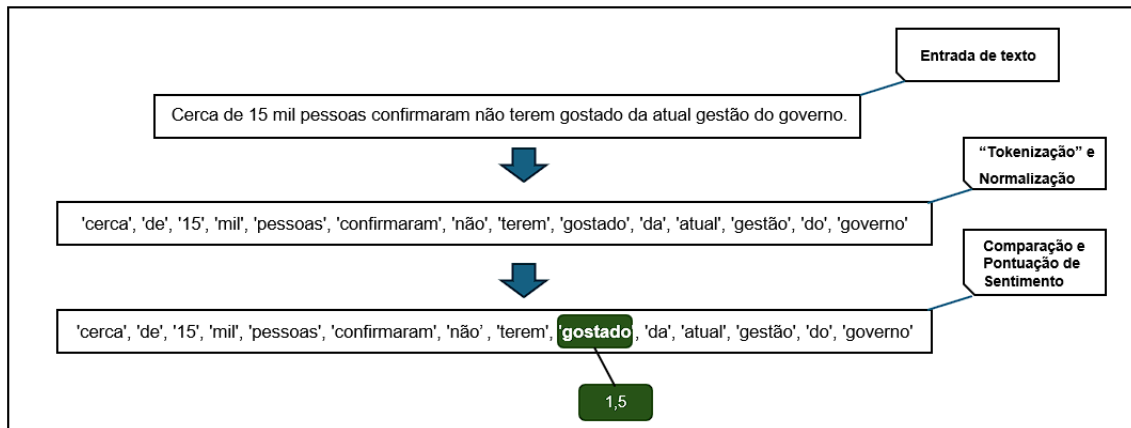
FIGURA 4 – ENTRADA E NORMALIZAÇÃO DE TEXTO



FONTE: Elaborado com base nas ideias de Hutto e Gilbert (2014).

Após as etapas de tokenização e normalização da frase, a etapa seguinte é a comparação de todas as palavras da lista, com os termos do dicionário VADER, processo que é ilustrado na Figura 5:

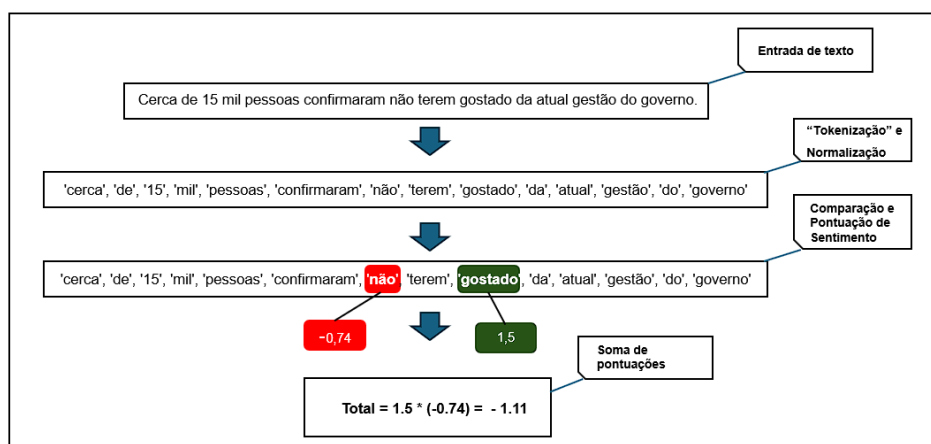
FIGURA 5 – PONTUAÇÃO DE SENTIMENTO



FONTE: Elaborado com base nas ideias de Hutto e Gilbert (2014).

Como foi dito anteriormente, as pontuações das palavras presentes no VADER variam no intervalo entre -4 e +4. A palavra “gostado”, apresenta uma pontuação de +1,5. Contudo, se a análise fosse concluída nessa etapa, a expressão da frase teria sentido positivo, quando sabemos que ela é negativa. Portanto, o que o algoritmo faz em seguida é verificar a existência de termos de negação – como a palavra “não” –, que possam estar contidos anteriormente na frase, e multiplica a pontuação de sentimentos por um fator de -0.74, conforme definido no algoritmo de Hutto e Gilbert (2014). Esta etapa está ilustrada na Figura 6.

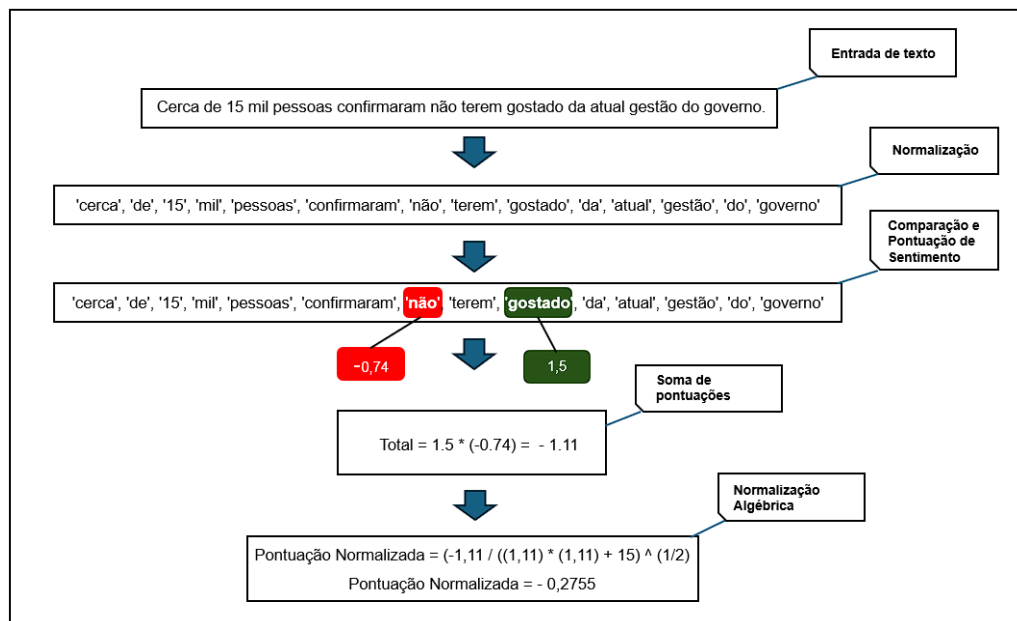
FIGURA 6 – SOMA DE PONTUAÇÕES



FONTE: Elaborado com base nas ideias de Hutto e Gilbert (2014).

Temos que a pontuação final de sentimento, é de -1.11 o que condiz com a emoção negativa expressa pela frase. Contudo, quanto maior for o tamanho do texto, teremos pontuações que podem ser numericamente muito negativas ou muito positivas, o que dificultaria a comparação entre informações de diferentes tamanhos, porque elas não teriam uma escala fixa de comparação. A solução encontrada pelos criadores foi a seguinte: a conversão da pontuação através de uma equação de tangente hiperbólica, pois a tangente hiperbólica tem a característica de oferecer uma transição suave na medida que os valores aumentam positivamente ou negativamente e, a pontuação de sentimentos final, fica no intervalo que varia de -1 a +1, conforme é mostrado na Figura 7, a seguir:

FIGURA 7 – NORMALIZAÇÃO ALGÉBRICA



FONTE: Elaborado com base nas ideias de Hutto e Gilbert (2014).

Essa propriedade matemática é desejável ao se realizar análises de sentimentos, uma vez que evita outliers e, conseqüentemente, a adição de incrementos tem um impacto que tende ser decrescente, o que evita grandes saltos numéricos na pontuação. Então, temos que a pontuação normalizada da frase A é igual a -0.2755, o que significa que a frase analisada possui um sentimento negativo, e indica que a gestão é predominantemente desfavorável.

3.3 ESTATÍSTICA DESCRITIVA, TESTES DE HIPÓTESES E CORRELAÇÃO

A seguir, serão demonstrados os cálculos matemáticos e as formas de interpretação dos resultados de algumas medidas de estatística descritiva, assim como de testes de hipóteses.

3.3.1 Estatística Descritiva

Serão explicados os elementos de estatística descritiva utilizados para a construção do trabalho: desvio padrão, assimetria e curtose.

3.3.1.1 Desvio Padrão

O desvio padrão é uma medida para calcular a variação dos valores de uma amostra em relação a média. Quanto maior o desvio padrão, maior é a dispersão dos valores observados em relação à média. (FARIAS, 2020):

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

No qual:

- s , é o valor do desvio padrão;
- $\bar{X}$, é a média amostral;
- n , é o número de observações da amostra;
- X_i , representa cada valor individual da amostra.

3.3.1.2 Assimetria

Segundo o National Institute of Standards and Technology (2024), a assimetria de Fisher é uma medida estatística da simetria de distribuição de dados. Além disso, ela possibilita visualizar a distribuição dos dados de uma amostra em torno da média. No qual a concentração dos dados caracteriza a assimetria como negativa ou positiva. Segundo Gujarati (2011), a assimetria pode ser representada pela equação a seguir:

$$S = \frac{E(X - \mu)^3}{\sigma^3}$$

No qual:

- E, Operador de esperança, ou seja, a média dos valores.
- S, assimetria da distribuição.
- X, Variável aleatória.
- μ , Média da distribuição.
- σ , Desvio padrão da distribuição.

3.3.1.3 Curtose

A curtose é uma medida estatística para avaliar o grau de concentração dos dados amostrais nas caudas da curva de distribuição. Ela é útil para apontar se uma distribuição tem muitos valores extremos. Segundo Gujarati (2011), a equação de curtose é representada da seguinte forma:

$$K = \frac{E(X - \mu)^4}{[E(X - \mu)^2]^2}$$

No qual:

- K, Curtose da distribuição;
- E, Operador de esperança (esperança matemática ou valor esperado), que significa a média ponderada dos valores;
- X, Variável aleatória;
- μ , média da distribuição.

O valor de K sugere como os dados se distribuem. Se K for menor que 3, a curva de distribuição possui caudas mais curtas e é chamada de platicúrtica. Por outro lado, se K for maior que 3, as caudas são mais longas, sendo chamadas de leptocúrticas. Caso o valor de K seja igual a 3, a distribuição é considerada mesocúrtica, sendo semelhante à distribuição normal.

3.3.2 Teste de Hipótese: Normalidade Amostral

De acordo com Gujarati (2011) os testes de normalidade tem o objetivo de verificar se a distribuição de uma amostra é consistente com as características de uma distribuição normal.

3.3.2.1 Teste de Jarque-Bera (JB):

Segundo Gujarati (2011), o Teste de Jarque-Bera tem o objetivo de verificar se uma amostra segue uma distribuição normal. O teste é baseado nas estatísticas de curtose e assimetria dos dados. Cujas fórmulas são descritas abaixo:

$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right]$$

No qual:

- JB é a estatística Jarque-bera;
- n é o número de observações da amostra;
- S é a estatística de assimetria;
- K é a estatística de curtose.

O teste de Jarque-Bera tem as seguintes hipóteses: a Hipótese Nula (H_0), na qual os dados da amostra apresentam uma distribuição normal, e em contrapartida, a Hipótese Alternativa (H_1), na qual os dados da amostra não apresentam uma distribuição normal. A forma de interpretar o resultado e analisar as hipóteses, ocorre da seguinte forma: quando o p-valor calculado for menor que o nível de significância do teste, – o que ocorre quando o valor da estatística JB é elevado –, pode-se rejeitar a hipótese nula de que a amostra segue uma distribuição normal. Por outro lado, quando o p-valor obtido for maior que o nível de significância, o que ocorre quando o valor da estatística JB está próximo de zero, não se pode rejeitar a hipótese nula de normalidade. Além disso, segundo Gujarati (2011), o teste Jarque-Bera segue a distribuição da qui-quadrado e tem dois graus de liberdade.

3.3.2.2 Teste de Hipótese: Teste de Shapiro-Wilk:

Abaixo temos o teste de normalidade desenvolvido por Shapiro e Wilk (1965).

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

No qual:

- W , representa o valor da estatística do teste;
- $\bar{x}$, representa a média amostral;
- a_i , são coeficientes tabelados;
- x_i , são os valores da amostra;
- $x_{(i)}$, são os valores da amostra em ordem crescente;

O teste de Shapiro-Wilk segue as seguintes hipóteses: como Hipótese Nula (H_0), teremos que os dados amostrais seguem uma distribuição normal, e a Hipótese Alternativa (H_1) assume que os dados não seguem uma distribuição normal. A interpretação da estatística W , pode ser feita através do cálculo do P-Valor relacionado, no qual se o P-valor for menor que o nível de significância do teste, pode-se rejeitar a hipótese nula de que os dados da amostra seguem uma distribuição normal. Porém, se o P-valor for maior que o nível de significância, não se pode rejeitar a hipótese nula, o que indica a que os dados são compatíveis com uma distribuição normal.

3.3.3 Teste de Hipótese: Testes de Variância

De acordo com Gujarati (2011), a variância é uma medida da variabilidade de um conjunto de observações em relação à média. Para verificar a relação entre as variâncias de duas amostras, podemos utilizar os testes F de Fisher, assim como o teste de Levene.

3.3.3.1 Teste F de Fisher

Segundo Gujarati (2011), o Teste F de Fisher busca comparar as variâncias amostrais, com o objetivo de verificar se as variâncias estão com diferenças estatisticamente significantes entre si.

O valor F é calculado através da razão entre as duas variâncias:

$$F = \frac{S_1^2}{S_2^2}$$

Na qual:

- S_1^2 é a variância da amostra 1;
- S_2^2 é a variância da amostra 2;

Com relação as hipóteses do teste, temos que: de acordo com a Hipótese Nula (H_0), as variâncias das amostras são iguais e, na Hipótese Alternativa (H_1), as variâncias amostrais são diferentes. A partir do cálculo do P-valor da estatística do teste, é possível testar as hipóteses. Se o P-valor for menor que o nível de significância, é necessário rejeitar a hipótese nula, uma vez que as variâncias dos grupos não são iguais. Porém, se o P-valor for maior que o nível de significância, não se pode rejeitar a hipótese nula, pois não há evidências suficientes para afirmar que as variâncias são significativamente diferentes. Além disso, é importante ressaltar que o Teste F supõe a normalidade, independência e homoscedasticidade das amostras. O teste F segue distribuição própria e, o número de graus de liberdade, depende do número de observações da amostra 1 e 2.

3.3.3.2 Teste de Levene

Abaixo, o teste de variância criado por Levene (1960), cujo objetivo é o de analisar a igualdade de variâncias entre duas ou mais amostras, ou seja, tem o propósito de avaliar se a variabilidade dos dados de ambas as amostras é consistente.

$$W = \frac{(N - k)}{(k - 1)} \times \frac{\sum_{i=1}^k N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

Sendo que:

$$Z_{ij} = |Y_{ij} - \bar{Y}_i|$$

$$Z_{i.} = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$$

$$Z_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{N_i} Z_{ij}$$

No qual:

- k é o número de grupos.
- N_i é o número de casos para cada grupo “i”;
- N é o número total de observações;
- Y_{ij} é o valor mensurado da variável para o caso “j”, do grupo “i”.
- Z_{ij} é o valor da média para o grupo “i”;
- $Z_{..}$ é o valor da média de todos Z_{ij}

Com relação as hipóteses dos testes, temos que: de acordo com a Hipótese Nula (H_0), as variâncias das amostras são iguais, e como contrapartida temos a Hipótese Alternativa (H_1), na qual as variâncias são diferentes. O teste de Levene é interpretado de acordo com o P-valor obtido a partir da estatística W, no qual se o P-valor for menor que o nível de significância, podemos rejeitar a hipótese nula de que as variâncias das amostras são iguais. Porém, se o P-valor for maior que o nível de significância, isso sugere que não há evidências o suficiente para rejeitar a hipótese nula.

3.3.4 Testes de Correlação

Serão utilizados os testes de Correlação de Pearson, de Spearman e de Kendall. Portanto, serão explicados o cálculo de cada um dos testes, assim as possíveis interpretações dos resultados.

3.3.4.1 Teste de Correlação de Pearson

Segundo Gujarati (2011), o teste de correlação de Pearson tenta medir a relação linear entre duas variáveis. O teste leva em consideração algumas suposições: a linearidade (as duas variáveis possuem uma relação linear), a normalidade (as variáveis seguem uma distribuição normal), a homoscedasticidade (homogeneidade da variância) e a independência (as observações são independentes).

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

No qual:

- X_i e Y_i são os valores das observações de suas respectivas amostras;
- $\bar{X}$ e $\bar{Y}$ são as médias das variáveis X e Y respectivamente.
- r é o valor do coeficiente de correlação de Pearson.

Se r for igual a 1, isso significa que existe uma correlação linear perfeita, ou seja, uma unidade de aumento em uma variável, causa um aumento proporcional na outra. Em contrapartida, se r for igual a -1, indica que uma unidade de aumento em uma variável, causa uma diminuição proporcional na outra. Se r for igual a zero, isso indica que não foi observada uma correlação linear entre as variáveis.

3.3.4.2 Teste de Correlação de Spearman

Segundo Gujarati (2011), o teste de Spearman pode ser descrito como um teste não paramétrico utilizado para avaliar a correlação entre duas variáveis. O teste mede a intensidade da reação de uma variável em resposta à variação de outra, ou seja, a regularidade com que uma variável aumenta ou diminui em relação à outra. Este coeficiente é útil para variáveis ordinais. Além disso, o coeficiente de Spearman converte os dados em uma sequência ordenada em forma de ranks e, é indicado para dados que não seguem uma distribuição normal ou para dados que não seguem relações lineares.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

No qual:

- ρ = coeficiente de correlação de Spearman;
- d_i = a diferença entre os ranks das duas variáveis para os pares de observações;
- n = número de observações.

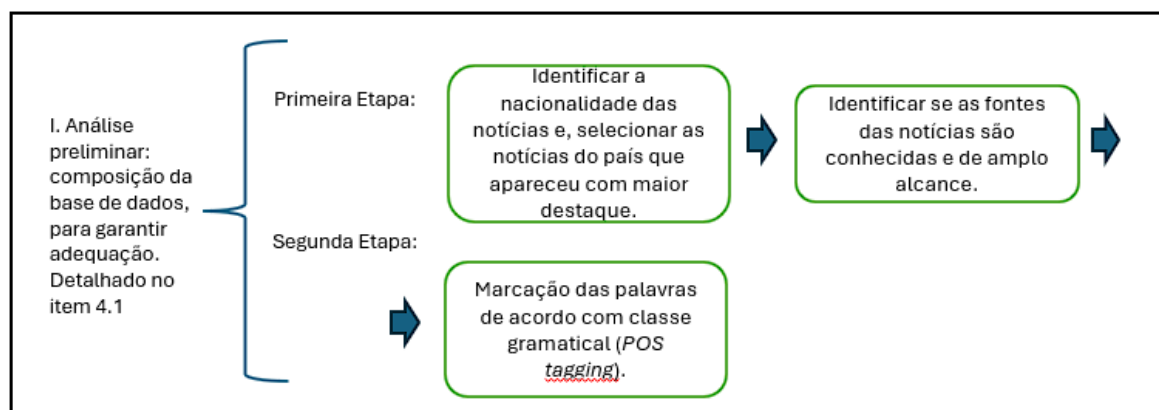
O teste de Spearman considera duas hipóteses para avaliação: a Hipótese Nula (H_0) considera que existe não uma correlação monotônica entre as variáveis, por outro lado, a Hipótese Alternativa (H_1) considera que existe uma correlação monotônica entre as variáveis.

Com relação a interpretação da estatística do teste, temos que ρ igual a 1, indica uma correlação monotônica perfeita positiva. Por outro lado, se ρ for igual a -1, teremos uma correlação monotônica perfeita negativa. Caso ρ seja igual a zero, isso indica que não foi observada uma correlação monotônica entre as variáveis.

4 ANÁLISE PRELIMINAR E ANÁLISE TEXTUAL DE NOTÍCIAS

Para avaliar a base de dados e saber se ela está adequada para o propósito deste trabalho, temos a análise preliminar do conteúdo, cujo processo está detalhado no item 4.1 e, é composta por duas etapas, conforme o fluxograma ilustrado na Figura 8 abaixo.

FIGURA 8 – FLUXOGRAMA: ANÁLISE PRELIMINAR

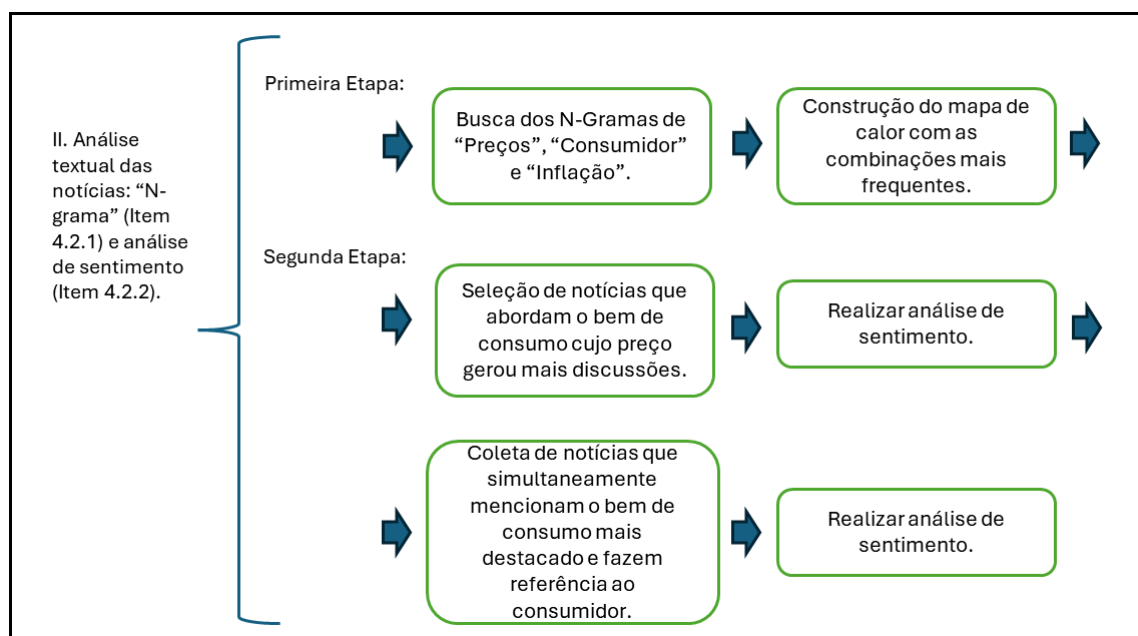


FONTE: elaborado pelo autor.

A primeira etapa foi identificar a nacionalidade das notícias que estavam presentes na base. O objetivo aqui, é que o trabalho se atenha as notícias apenas de um país, ou seja, das notícias cujo país apareceu com a maior frequência. Após essa seleção, teremos uma amostra de notícias inicial para ser manipulada. A partir dessa amostra, é feita a identificação das fontes dessas notícias, de modo a verificar se elas foram publicadas por veículos de comunicação reconhecidos mundialmente. Isso confere robustez a análise, pois se os veículos de comunicação são famosos, isso sustenta a ideia de que as informações são de fácil acesso aos consumidores, e, portanto, podem influenciar as percepções e posteriormente, as decisões de consumo e expectativas dos consumidores em relação à economia. De acordo com o fluxograma, a segunda etapa da análise preliminar, consistiu na aplicação da técnica de *POS tagging*, que permite a classificação das palavras conforme suas respectivas classes gramaticais. O objetivo dessa técnica, é identificar os temas mais frequentemente abordados pelos autores das notícias. Para isso, optou-se por focar na extração de substantivos, pois essa classe de palavras é fundamental para nomear os agentes econômicos, serviços e bens de consumo. Conforme os resultados que serão mostrados adiante, constatou-se que é uma base sólida para o propósito do trabalho. Então, foi

realizado, um processo de análise textual, conforme a Figura 9, que é composto por duas etapas de processamento de texto.

FIGURA 9 – FLUXOGRAMA: ANÁLISE TEXTUAL



FONTE: elaborado pelo autor.

Na primeira etapa da Análise Textual, temos a pesquisa por N-gramas envolvendo os termos “preços”, “consumidor” e “inflação”. Além disso, será construída uma representação gráfica sobre como os termos que surgiram com mais frequência, aparecem de forma correlacionada. Isso permitirá a identificação dos temas que foram mais discutidos pelos jornalistas envolvendo “preços” e, uma posterior análise sobre como este bem de consumo ou serviço, afeta as expectativas dos consumidores – que é justamente o primeiro objetivo do trabalho. Na segunda etapa, ocorre a aplicação da técnica de análise de sentimentos sobre as notícias que apareceram mais correlacionadas ao consumidor. O propósito é verificar qual foi o “framing” que os jornalistas atribuíram às informações e, se este é compatível com o que a teoria empírica econômica tradicional relata – sendo este o terceiro objetivo do trabalho.

4.1 ANÁLISE PRELIMINAR DA BASE DE DADOS

Conforme o fluxograma da Figura 8, a análise preliminar é composta tanto pela análise da origem das notícias, como pela análise dos substantivos mais utilizados nas notícias. A base de dados construída pela Aylien (2020), apresenta o nome dos autores das notícias, o texto (corpo) e título das notícias, a linguagem na qual a notícia foi escrita, assim como o domínio no qual ela está disponível e, também disponibiliza o link para que o pesquisador possa conferir o texto original da notícia.

A primeira etapa é uma investigação inicial para saber como a base de dados foi construída. Para isso, a partir dos dados disponíveis na própria base, foi contado o número total de notícias por país, e o resultado obtido foi que de cerca 1,6 Milhão de notícias, aproximadamente 500 mil notícias não estavam identificadas de acordo com o país de origem. Contudo, cerca de 1,1 Milhão de notícias estavam catalogadas de acordo com o país de origem e, descobriu-se que a maior parte das notícias, como demonstrado na Tabela 1, foram extraídas de portais mídia dos Estados Unidos (US), e em seguida da Grã-Bretanha (GB) e em terceiro lugar da Índia (IN).

TABELA 1 – CONTAGEM DE NOTÍCIAS POR PAÍS DE ORIGEM

Ordem	Código	País	Contagem
1	US	Estados Unidos	435.849
2	GB	Reino Unido	244.198
3	IN	Índia	228.670
4	CA	Canadá	59.643
5	AU	Austrália	43.496
6	PH	Filipinas	25.429
7	PK	Paquistão	24.017
8	IE	Irlanda	21.390
9	RU	Rússia	13.883
10	NZ	Nova Zelândia	13.230

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

A segunda etapa da análise preliminar, é a identificação da composição dos portais de notícias. Como este trabalho irá analisar notícias que mencionam “consumidor”, “inflação” e “preços”, foram pesquisados quais foram os portais de notícias que mais mencionaram esses três termos, o resultado segue na Tabela 2.

TABELA 2 – AS 20 FONTES DE NOTÍCIAS MAIS COMUNS

Fonte de Notícia	Número de notícias que mencionam...		
	Consumidor	Preços	Inflação
Forbes	4595	2221	389
Business Insider	2529	1927	218
CNBC	1460	1191	220
Market Watch	1073	1193	247
The Washington Post	853	767	126
CNN	621	674	34
Biz Journals	881	431	16
The Hill	544	451	32
SF Gate	407	498	78
USAToday.com	527	299	34
ABC News	312	249	50
USA Today	356	203	25
CBS Local	354	210	0
Fox News	308	225	29
New York Post	294	226	17
The Los Angeles Times	298	188	24
CBS News	301	127	15
The Chicago Tribune	250	122	13
Breitbart	149	133	28
Newsweek	168	100	19

FONTE: Elaborado pelo autor, a partir dos dados de Aylien (2020).

A partir disso, contatou-se que os objetos de estudo do trabalho foram mencionados por fontes de notícias de amplo alcance, uma vez que são fontes mundialmente conhecidas, cujas narrativas poderiam facilmente ser copiadas por outros portais ao redor do mundo, o que confere o poder de influenciar as expectativas dos consumidores que entrem em contato com o conteúdo econômico que é escrito por elas.

A seguir, será feita uma análise textual para identificar quais foram os agentes econômicos, bens de consumo ou serviços, assim como possíveis elementos de importância econômica, que apareceram com mais frequência juntamente aos objetos de estudo. A análise se concentrará exclusivamente em substantivos presentes na base de dados. O objetivo é identificar e classificar os principais agentes econômicos, bens e serviços mencionados nas notícias, utilizando apenas substantivos como indicadores. Essa escolha é fundamentada no fato de que os substantivos representam os elementos centrais do discurso econômico, permitindo uma compreensão clara dos tópicos mais relevantes, como consumidores, empresas e produtos, para entender se as palavras-chave foram contextualmente bem escolhidas. Para isso, foram aplicados três filtros na base de notícias que mencionavam “consumidor”, “preços” e “inflação”, visando analisar os substantivos que

apareceram com mais frequência. Abaixo, estão os resultados obtidos, revelando os substantivos mais recorrentes na composição das notícias e destacando os principais agentes econômicos, assim como os bens e serviços mais mencionados.

TABELA 3 – OS 10 SUBSTANTIVOS MAIS COMUNS

Os 10 substantivos mais frequentes em notícias que mencionam...					
O consumidor		Os preços		A inflação	
Palavra	Frequência	Palavra	Frequência	Palavra	Frequência
pessoas	43.448	preços	32.330	mercado	4.033
negócios	36.623	petróleo	29.610	ano	3.781
empresa	32.876	mercado	27.889	inflação	3.566
tempo	29.280	pessoas	24.532	economia	3.541
empresas	26.497	ano	20.207	pessoas	3.509
ano	25.600	tempo	19.667	taxas	3.050
consumidores	23.276	preço	18.223	tempo	2.995
saúde	22.823	negócio	16.825	dinheiro	2.651
mercado	22.426	empresa	16.165	anos	2.597
vendas	20.436	ações	16.135	preços	2.491

FONTE: Elaborado pelo autor, a partir dos dados de Aylien (2020).

Ao analisarmos os termos que aparecem com mais frequência nas notícias que mencionam o consumidor, o substantivo "pessoas" emerge com frequência, e sugere que os relatos das notícias se concentram amplamente nos comportamentos e experiências dos consumidores no mercado. A grande frequência de termos como "negócios" e "empresa" ressalta uma estreita conexão entre as ações dos consumidores e as estratégias empresariais, e indica um cenário onde as decisões dos consumidores têm uma influência direta.

Dentro da categoria de “Preços”, a dominância do termo "petróleo" destaca uma preocupação central com o custo dos bens, especialmente de matérias primas essenciais. A ênfase em "mercado" indica que grande parte da cobertura midiática, investiga as condições de mercado que moldam as tendências de preços. A repetida aparição de "pessoas" e "tempo" mais uma vez aponta para um foco em como os consumidores podem perceber as flutuações de preços ao longo do tempo, o que pode mudar padrões de consumo.

Por último, ao analisar a coluna “inflação”, termos como "mercado" e "ano" sugere que o fenômeno inflacionário está contextualizado sob uma perspectiva temporal e mercadológica. A presença de termos como "economia" e "taxas" sugere que as informações contidas nas notícias, não abordam apenas a inflação em si, mas também suas implicações

macroeconômicas, o que inclui possíveis variações nas taxas de juros. Novamente, a palavra "pessoas" reflete uma preocupação com o impacto direto da inflação na vida cotidiana, ressaltando a importância de compreender seus efeitos no mundo real.

É possível identificar uma interconexão entre o comportamento do consumidor, os preços e a inflação, que se origina das informações econômicas que envolvem tanto as pessoas, quanto as empresas ao longo do tempo. Portanto, a seguir, será feita a análise textual para entender como esses elementos econômicos se relacionam entre si dentro das notícias.

4.2 ANÁLISE TEXTUAL: BUSCA POR TERMOS MAIS RECORRENTES

O jornalismo econômico tende a ser uma fonte confiável de informação, e notícias de teor econômico entraram em alta devido à crise econômica instaurada durante o início da COVID-19 em 2020. Então, foi necessário fazer um levantamento de quais temas econômicos foram os mais discutidos neste período sob o contexto em que apareceram.

Conforme foi explicado anteriormente, para a análise N-grama, primeira etapa do processo de análise textual, foram selecionadas as palavras “preços” e “inflação”, por serem dois tópicos que moldam as expectativas dos consumidores, de modo a saber quais foram os termos, ou até mesmo expressões que apareceram conjuntamente com mais frequência. A importância desta pesquisa, está em saber quais foram as ideias mais disseminadas para essas duas variáveis econômicas. Além disso, foi feita uma pesquisa com o termo “consumidor”, para identificar quais foram as narrativas econômicas que estavam sendo propagadas pela mídia jornalística aos consumidores.

Como foi explicado na seção de metodologia, a técnica N-grama consiste em buscar quais foram as palavras ou expressões, que mais apareceram em conjunto dada uma palavra-chave. A seguir, serão mostrados os resultados das combinações que apareceram com mais frequência para “Inflação”, “Preços” e “Consumidor”.

Dentro da literatura econômica, a inflação é apontada como um dos principais indicadores macroeconômicos. Além disso, é uma das variáveis que os consumidores mais prestam atenção ao elaborarem expectativas econômicas. Abaixo, estão os bigramas ($n = 2$), mais frequentes dentro dos textos (corpos) das notícias:

TABELA 4 – INFLAÇÃO: N-GRAMAS (CORPO DAS NOTÍCIAS)

Trigrama (n = 3)		Quadrigrama (n = 4)		Hexagrama (n = 6)	
Expressão	Frequência	Expressão	Frequência	Expressão	Frequência
“ajustado pela inflação”	61	“proteção contra a inflação e”	38	“uma proteção contra a inflação e a moeda”	32
“inflação no”	55	“uma proteção contra a inflação”	36	“proteção contra a inflação e a desvalorização da moeda”	28
“taxa de inflação”	55	“contra a inflação e a moeda”	32	“como uma proteção contra a inflação e”	27
“proteção contra a inflação”	46	“inflação e desvalorização da moeda”	28	“contra a inflação e a desvalorização da moeda”	25
“contra a inflação e”	41	“a taxa de inflação”	24	“retorno da inflação que alterará o”	20
“inflação e moeda”	35	“retorno da inflação que irá”	20	“expectativas para a economia e a inflação”	18
“ajustado pela inflação”	34	“a economia e a inflação”	19	“um retorno crescente da inflação que”	17
“para a inflação”	30	“quando ajustado pela inflação”	18	“de um retorno crescente da inflação que”	17
“e inflação”	25	“de uma inflação crescente”	17	“lucrar com uma inflação crescente”	17

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Os trigramas “ajustado pela inflação” (61) e “taxa de inflação” (55) sugerem uma preocupação com o poder de compra. Além disso, a expressão “proteção contra a inflação” (46) indica a necessidade de medidas para enfrentar os efeitos da inflação sobre consumidores e investidores. A presença de “proteção contra a inflação e a moeda” tanto nos quadrigramas (n = 4) quanto nos hexagramas (n = 6) reforça a preocupação com a desvalorização da moeda. A expressão “expectativas para a economia e a inflação” destaca a importância de entender como as percepções futuras sobre a inflação e, a inflação corrente, estão interligadas ao desempenho econômico futuro. Portanto, as narrativas econômicas discutidas sobre a inflação sugerem uma preocupação em mitigar seus efeitos nas expectativas dos agentes econômicos, assim como nas percepções sobre a desvalorização da moeda. A seguir, estão os trigramas para os títulos que mencionam “inflação”.

TABELA 5 – INFLAÇÃO: N-GRAMAS (TÍTULO DAS NOTÍCIAS)

Trigrama (n = 3)	
Expressão	Frequência
“salto na inflação”	2
“preocupado com a inflação”	2
“inflação no”	2
“sem inflação se”	2
“não haverá inflação”	2
“inflação se o”	2
“combater a inflação do coronavírus”	1
“Powell minimiza a inflação”	1
“minimiza ameaça de inflação”	1
“ameaça de inflação”	1

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Ao contrário dos resultados obtidos para o corpo da notícia, os títulos não apresentaram um padrão de repetição na forma como a inflação é escrita. Isso significa que os autores das notícias, utilizaram uma ampla gama de expressões para se referir a inflação. A seguir, serão explorados os trigramas de preços, que permitem uma análise das combinações mais frequentes. Este termo foi escolhido, dada a importância dentro do cotidiano do consumidor, no qual o preço é a medida de comparação para que os consumidores possam tomar uma decisão. Dentro do contexto macroeconômico, os preços são incorporados dentro de indicadores, que posteriormente dão origem a formulação de políticas econômicas. A seguir, estão os resultados para os N-Gramas encontrados para “preços”.

TABELA 6 – TERMOS RECORRENTES: “PREÇOS”

Bigrama (n = 2)		Trigrama (n = 3)		Quadrigrama (n = 4)	
Expressão	Frequência	Expressão	Frequência	Expressão	Frequência
“preços do petróleo”	5.217	“preços do petróleo bruto”	390	“os preços eram precisos em”	259
“preços das ações”	923	“os preços do petróleo têm”	374	“os preços acima refletem”	217
“preços de gasolina”	705	“preços do petróleo e”	349	“os preços acima refletem o”	215
“preços de casas”	626	“como os preços do petróleo”	271	“queda nos preços do petróleo”	118
“preços do ouro”	604	“os preços do petróleo estão”	260	“preços e disponibilidade sujeitos”	104
“preços do petróleo bruto”	479	“os preços eram precisos”	259	“preços do gás natural”	98
“os preços subiram”	279	“os preços acima refletem”	217	“preços da dívida do governo dos EUA”	98
“preços de medicamentos”	269	“preços baixos do petróleo	203	“os preços do petróleo têm”	88
“preços de títulos”	241	“preços do petróleo para”	190	“os preços eram precisos”	84
“preços da energia”	228	“os preços foram”	180	“os preços da dívida pública eram”	83

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Ao pesquisar os bigramas e quadrigramas mais comuns, notou-se que os termos que apareceram com mais frequência foram os que se referem a matérias primas, como o petróleo, a gasolina e o ouro. Para os bigramas (n = 2), temos que o resultado mais frequente é “preços de petróleo”, com 5.217 ocorrências, o que revela o foco das notícias sobre a discussão deste item. Outras expressões como “preços de ações” e “preços de gasolina” – com 923 e 705 ocorrências respectivamente –, foram amplamente mencionadas neste período. Em seguida, temos os trigramas (n = 3), no qual a maioria dos resultados na tabela, contém expressões como “preços de petróleo bruto” e “preços baixos do petróleo”, o que reforça o destaque desse item para a discussão econômica jornalística. Ao avançarmos para os quadrigramas (n = 4), vemos que a “queda nos preços do petróleo”, reforça novamente a importância da discussão sobre a variação de preços da matéria prima. Segundo Quintas, Fona e Lima (2020), houve um choque nos preços de petróleo no início de 2020, que foi decorrente de uma acentuada queda na demanda provocada pela menor utilização da gasolina. Além disso, a queda na demanda foi acompanhada por um excesso de oferta, impulsionada por uma guerra de preços entre a Arábia Saudita e a Rússia no âmbito da

OPEP+, com a Arábia Saudita aumentando sua produção. Abaixo, na Tabela 7, temos os resultados obtidos para a análise dos títulos das notícias.

TABELA 7 – PREÇOS: N-GRAMAS (TÍTULO DAS NOTÍCIAS)

Bigrama (n = 2)		Trigrama (n = 3)	
Expressão	Frequência	Expressão	Frequência
“preços do petróleo”	478	“como os preços do petróleo”	58
“preços do gás”	128	“ações: preços do ouro”	28
“preços do ouro”	119	“os preços do petróleo são”	26
“preços caem”	92	“preços do petróleo caem”	23
“preços de casas”	55	“preços do petróleo nos EUA”	22
“preços despencam”	53	“os preços do petróleo poderão”	20
“os preços sobem”	34	“preços do petróleo despencam”	20
“preços ao consumidor”	31	“movimentadores: preços do petróleo”	20
“preços do petróleo bruto”	25	“preços do petróleo nos EUA”	17
“os preços caem”	23	“preços do petróleo caem”	14

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Os títulos das notícias referentes aos preços, também refletem a predominância das discussões sobre os preços de petróleo em relação a outros assuntos. Portanto, pode-se dizer que a principal narrativa econômica encontrada sobre preços, foi sobre os choques de petróleo que ocorreram no período. É necessário verificar quais termos foram associados ao consumidor com mais frequência, de modo a se ter uma visão panorâmica sobre como este agente econômico foi retratado pelos jornalistas. A seguir, estão os resultados da análise de n-gramas relacionados a “consumidor”, realizada no corpo das notícias.

TABELA 8 – TERMOS RECORRENTES: “CONSUMIDOR”

Bigrama (n = 2)		Trigrama (n = 3)		Quadrigrama (n = 4)	
Expressão	Frequência	Expressão	Frequência	Expressão	Frequência
“gastos do consumidor”	1.550	“proteção financeira do consumidor”	253	“agência de proteção financeira ao consumidor”	148
“para os consumidores”	892	“o consumidor financeiro”	225	“a lei nacional do consumidor”	136
“demanda do consumidor”	835	“nos gastos do consumidor”	218	“a federação de consumidores de”	69
“confiança do consumidor”	802	“consumidores e empresas”	204	“o consumidor e o varejo”	60
“comportamento do consumidor”	524	“demanda do consumidor por”	201	“o futuro do consumidor”	60
“sentimento do consumidor”	375	“mudanças no consumidor”	154	“indústrias de consumo e varejo”	58
“proteção ao consumidor”	357	“o consumidor nacional”	151	“abrangendo o consumidor e”	57
“mais consumidores”	355	“direito nacional do consumidor”	150	“colunista que cobre o consumidor”	57
“os consumidores podem”	338	“consumidor e negócios”	143	“indicado para segurança do consumidor de Trump”	56
“produtos de consumo”	311	“no comportamento do consumidor”	126	“assuntos e proteção ao consumidor”	55

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Ao explorar quais foram os termos ou expressões mais frequentes para “consumidor”, temos que combinações como “gastos do consumidor” (1.550), “demanda do consumidor” (835) e “confiança do consumidor” (802), sugerem que os autores das notícias estavam focados em analisar como os gastos e demanda dos consumidores estavam mudando. Além disso, a frequente ocorrência de “confiança do consumidor”, sugere que os jornalistas estavam interessados em discutir as expectativas dos consumidores em relação a economia.

Em seguida, ao olharmos para os trigramas mais frequentes, as ocorrências de “proteção financeira do consumidor” (253) e “o consumidor financeiro” (225), sugerem uma preocupação em saber se os consumidores estão tomando medidas para se proteger financeiramente. Por fim, ao avançarmos para os quadrigramas, expressões como “agência de proteção financeira ao consumidor” (148), e “a lei nacional do consumidor” (136), indicam a discussão de regulamentações de proteção financeira para o consumidor. A expressão “o futuro do consumismo” (60), pode sugerir reflexões a respeito de tendências de

longo prazo e, sobre como o consumo poderia evoluir após a pandemia. A seguir, temos a Tabela 9 com os resultados de análise para os títulos das notícias.

TABELA 9 – CONSUMIDOR: N-GRAMAS (TÍTULO DAS NOTÍCIAS)

Bigrama (n = 2)		Trigrama (n = 3)	
Expressão	Frequência	Expressão	Frequência
“gastos do consumidor”	70	“confiança do consumidor dos eua”	16
“confiança do consumidor”	67	“gastos do consumidor dos EUA”	15
“preços ao consumidor”	31	“os gastos do consumidor são”	9
“sentimento do consumidor”	27	“preços ao consumidor caem”	9
“consumidor dos eua”	17	“como os consumidores permanecem”	8
“demanda do consumidor”	17	“confiança do consumidor dos eua”	7
“comportamento do consumidor”	16	“os gastos do consumidor caem”	6
“gabinete do consumidor”	8	“para ajudar os consumidores”	5
“mais consumidores”	8	“confiança do consumidor aumenta”	5
“mudança do consumidor”	8	“a confiança do consumidor cai”	5

FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Ao olharmos os resultados da Tabela 9, encontramos que “gastos do consumidor”, com 70 ocorrências, também foi o principal resultado encontrado para o corpo das notícias. Ao compararmos os resultados encontrados para os corpos das notícias, com os resultados acima, constatou-se que existe uma homogeneidade das expressões que foram encontradas, porém nos títulos, elas apresentam um menor volume de ocorrências. Além disso, parece existir uma grande preocupação com a “confiança do consumidor”, resultado que aparece com 67 ocorrências nos bigramas e predomina as expressões encontradas para os trigramas.

4.2.1 Resultados de Co-ocorrência de Termos em Notícias

Foi criado um mapa de calor para visualizar os bigramas econômicos mais frequentes, com o objetivo de identificar com que frequência eles ocorrem simultaneamente. Isso nos ajudará a compreender melhor como determinados temas foram discutidos.

Além disso, foram arbitrados alguns itens considerados relevantes para o consumidor, utilizando conceitos da composição do Índice de Sentimento do Consumidor (ISC). O ISC é uma métrica essencial que reflete as expectativas de consumidores em relação ao estado presente e, futuro da economia de um país. O ISC é calculado através das respostas de 5 perguntas, que avaliam as percepções dos consumidores sobre: as condições econômicas, as próprias finanças pessoais, os preços de bens de consumo essenciais, assim

como as condições de compra de bens duráveis (University of Michigan, 2024). É através destes componentes que os consumidores criam expectativas econômicas.

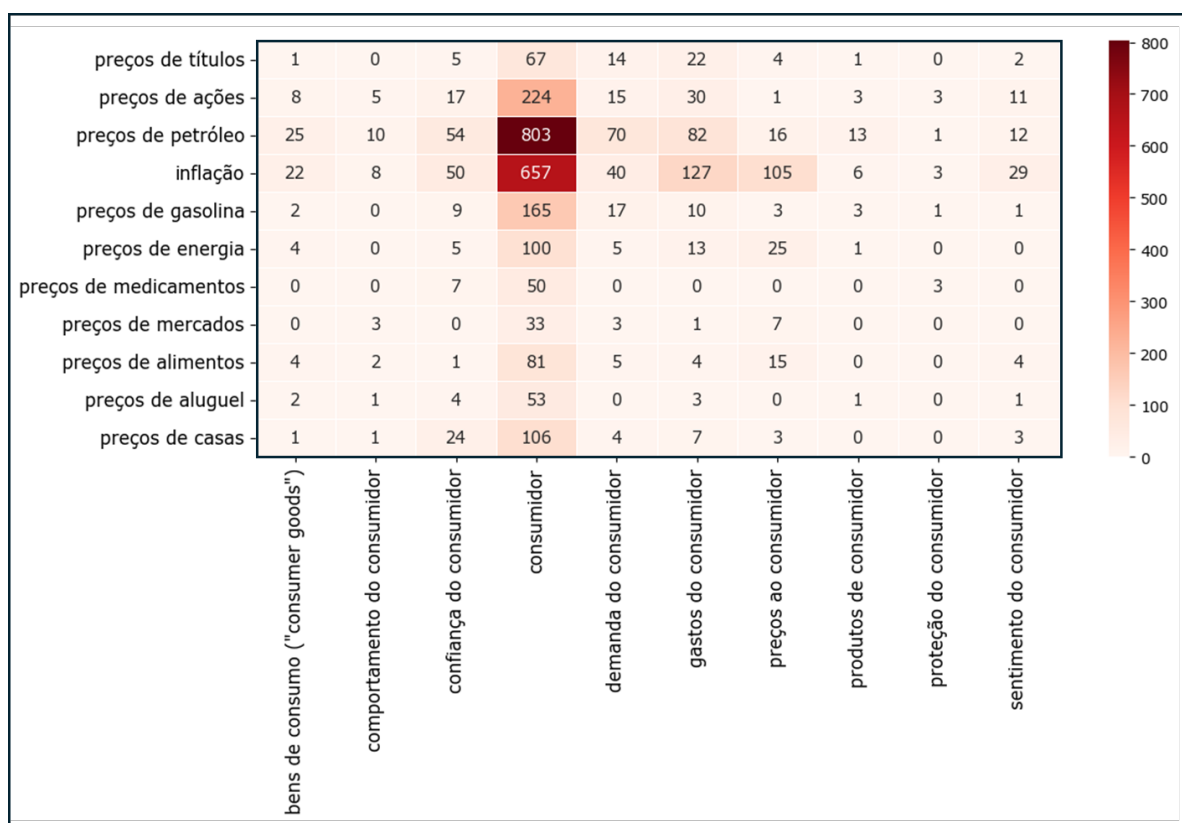
Os itens selecionados para análise – preços de alimentos, preços de aluguel e preços de mercados, – foram considerados itens críticos porque moldam as expectativas dos consumidores. Cada um destes itens impacta diretamente a percepção de estabilidade econômica e o custo de vida, que são elementos fundamentais para a formação do sentimento do consumidor.

No eixo vertical, estão os itens que aparecem com maior frequência dentro da categoria de preços, assim como itens que são do interesse do consumidor. O aumento nos preços de itens essenciais – como os de alimentos –, resulta em uma perceptível redução do poder de compra, o que leva a diminuição da confiança do consumidor. Ambos são itens frequentemente monitorados para avaliar o impacto direto na inflação percebida e esperada pelos consumidores. De acordo com USDA ERS (2022, apud Hsu, 2024) levando em conta que consumidores chegam a gastar cerca de 10% da renda em alimentação, mas para indivíduos de baixa renda, esse valor pode chegar a um terço do total da renda.

Também foram selecionados os “preços de aluguel”, assim como “preços de casas”, pois as flutuações nesses preços, afetam a segurança financeira dos consumidores no longo prazo, influenciando decisões de compra futuras, a poupança e a disposição para investir em maiores aquisições. A categoria “preços de mercado” foi inserida com o intuito de capturar notícias que mencionem os preços de mercearias (*grocery prices*), ou seja, os preços dos estabelecimentos nos quais os consumidores realizam suas compras domésticas.

Cada unidade representa uma notícia na qual os termos em questão apareceram simultaneamente. Optou-se por utilizar apenas “inflação” no cálculo de co-ocorrência, porque os resultados de bigramas para “inflação” possuíam uma baixa diferenciação entre si. Além disso, também foi inserido o termo “consumidor”, com o objetivo de ampliar os resultados, pois o eixo horizontal do diagrama contém expressões que mencionam apenas o consumidor.

FIGURA 10 – FREQUÊNCIA DE CO-OCORRÊNCIA DE BIGRAMAS E TERMOS



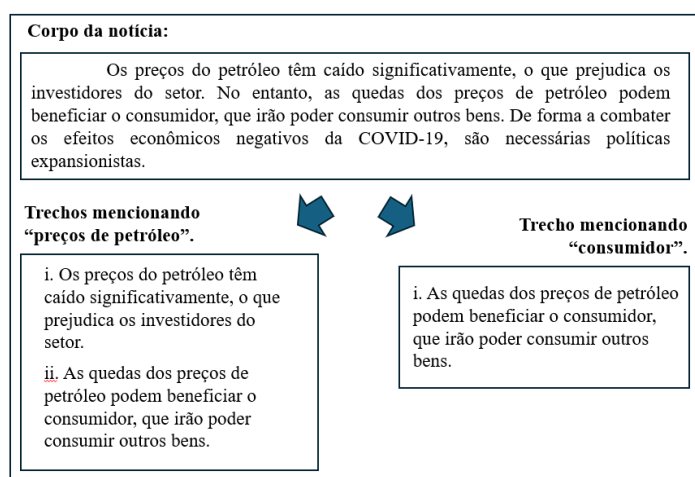
FONTE: elaborado pelo autor, a partir dos dados de Aylien (2020).

Com base nos resultados auferidos na figura acima, constatou-se a importância de utilizar a análise de sentimentos para entender qual o sentimento médio expresso pelas notícias que associam o consumidor aos preços de petróleo. A importância dessa análise visa identificar qual o enquadramento das informações relacionadas aos preços de petróleo que os jornalistas criaram para o consumidor, dado que estas viralizaram. Em uma entrevista para a CNBC (2024), Shiller afirmou que os preços de petróleo podem mudar o comportamento do consumidor e o mercado, porque historicamente, importantes ciclos econômicos estavam associados com os preços do petróleo. O autor justifica que as maiores recessões contemporâneas dos Estados Unidos, foram durante os dois choques de petróleo. Além disso, os choques de petróleo tiveram consequências significativas para os consumidores não apenas dos Estados Unidos, mas mundialmente.

4.3 ANÁLISE TEXTUAL: ANÁLISE DE SENTIMENTOS

Para as 803 notícias que mencionam tanto “preços de petróleo” quanto “consumidor”, foi realizada uma análise textual mais profunda, para compreender o contexto em que essas expressões estavam inseridas. O processo envolveu a extração das frases contendo a expressão “preços de petróleo” e aquelas que mencionam o termo “consumidor” dessa amostra de notícias. Para isso, foi desenvolvido um algoritmo capaz de capturar os trechos das notícias que incluíam essas duas expressões, conforme ilustrado no fluxograma apresentado na Figura 11 abaixo:

FIGURA 11 – EXTRAÇÃO DE FRASES DO CORPO DA NOTÍCIA



FONTE: elaborado pelo autor.

É importante ressaltar casos como o do exemplo acima, no qual o texto indicado como “Corpo da notícia”, contém expressões que mencionam “preços de petróleo” com “consumidor” em uma mesma frase. Para casos assim, não foi feito um corte total entre os trechos que mencionam “preços de petróleo” e “consumidor” em uma mesma frase, porque isso alteraria significativamente o sentido original da frase e, ignorar a existência de frases que mencionam os objetos de pesquisa simultaneamente, não seria realista. Outro aspecto importante de ressaltar, é que se uma notícia menciona “preços de petróleo” e “consumidor” mais de uma vez dentro do texto, todos esses trechos são extraídos, de forma a aumentar a precisão da análise. Após isso, foi feita a análise de sentimentos VADER para cada uma dessas amostras, cujos resultados estatísticos estão na Tabela 10, a seguir:

TABELA 10 – RESULTADOS DE ESTATÍSTICA DESCRITIVA

Estatística Descritiva	Preços de Petróleo	Consumidor	Corpo das Notícias	Título das Notícias
Média	-0,264	-0,013	-0,275	-0,159
Desvio Padrão	0,488	0,507	0,870	0,383
Curtose	0,525	0,040	0,591	0,132
Assimetria	-0,623	-1,002	-1,529	-0,565

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

A média dos sentimentos para as frases que mencionam 'preços de petróleo' é negativa (-0,26), enquanto a média dos sentimentos observados para 'consumidor' também é negativa (-0,01), mas próxima de zero, o que sugere que, em média, os jornalistas mantiveram neutralidade ao mencionar o consumidor.

Ao analisarmos os resultados de desvio padrão, temos que para os títulos das notícias, foram registrados os menores valores, o que indica que as observações desta amostra estão mais concentradas em torno da média. Para os trechos das notícias que mencionam “preços de petróleo” e para os trechos que mencionam “consumidor”, os valores indicam uma ligeira variação em relação à média, o que indica que as opiniões podem ser mais diversas. Contudo, ao analisarmos os resultados obtidos para os corpos das notícias, temos um elevado desvio padrão, o que indica a presença de sentimentos muito positivos ou muito negativos, consequente de uma variedade de opiniões.

A curtose foi positiva para todas as amostras, contudo ela foi maior para os corpos das notícias e para os trechos que mencionam “preços de petróleo”, o que indica que a concentração dos dados está em torno da média, porém existe uma notável presença de valores extremos. Por fim, temos os resultados de assimetria negativa para todas as amostras, o que revela que a maioria dos dados observados é maior que a média, mas existem algumas observações com valores menores que estão puxando a média para baixo.

Ao levar em conta o contexto do trabalho, avaliar a normalidade amostral das pontuações de sentimentos é importante, porque ao verificar as assimetrias negativas na distribuição das opiniões, juntamente com as médias negativas, isso indica que há uma concentração de sentimentos negativos moderados em relação ao que foi dito sobre os preços de petróleo, ou sobre as menções aos consumidores nas notícias. Essa assimetria à esquerda, considerando a escala (-1 a +1) que inclui valores negativos e positivos, sugere que os textos tendem a apresentar um viés mais pessimista quando abordam esses temas, com uma concentração de valores negativos moderados e alguns valores ainda mais

negativos que formam a cauda à esquerda da distribuição. Este padrão de distribuição revela uma tendência consistente de cobertura com tom mais crítico ou desfavorável sobre estes assuntos nas notícias analisadas.

Para testar a normalidade das amostras, foram utilizados os Testes de Jarque-Bera, assim como o teste de Shapiro-Wilk. Ambos os testes avaliam as mesmas hipóteses: a hipótese nula (H_0), considera que a amostra possui distribuição normal, por outro lado, a hipótese alternativa (H_1) considera que a amostra não possui distribuição normal.

Foi adotado um nível de significância de 5%. Abaixo, a Tabela 11 com os resultados dos testes.

TABELA 11 – RESULTADOS: TESTES DE NORMALIDADE

Testes de Normalidade	Preços de Petróleo (P-Valor)	Consumidor (P-Valor)	Corpo das Notícias (P- Valor)	Título das Notícias (P- Valor)
Jarque-Bera	$1,473 \times 10^{-11}$	$4,611 \times 10^{-8}$	$7,105 \times 10^{-28}$	0,001
Shapiro-Wilk	$9,326 \times 10^{-16}$	$5,013 \times 10^{-12}$	$2,922 \times 10^{-35}$	$1,929 \times 10^{-13}$

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

Através dos resultados de P-Valor obtidos para os respectivos testes, é possível afirmar que em todos os casos, o P-valor foi menor que o nível de significância. Portanto, a hipótese nula (H_0) é rejeitada e conclui-se que nenhuma amostra possui distribuição normal. Apesar das amostras não seguirem uma distribuição normal, a utilização do coeficiente de correlação de Pearson pode indicar correlações escondidas (Rovetta, 2020). A partir daqui, serão feitos testes entre a amostra que contém os trechos que principalmente mencionam “preços de petróleo” e a amostra que contém trechos que mencionam o “consumidor” (Teste 1), assim como são feitos os testes entre os corpos das notícias e o título (Teste 2). Abaixo, na Tabela 12, estão presentes os resultados da correlação de Pearson entre as amostras.

TABELA 12 – RESULTADOS: CORRELAÇÃO DE PEARSON

Teste de Correlação	Teste 1: Entre Preços de Petróleo e Consumidor	Teste 2: Entre Corpo da Notícia e Título da Notícia
Correlação de Pearson	0,117	0,378

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

Para os dois testes de correlação de Pearson, foi identificado uma baixa correlação linear positiva. Por isso, o segundo teste de correlação a ser calculado, será o de Spearman,

que é um teste robusto para avaliar dados que não possuem normalidade, como é o caso das amostras. O teste de Spearman considera duas hipóteses: a Hipótese Nula (H_0) considera que não existe uma correlação monotônica entre as duas variáveis, por outro lado, a Hipótese Alternativa (H_1) considera que existe uma correlação monotônica entre as duas variáveis. Foi adotado um nível de significância de 5%. Os resultados seguem conforme a Tabela 13 abaixo:

TABELA 13 – RESULTADOS: TESTE DE SPEARMAN

Testes de Correlação	Teste 1: Entre preços de petróleo e consumidor		Teste 2: Entre corpo das notícias e título das notícias	
	Estatística do Teste	P-Valor	Estatística do Teste	P-Valor
Correlação de Spearman	0,120	0,0006	0,39	$1,055 \times 10^{-30}$

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

Através dos testes de Kendall e Spearman, identificou-se um p-valor menor que 0.05, para ambos os testes, o que indica que existe uma forte correlação entre as variáveis, por mais que ela não seja linear. Portanto, rejeita-se a Hipótese Nula (H_0), pois existem evidências de que ambos os testes apresentam correlação monotônica.

Com o objetivo de testar a igualdade entre as variâncias das amostras, foram calculados o Teste F de Fisher e o Teste de Levene. Apesar das amostras não serem normalmente distribuídas, optou-se por fazer o Teste F para fins exploratórios, de modo a possibilitar uma comparação com os resultados do Teste de Levene, que é um teste robusto para desvios de normalidade amostral. Ambos os testes seguem as mesmas hipóteses, sendo que a Hipótese Nula (H_0) considera que as variâncias das amostras são iguais, por outro lado, a Hipótese Alternativa (H_1) considera que as variâncias das amostras são diferentes. A seguir, seguem os resultados na Tabela 14:

TABELA 14 – RESULTADOS: TESTE F E DE LEVENE

Testes de Hipóteses	Teste 1: Entre preços de petróleo e consumidor		Teste 2: Entre corpo das notícias e título das notícias	
	Estatística do Teste	P-Valor	Estatística do Teste	P-Valor
Teste de Levene	1,387	0,239	166,192	$2,964 \times 10^{-36}$
Teste F de Fisher	102,114	0,0005	12	0,0005

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

Com relação ao teste F, temos que o p-valor é menor que o nível de significância para os dois testes, o que implica na rejeição da Hipótese Nula (H_0), e conclui-se que as

variâncias são diferentes. Para o teste de Levene, temos que para o resultado para o Teste 1, o p-valor foi maior que o nível de significância e então não podemos rejeitar a Hipótese Nula (H_0). Além disso, utilizando Levene temos que no resultado para o Teste 2, o p-valor é menor que o nível de significância, nesse caso rejeitamos (H_0), pois existem evidências de que as variâncias são significativamente diferentes.

4.3.1 Resultados de análise de sentimento entre os trechos de texto

Para avaliar como as informações sobre os preços de petróleo foram enquadradas no comportamento do consumidor, com base no conceito do enquadramento (*framing*) comportamental de Kahneman (2012), foi feita uma nova análise dos dados. Para a amostra que contém os trechos das notícias que mencionam o consumidor, foram aplicados filtros para encontrar as expressões com os bigramas apresentados na Figura 9. Por exemplo, foi feito um filtro para “bens de consumo”, calculando-se o sentimento associado a essa expressão e comparando com o trecho que menciona “preços de petróleo”. Esse processo foi repetido para cada uma das expressões relacionadas ao consumidor, como: bens de consumo, comportamento do consumidor, confiança do consumidor, demanda do consumidor, gastos do consumidor, preços ao consumidor, produtos do consumidor, proteção do consumidor e sentimento do consumidor. A única exceção, foi “proteção do consumidor” por ter apresentado apenas uma observação e, por isso, não foi incluída. Esses resultados seguem o padrão mostrado na Figura 10.

FIGURA 12 – RESULTADOS DE ANÁLISE SENTIMENTOS POR EXPRESSÃO

Expressão que menciona o consumidor...	Número de Notícias	Para uma mesma notícia que mencione "preços de petróleo" e "consumidor", temos que....		Com relação ao sentimento transmitido pela notícia de forma geral...	
		Sentimento do trecho que menciona "Preços de Petróleo"	Sentimento do trecho que menciona o consumidor...	Corpo da notícia	Título da Notícia
Gastos do Consumidor	82	-0,233	-0,140	-0,507	-0,195
Demanda do Consumidor	70	-0,266	-0,378	-0,388	-0,203
Confiança do Consumidor	54	-0,261	0,400	-0,241	-0,132
Bens de Consumo	25	-0,517	0,236	-0,392	-0,125
Preços ao Consumidor	16	-0,426	-0,475	-0,345	-0,070
Produtos ao Consumidor	13	-0,348	-0,054	0,425	0,018
Sentimento do Consumidor	12	-0,366	0,121	-0,684	-0,098
Comportamento do Consumidor	10	-0,266	0,213	0,137	-0,150
Total	282				

FONTE: Elaborado pelo autor a partir dos dados de Aylien (2020).

Com base nos resultados apresentados na Figura 12, a média dos sentimentos associados à expressão "preços de petróleo" é consistentemente negativa em comparação aos sentimentos das expressões que mencionam o consumidor. Além disso, mesmo quando o consumidor é citado em um contexto positivo – o que acontece para “confiança do consumidor”, “bens de consumo”, “sentimento do consumidor” e “comportamento do consumidor” –, o impacto do trecho que contém "preços de petróleo" parece prevalecer no sentimento geral da notícia.

Embora as manchetes – os títulos das notícias – apresentem sentimentos próximos à neutralidade, a mensagem subjacente tende a transmitir uma perspectiva negativa ao consumidor. Isso contraria a expectativa econômica tradicional, que sugere que uma queda nos preços de petróleo deveria beneficiar o consumidor (Husain et al., 2015), gerando sentimentos mais positivos nas notícias em relação a itens como gastos, demanda e preços ao consumidor. Esse enquadramento negativo nas notícias sugere que, embora a queda nos preços de petróleo teoricamente possa trazer vantagens econômicas ao consumidor, a narrativa apresentada pela mídia não reflete o impacto positivo esperado.

5 AS NOTÍCIAS ECONÔMICAS REFLETIRAM AS PRINCIPAIS PREOCUPAÇÕES DOS CONSUMIDORES?

Com base na análise de frequência dos N-gramas, conclui-se que a narrativa sobre os choques de petróleo dominou a produção de notícias, possivelmente ofuscando discussões mais relevantes para os consumidores. Portanto, é necessário compreender mais profundamente como quedas dos preços de petróleo influenciam a economia. Em seguida, será feita uma análise para saber quais bens de consumo foram mais afetados pela inflação em 2020, a fim de identificar possíveis mudanças na demanda e expectativas. E então, poderemos avaliar se a queda dos preços do petróleo naquele período foi de fato a principal preocupação dos consumidores naquele momento.

5.1 QUAL O EFEITO DE UMA QUEDA NOS PREÇOS DE PETRÓLEO SOBRE O CONSUMO?

De acordo com a análise econômica tradicional, existem dois canais que são afetados pela variação do preço do petróleo dentro de uma economia. O primeiro é o canal da renda discricionária, no qual uma queda do preço do petróleo afeta os gastos dos consumidores, que terão mais dinheiro disponível para gastar em bens de consumo não relacionados ao petróleo. Em contrapartida, quando o preço começa a aumentar, os consumidores tendem a diminuir os gastos em outros bens de consumo em função da despesa adicional em gasolina. (Melek; Vigfusson, 2021)

O segundo é o canal do produtor, no qual a queda nos preços do petróleo tende a diminuir as receitas de empresas petrolíferas. O que acontece é que, com a diminuição da renda dos produtores, estes terão um estímulo a consumir menos. Além disso, a queda de rentabilidade do setor pode fazer com que haja demissões nesse setor, e estes trabalhadores podem enfrentar obstáculos para se realocar dentro do mercado de trabalho, em função da dificuldade em encontrar empregos que sejam compatíveis com as habilidades técnicas que possuem. Apesar disso, ao desconsiderar uma possível rigidez de preços, a vantagem da queda nos preços de petróleo, é mais rápida ao beneficiar a renda discricionária dos consumidores do que as consequências negativas de empregabilidade dentro do setor petrolífero. Segundo Hamilton (2016, apud Melek e Vigfusson, 2021), existe um efeito distributivo entre os benefícios e prejuízos de forma desigual na população, uma vez que

mesmo que algumas pessoas percam renda – investidores ou trabalhadores no setor petrolífero -, outros ganham um poder de compra adicional.

Ao olharmos para um ponto de vista macroeconômico acerca dos efeitos de choques dos preços, temos que para países importadores de petróleo, segundo Husain et al. (2014), uma queda nos preços do petróleo poderia resultar em um aumento de atividade econômica devido à queda nas taxas de juros e do risco do país, o que consequentemente, diminuiria o custo de capital.

No entanto, para nações exportadoras, isso poderia representar o oposto. Durante a elevação dos preços do petróleo, é habitual que empresas ligadas ao setor petrolífero acumulem dívidas de financiamento. Isso implica que, à medida que o risco de crédito para países ou corporações cresce, as condições para refinar suas dívidas se deterioram, gerando um ciclo vicioso onde o refinanciamento se torna mais oneroso e eleva o risco de crédito. Ademais, quando as nações produtoras de petróleo apresentam superávits, é habitual que estas apliquem em mercados internacionais. Uma redução no superávit levaria a uma redução na oferta de crédito, bem como na disponibilidade de capital nesses mercados, elevando os custos de empréstimos e financiamentos. (Husain et al., 2014)

No entanto, segundo Melek e Vigfusson (2021) é possível argumentar que os efeitos de variações nos preços do petróleo sob o consumo diminuíram com o tempo. Dados mostrados no estudo indicam que desde 2014, houve uma queda dos gastos dos consumidores em gasolina, que se acentuou devido as condições de trabalho remoto no ano de 2020. Além disso, segundo Melek e Vigfusson (2021), isso implica no fator de que a manutenção do trabalho remoto pós-Covid, pode impulsionar os gastos de consumidores em outros bens e serviços.

5.2 QUAIS BENS DE CONSUMO FORAM MAIS AFETADOS PELA INFLAÇÃO?

A seguir, na Figura 13, temos a evolução do Índice de Preços ao Consumidor para os Estados Unidos, calculado pelo U.S. Bureau of Labor Statistics (2021). Este estudo é importante, porque permite identificar quais foram os itens que sofreram maior acréscimos de preços e consequentemente afetaram o poder de compra do consumidor, antes e durante o ano da pandemia – de 2016 a 2020.

FIGURA 13 – ÍNDICE DE PREÇOS AO CONSUMIDOR (2016 – 2020)

Categoria		Período 12/2019 - 12/2020	Período 12/2018 - 12/2019	Período 12/2017 - 12/2018	Período 12/2016 - 12/2017
Alimentos e Relacionados	Alimentos	3,9%	1,8%	1,6%	1,6%
	Alimentos em casa	3,9%	0,7%	0,6%	0,9%
	Outros alimentos em casa	3,9%	0,3%	0,2%	0,5%
	Alimentos fora de casa	3,9%	3,1%	2,8%	2,5%
	Carnes, aves, peixes e ovos	4,6%	2,3%	-0,4%	2,8%
	Laticínios e produtos relacionados	4,4%	2,4%	-0,1%	-0,5%
	Frutas e vegetais	3,2%	-1,3%	1,6%	1,5%
	Cereais e produtos de panificação	3,2%	0,3%	1,7%	-0,6%
	Bebidas não alcoólicas	4,4%	1,0%	1,4%	0,0%
Combustível	Gasolina	-15,2%	7,9%	-2,1%	10,7%
	Serviço de gás (canalizado)	4,1%	-3,5%	2,3%	4,7%
Despesas relacionadas a transporte	Carros e caminhões usados	10,0%	-0,7%	1,4%	-1,0%
	Seguro de veículos	-4,8%	0,0%	4,6%	7,9%
	Veículos novos	2,0%	0,1%	-0,3%	-0,5%
	Tarifas aéreas	-18,4%	1,7%	-2,6%	-4,0%
Saúde	Cuidados médicos	1,8%	4,6%	2,0%	1,8%
	Medicamentos com receita	-2,4%	3,0%	-0,6%	2,8%
	Serviços médicos	1,7%	1,4%	0,6%	-1,8%
	Serviços hospitalares	3,0%	3,0%	3,7%	5,1%
Energia	Energia	-7,0%	3,4%	-0,3%	6,9%
	Eleticidade	2,2%	-0,4%	1,1%	2,6%
Bens de consumo regulados	Bebidas Alcoólicas	2,8%	0,5%	1,8%	1,4%
	Tabaco e produtos de fumo	5,1%	5,5%	3,4%	6,5%
Outros bens e serviços essenciais	Comunicação	2,4%	0,7%	-1,8%	-4,9%
	Habitação	1,8%	3,2%	3,2%	3,2%
	Educação	1,4%	2,1%	2,6%	2,0%
	Cuidados pessoais	1,8%	1,5%	1,8%	0,9%
	Vestuário	-3,9%	-1,2%	-0,1%	-1,6%
	Recreação/Lazer	0,9%	1,5%	1,2%	1,5%
	Móveis e operações domésticas	3,2%	1,0%	2,1%	-0,8%
Outros	Todos os itens, exceto alimentos e energia	1,6%	2,3%	2,2%	1,8%

FONTE: adaptado de U.S. Bureau of Labor Statistics (2021)

É possível identificar que de 2019 para 2020, houve um aumento de preços para todas as categorias relacionadas a alimentação, e em comparação com outros anos, nota-se que havia uma variabilidade nos níveis de preços entre os itens relacionados a alimentos. Além disso, em anos anteriores, ocorreu deflação em pelo menos uma das categorias de alimentos. Conforme o que será mostrado a seguir, esse aumento de preços nos alimentos seria sentido pelo consumidor, que passaria a mencionar os preços de alimentos espontaneamente durante a pesquisa do Índice de Sentimento do Consumidor (ISC) da Universidade de Michigan. Durante o ano de 2020, houve uma deflação para a gasolina, que foi consequência dos eventos econômicos explicados anteriormente. Além disso, também houve uma deflação dentro da indústria de energia, que foi afetada pelas quedas dos preços de petróleo. Com base nisso, foi analisado se o foco sobre a queda dos preços de petróleo, acabou ofuscando a importância da discussão sobre os aumentos de preços dos alimentos. De acordo com o relatório da pesquisa

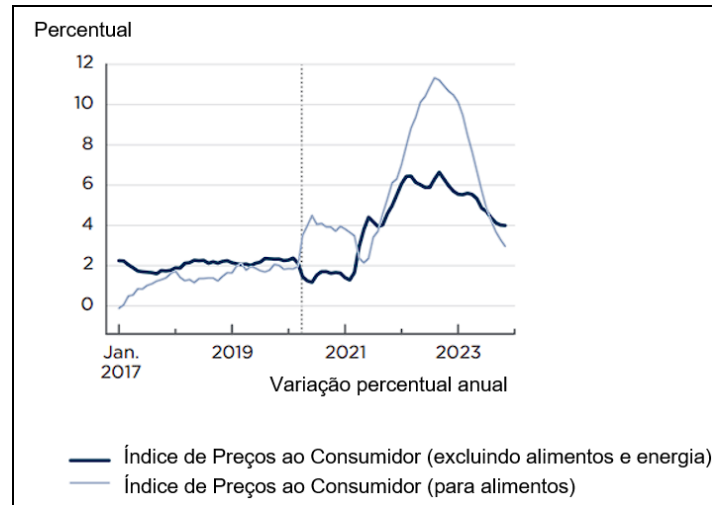
de consumidores da Universidade de Michigan, de junho de 2020, Curtin (2020) estabelece que as expectativas de inflação formadas naquele período eram reflexo de um aumento nos preços dos alimentos. Além disso, o autor observa que os consumidores ajustam suas expectativas inflacionárias com base nos preços dos itens de mercado. No relatório seguinte, de julho de 2020, Curtin (2020) aponta que houve um aumento nos preços dos alimentos logo após o início da pandemia, sendo que as maiores expectativas de inflação foram registradas entre as famílias que destinam a maior parte de seus orçamentos ao consumo de alimentos.

5.2.1 Qual é a relação entre os preços de gasolina e de alimentos?

Segundo dados do U.S. Energy Information Administration (2022), no início do ano de 2021, os preços do petróleo começaram a subir devido ao aumento da demanda, resultante da flexibilização das restrições da quarentena, quando as pessoas voltaram a usar mais seus veículos e a consumir mais gasolina. Apesar disso, um item que adquiriu peso dentro do consciente coletivo dos consumidores, foram os preços de alimentos. Ao olharmos para o mapa de calor na Figura 10, observa-se que a categoria “preços de alimentos”, apareceu em conjunto com “consumidor”, em 81 notícias. Este valor foi muito mais abaixo do esperado, uma vez um estudo do Centro de Pesquisa de Michigan da autoria de Hsu (2024), mostrou que os preços de alimentos tiveram uma alta decorrente da pandemia durante o ano de 2020 e, essa categoria foi mencionada espontaneamente em entrevistas de pesquisa pelos consumidores durante 2021.

A seguir, na Figura 14, estão os dados a respeito da inflação de preços de alimentos em comparação com o IPC-Central (ou IPC-Núcleo), que segundo o U.S. Bureau of Labor Statistics (2024), é o Índice de Preços ao Consumidor que desconsidera os preços de alimentos e gasolina, por serem considerados altamente voláteis.

FIGURA 14 – COMPARAÇÃO ENTRE “IPC-NÚCLEO” E “IPC-ALIMENTOS”

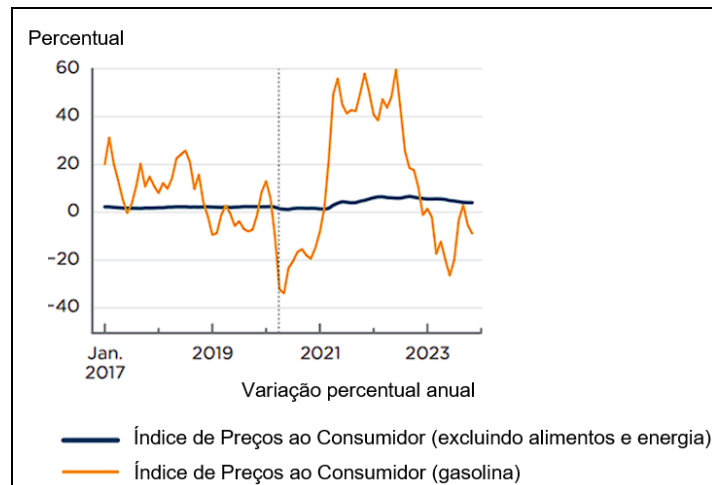


FONTE: Gráfico adaptado de HSU (2024).

NOTA: Dados da Universidade de Michigan, Survey Research Center, Survey of Consumers.

A partir da Figura 14, é possível ver que no ano de 2020, houve um aumento do índice de preços de alimentos de 2% ao final do ano de 2019, para acima de 4% em 2020. Apesar disso, no ano de 2020, como demonstrado na Figura 15, houve a deflação dos preços de gasolina.

FIGURA 15 – EVOLUÇÃO ENTRE “IPC-NÚCLEO” E “IPC-GASOLINA”



FONTE: Gráfico adaptado de HSU (2024).

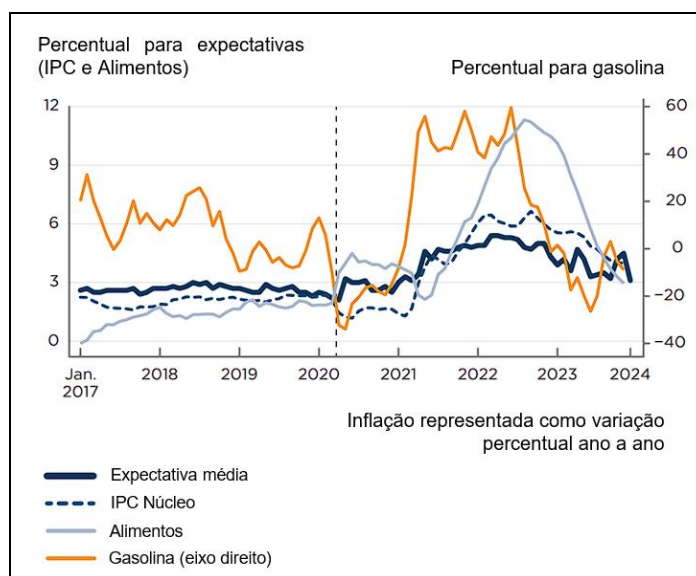
NOTA: Dados da Universidade de Michigan, Survey Research Center, Survey of Consumers.

No gráfico da Figura 15, a curva laranja mostra que os preços de gasolina variaram de forma significativa, principalmente em 2021, seguidos de quedas em 2022 e 2023, o que sugere que a inflação relacionada a indústria de petróleo, estava sensível às condições econômicas. Por outro lado, a curva azul no gráfico demonstra que, apesar da instabilidade

dos preços de gasolina, o núcleo da inflação – excluindo alimentos e energia – manteve-se estável.

Além disso, temos a Figura 16, que mostra a sobreposição da evolução histórica dos preços de gasolina (eixo direito), com a evolução da expectativa de inflação, assim como o IPC-Central, e a inflação dos preços de gasolina.

FIGURA 16 – EXPECTATIVA E INFLAÇÃO OBSERVADA (2017 – 2023)



FONTE: Gráfico adaptado de HSU (2024).

NOTA: Dados da Universidade de Michigan, Survey Research Center, Survey of Consumers.

A partir da Figura 16 e dos resultados apresentados até o momento, é possível afirmar que o ano de 2020, foi caracterizado por uma mudança abrupta de demanda na economia dos Estados Unidos. Ao mesmo tempo que houve uma queda histórica dos preços de gasolina, houve um aumento expressivo do nível de preços dos alimentos.

De acordo com Armantier *et al.* (2017) é importante considerar as expectativas dos consumidores em relação ao futuro, pois são justamente as expectativas que influenciam decisões econômicas das famílias, como a aquisição de imóveis, bens de consumo duráveis, poupança ou demais investimentos. Estes fatores podem influenciar resultados macroeconômicos como a inflação presente e, as expectativas de inflação projetadas para o futuro. Portanto, ao realizar estimativas do sentimento dos consumidores, os resultados poderão ser utilizados para a formulação de políticas fiscais ou monetárias de um país. Com base nisso, o centro de Pesquisa de Consumidores da Universidade de Michigan, realiza o cálculo do Índice de Sentimento do Consumidor (ISC), que é utilizado por uma ampla gama de formuladores de políticas econômicas e pesquisadores.

A Tabela 15 foi composta a partir dos resultados de regressões econométricas estimadas no estudo de Hsu (2024). Ela mostra os efeitos do aumento de 1% na inflação da gasolina, na inflação dos preços de alimentos e na inflação central sobre três indicadores econômicos importantes: i) a variação anual esperada nas expectativas de inflação; ii) a variação nas expectativas de inflação de longo prazo; iii) a variação associada no Índice de Sentimento do Consumidor.

TABELA 15 – INFLAÇÃO E EXPECTATIVAS DO CONSUMIDOR

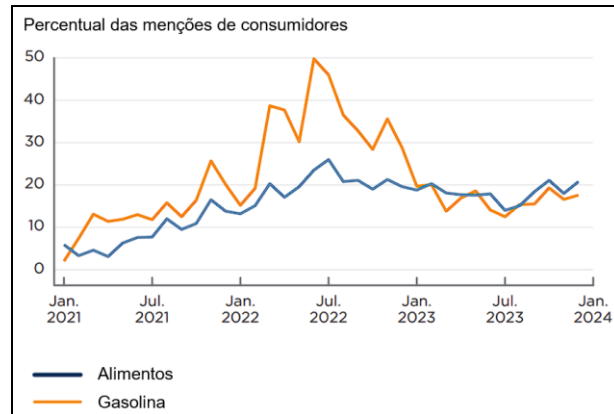
	Variação anual esperada de expectativas de inflação		Variação nas expectativas de inflação de longo prazo.		Variação associada no Índice de Sentimento do Consumidor (ISC)	
	1978-2023	2020-2023	1978-2023	2020-2023	1978-2023	2020-2023
Inflação de Gasolina	0,02 p.p.	0,01 p.p.	-0,00 p.p.	-0,00 p.p.	-0,06 p.p.	0,08 p.p.
Inflação de Alimentos	0,19 p.p.	0,06 p.p.	-0,10 p.p.	-0,02 p.p.	-1,82 p.p.	-2,06 p.p.
Inflação Central	0,38 p.p.	0,27 p.p.	0,49 p.p.	0,12 p.p.	-0,73 p.p.	-2,47 p.p.

FONTE: Adaptado de HSU (2024).

NOTA: “p.p.” significa pontos percentuais.

Com base nos resultados acima, constatou-se que historicamente, os preços de alimentos são uma preocupação maior para os consumidores do que os preços de gasolina, pois o ISC apresenta uma perceptível variação decorrente do aumento da inflação nos preços de alimentos. Além disso, outro aspecto importante para ser mencionado, é que o Centro de Pesquisa do Consumidor de Michigan, começou a detectar menções espontâneas acerca dos preços de gasolina, assim como dos preços de alimentos a partir de janeiro de 2021, conforme a evolução mostrada na Figura 17. Segundo Hsu (2024), os dados revelam que a maioria das menções, são comentários negativos a respeito dos preços de qualquer um desses dois bens.

FIGURA 17 – MENÇÕES: GASOLINA E ALIMENTOS (2021 – 2023)



FONTE: Gráfico adaptado de HSU (2024).

NOTA: Dados da Universidade de Michigan, Survey Research Center, Survey of Consumers.

Com base nos resultados apresentados até o presente momento, é possível dizer que temos evidências de uma falta de cobertura mais ampla nas notícias, para os itens que compõem a cesta de produtos do consumidor. A captura de menções sobre os preços de alimentos no estudo de Hsu (2024), mostra que o consumidor é sensível às variações de preços para os itens de consumo mais básicos. A Figura 17 mostra que as menções aos preços de gasolina aumentaram logo no primeiro semestre 2021, o que é coincidente com o aumento dos preços na gasolina, conforme a Tabela 16.

TABELA 16 – VARIAÇÃO MENSAL DE PREÇOS DA GASOLINA (2020-2021)

Ano 2020	Preço	Ano 2021	Preço	Variação
Janeiro	2,64	Janeiro	2,42	-8,19%
Fevereiro	2,53	Fevereiro	2,59	2,13%
Março	2,33	Março	2,90	24,43%
Abril	1,94	Abril	2,95	52,12%
Maio	1,96	Maio	3,08	56,86%
Junho	2,17	Junho	3,16	45,48%
Julho	2,27	Julho	3,23	42,21%
Agosto	2,27	Agosto	3,26	43,27%
Setembro	2,27	Setembro	3,27	43,89%
Outubro	2,25	Outubro	3,38	50,53%
Novembro	2,20	Novembro	3,49	58,68%
Dezembro	2,28	Dezembro	3,41	49,12%

FONTE: Elaborado a partir dos dados de U.S. Energy Information Administration (2024).

De acordo com os dados do U.S. Energy Information Administration (2024), os preços subiram drasticamente de fevereiro para março de 2021 e permaneceram elevados durante o ano. Esse aumento fez com que os consumidores mencionassem mais a gasolina, o que sugere

que, em 2020, as notícias sobre os preços do petróleo (que estavam baixos) podem ter desviado a atenção de discussões mais relevantes levando em conta as necessidades mais urgentes do consumidor naquele período.

5.3 DISCUSSÃO SOBRE RESULTADOS:

Os resultados encontrados ao longo deste trabalho, indicam que o predomínio da narrativa acerca dos preços de petróleo ofuscou a discussão sobre temas que teriam sido mais pertinentes para o consumidor durante a pandemia de COVID-19. Isso é reforçado pela pandemia ser uma grave crise de saúde pública, e por isso, também esperava-se uma maior presença de discussões de preços relacionadas a recursos médicos ou hospitalares – como máquinas, medicamentos e antissépticos como o Álcool em gel, que foi extensivamente utilizado durante 2020. Abaixo, está um trecho de uma entrevista de Shiller para a Bloomberg, na qual o economista alega que a narrativa da pandemia era mais crítica do que a narrativa sobre a queda dos preços de petróleo.

O que surgiu nas últimas semanas é uma narrativa sobre o vírus COVID, que sugere que isso poderia ser uma epidemia recorde, pelo menos na narrativa. [...] E isso tem deixado as pessoas realmente assustadas, sabe? [...] Acho que isso domina a narrativa do petróleo por um fator de 10. É realmente algo. E faz com que as pessoas queiram segurar os gastos e adiar novos investimentos e novos empreendimentos, o que é ruim para a economia. (Bloomberg Television, 2020, 1 min 16s, tradução nossa).

Com relação aos resultados de análise de sentimentos, temos que a média obtida das frases que mencionam preços do petróleo, induzem a um enquadramento (*framing*) negativo sobre as quedas dos preços, que como vimos anteriormente, tendem a beneficiar os consumidores. Outra característica observada nas notícias, é que mesmo quando os trechos que mencionam o consumidor apresentaram características positivas, este ainda era afetado pelos efeitos de enquadramento dos preços de petróleo, no qual quase todos os componentes restantes da notícia transmitem sentimentos negativos e propiciam uma péssima percepção para o consumidor.

No que se refere aos resultados de Hsu (2024), as menções dos consumidores sobre os preços de petróleo e sobre os preços de alimentos são reflexo da sensibilidade dos consumidores a volatilidade destes preços. Contudo, os preços de alimentos tendem a afetar ainda mais as expectativas dos consumidores, do que os preços de petróleo sobre o percurso da economia ao longo do tempo.

A contradição entre o medo midiático sobre a queda dos preços de petróleo e as consequências empíricas de quando isso acontece, fica evidente na mesma entrevista de Shiller para a Bloomberg, na qual o autor afirmou que:

É disruptivo ver esses baixos preços de petróleo. Por outro lado, altos preços de petróleo desencadearam uma das nossas maiores recessões econômicas, como as de 1974 e 1975, ou as de 1980 e 1982. Aquelas foram épocas de crises de petróleo. Agora temos petróleo abundante e barato, o que é uma boa notícia. (BLOOMBERG TELEVISION, 2020, 25s, tradução nossa).

A resposta do autor dá a entender que é estranho ver estes baixos preços de petróleo, porém ele ainda ressalta o saldo positivo da queda destes preços. A “boa notícia” faz a alusão de que a queda nos preços de petróleo seria um fator positivo para os consumidores. Porém, é notável ver como a narrativa dos preços de petróleo apareceu de forma dominante sobre outros assuntos, o que é característica do pensamento econômico de tentar formular previsões e políticas econômicas com base em acontecimentos passados, mas que ignora o contexto econômico atual.

6 CONCLUSÃO

Através da teoria de narrativas econômicas, a contribuição de Shiller (2019) reside em proporcionar um entendimento sintético de como o fluxo de informações influencia as dinâmicas econômicas. O caráter viral de algumas narrativas é um fator crítico, pois narrativas que são amplamente disseminadas tendem a desviar a atenção do público de questões socioeconômicas de maior relevância. Ao considerarmos esse fenômeno dentro do contexto de crises econômicas, como foi o caso da pandemia de COVID-19, ele se torna preocupante. Problemas urgentes, como o aumento dos preços de alimentos – ao considerar que a pandemia agravou a vulnerabilidade econômica de muitas famílias –, podem ser ofuscados por narrativas dominantes, como os choques de petróleo.

O estudo das narrativas econômicas vai além dos consumidores e abrange outros agentes econômicos. As empresas podem antecipar tendências de consumo e ajustar suas estratégias comerciais para serem mais eficazes em determinados momentos. Em relação aos investidores, um cenário de aumento nos preços dos alimentos impacta diretamente suas decisões sobre as alocações de recursos.

Os alimentos são geralmente considerados bens de consumo inelásticos, pois a demanda por eles varia pouco em relação ao preço, uma vez que são essenciais. Quando os preços dos alimentos aumentam significativamente, as famílias tendem a cortar gastos em bens de menor importância. A compra de alimentos não pode ser eliminada ou adiada. Empresas que operam fora do mercado alimentício — como as de serviços, entretenimento e turismo — são, muitas vezes, as primeiras a sentir os efeitos negativos de uma reorganização orçamentária familiar. Isso impacta não apenas suas receitas, mas também os investimentos que recebem, afetando, assim, outro agente econômico importante: o investidor.

Na esfera governamental, a compreensão das narrativas econômicas permite que gestores públicos desenvolvam políticas econômicas mais coerentes e eficazes. Essas políticas podem proteger tanto as empresas que sofrem com a queda de consumo quanto os consumidores mais vulneráveis aos aumentos nos preços dos alimentos. Isso reflete a necessidade de uma cobertura midiática que aborde as reais demandas dos consumidores, e não apenas narrativas clássicas que dominem o cenário econômico, como é o caso dos choques nos preços de petróleo.

A mídia jornalística desempenha um papel fundamental na formação de percepções do público e pode alterar o foco para questões que nem sempre estão alinhadas com as prioridades de políticas econômicas. Nesse sentido, a teoria sobre a heurística de

disponibilidade de Kahneman (2012) deve ser aplicada em âmbito macroeconômico: quanto mais uma narrativa estiver presente no discurso informacional e midiático, maiores são as chances de ela ser percebida como um problema central, independentemente da urgência de outros assuntos. Portanto, narrativas econômicas dominantes, como a queda dos preços de petróleo, podem desviar a atenção do público e do governo da necessidade de políticas públicas, uma vez que não são as necessidades mais prementes. Conforme Shiller (2019), os jornalistas prestam atenção no que viraliza em outros países para escrever notícias e, isso induz à possibilidade de narrativas econômicas dominantes influenciarem as discussões globais e alterarem debates econômicos em outros países.

Este trabalho contribui ao aplicar a análise de sentimentos, como uma técnica para mensurar os enquadramentos das narrativas econômicas relacionadas às expectativas dos consumidores. Shen e Zhang (2024) assim como Zhang, Yang e Liu (2023), entendem que a técnica é imperativa para a tomada de decisões dentro do mercado financeiro. Outros campos de aplicação da técnica incluem, de acordo com Cochrane et al. (2021), o âmbito da ciência política, como é mencionado no artigo de Haselmayer e Jenny (2016), que destaca o uso da técnica em debates parlamentares. Outro estudo que aponta a aplicação em contexto político é o de Kampić e Babać (2021), que realizaram uma análise de tweets de Donald Trump para entender a influência e os aspectos das opiniões controversas do político. Martín e Valentín (2023) também ressaltam a importância da técnica no contexto dos discursos políticos e, assim como Haselmayer e Jenny (2016), indicam que o campo ainda é amplamente dominado por pesquisas voltadas à língua inglesa, o que representa um desafio para estudos em outros idiomas.

Além disso, o trabalho ampliou as contribuições de Shiller, Katona e Kahneman acerca dos elementos comportamentais que podem afetar a economia, mas que não são comumente considerados pelas abordagens econômicas convencionais. Enquanto Shiller (2019) enfatiza como as narrativas econômicas moldam o discurso econômico, Katona (1975) ressalta a importância das expectativas dos consumidores, e Kahneman (2012) explora heurísticas como a da disponibilidade e o viés de enquadramento. Espera-se que os resultados das análises textuais apresentadas aqui incentivem a elaboração de outros estudos, não apenas sobre períodos de crise econômica, mas também em outros campos da ciência econômica que possam se beneficiar dessa abordagem.

Entre as limitações deste trabalho, destaca-se a mudança nos canais pelos quais a população norte-americana consome notícias, o que indica uma crescente preferência por fontes informais. De acordo com o World Economic Forum (2018), um em cada cinco norte-

americanos afirma consumir notícias por meio de redes sociais — como WhatsApp, Facebook e Instagram — em vez de recorrer aos jornais tradicionais, como foi o caso das notícias da base de dados Aylien (2020) que foi utilizada para o desenvolvimento do trabalho. Além disso, uma reportagem da NBC News (2024) aponta que um em cada cinco norte-americanos busca informações através de influenciadores nas redes sociais, o que pode acentuar a propagação de informações e narrativas que não sejam verdadeiras. Isso sugere que os próximos estudos sobre as narrativas econômicas, podem levar essas questões em consideração.

REFERÊNCIAS

- AKERLOF, G. A.; SHILLER, R. J. **Animal spirits: how human psychology drives the economy, and why it matters for global capitalism**. 9. ed. Princeton: Princeton University Press, 2010.
- AL-OBEIDAT, F.; ADEDUGBE, O.; HANI, A. B.; BENKHELIFA, E.; MAJDALAWIEH, M.. *Cone-KG: A Semantic Knowledge Graph with News Content and Social Context for Studying Covid-19 News Articles on Social Media*. In: 2020 **Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)**. Zayed University, United Arab Emirates, 2020. Disponível em: <https://www.academia.edu/101785405/>. Acesso em: 11 set. 2024. DOI: 10.1109/SNAMS52053.2020.9336541
- AMAZON. O que é Apache Spark? Disponível em: <https://aws.amazon.com/pt/big-data/what-is-spark/>. Acesso em: 15 nov. 2024.
- ARMANTIER, O.; TOPA, G.; VAN DER KLAAUW, W.; ZAFAR, B.. *An Overview of the Survey of Consumer Expectations*. **Federal Reserve Bank of New York Economic Policy Review**, v. 23, n. 2, p. 51-72, dez. 2017. Disponível em: https://www.newyorkfed.org/medialibrary/media/research/epr/2017/epr_2017_survey-consumer-expectations_armantier.pdf. Acesso em: 2 set. 2024.
- AYLIEN. 2020. Free Coronavirus News Dataset. Disponível em: <https://aylien.com/resources/datasets/coronavirus-dataset>. Acesso em: 2 mar. 2024.
- BASK, M.; FORSBERG, L.; ÖSTLING, A. Media sentiment and stock returns. **The Quarterly Review of Economics and Finance**, v. 94, p. 303-311, 2024. Disponível em: <https://doi.org/10.1016/j.qref.2024.02.008>. Acesso em: 10 set. 2024.
- BAYRAM, U. Revealing the reflections of the pandemic by investigating COVID-19 related news articles using machine learning and network analysis. **Bilişim Teknolojileri Dergisi**, v. 15, n. 2, p. 209-220, abr. 2022. DOI: 10.17671/gazibtd.949599. Disponível em: <https://doi.org/10.17671/gazibtd.949599>. Acesso em: 7 set. 2024.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. Sebastopol: O'Reilly Media, 2009.
- BLOOMBERG TELEVISION. **Yale's Shiller on Economic Fallout From Virus, Fed Policy and Housing**. YouTube, 30 jul. 2020. Disponível em: <https://www.youtube.com/watch?v=u7X75PGmFAA>. Acesso em: 09 set. 2024.
- CNBC. SHILLER: OIL PRICES COULD SHIFT MARKET PSYCHOLOGY. Disponível em: <https://www.cnbc.com/2015/01/14/shiller-oil-prices-could-shift-market-psychology.html>. Acesso em: 02 ago. 2024.
- COCHRANE, C.; RHEAULT, L.; GODBOUT, J. F.; WHYTE, T.; WONG, M. W. C.; BORWEIN, S. The Automatic Analysis of Emotion in Political Speech Based on Transcripts.

Political Communication, v. 39, n. 1, p. 98–121, 2021. DOI: 10.1080/10584609.2021.1952497.

CURTIN, R. George Katona: A founder of behavioral economics. In: FRANTZ, Roger et al. (orgs.). **Routledge handbook of behavioral economics**. New York: Routledge, 2016. Disponível em: <https://data.sca.isr.umich.edu/fetchdoc.php?docid=56663>. Acesso em 15 mar. 2024.

CURTIN, R. June 2020 survey results June 26, 2020. Survey of Consumers, University of Michigan, 26 jun. 2020. Disponível em: <https://data.sca.isr.umich.edu/fetchdoc.php?docid=65588>. Acesso em: 25 out. 2024.

CURTIN, R. Subject: July 2020 survey results. July 31, 2020. Survey of Consumers. The University of Michigan. Disponível em: <https://data.sca.isr.umich.edu/fetchdoc.php?docid=65773>. Acesso em: 25 out. 2024.

DORN, R.; MA, Y.; MORSTATTER, F.; LERMAN, K. Gender and prestige bias in coronavirus news reporting. arXiv, 27 jan. 2023. Disponível em: <https://arxiv.org/pdf/2301.11994>. Acesso em: 10 mar. 2024.

FARIAS, A. M. L. Estatística Descritiva. Niterói: **Universidade Federal Fluminense, Instituto de Matemática e Estatística, Departamento de Estatística**, 2020. 67 p. Disponível em: <https://www.professores.uff.br/anafarias/wp-content/uploads/sites/210/2021/01/estdesc-0-light.pdf>. Acesso em: 29 set. 2024.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**, 5. ed. Bookman: Porto Alegre, 2011

GUPTA, A.; KATARYA, R. PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning. **Computers in Biology and Medicine**, v. 138, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0010482521007149>. Acesso em: 11 set. 2024. <https://doi.org/10.1016/j.compbiomed.2021.104920>.

HASELMAYER, M.; JENNY, M. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. **Qual Quant**, v. 51, p. 2623–2646, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11135-016-0412-4>. DOI: 10.1007/s11135-016-0412-4.

HSU, J. Policy Brief 24-3: The influence of gasoline and food prices on consumer expectations and attitudes in the COVID era. Washington, D.C.: **Peterson Institute for International Economics**, p. 1-18, 2024. Disponível em: <https://www.piie.com/sites/default/files/2024-05/pb24-3.pdf>. Acesso em: 8 ago. 2024.

HUSAIN, A. M. et al. Global implications of lower oil prices. Washington, D.C.: **International Monetary Fund**, p. 1-41, 2015. Disponível em: <https://www.imf.org/external/pubs/ft/sdn/2015/sdn1515.pdf>. Acesso em: 10 ago. 2024.

HUTTO, C. J.; GILBERT, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: ICWSM 2014: Proceedings of the Eighth International

Conference on Weblogs and Social Media. Atlanta, GA: **Georgia Institute of Technology**, p. 216-225, 2014. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399>. Acesso em: 12 mar. 2024. DOI: 10.1609/icwsml.v8i1.14550.

JURAFSKY, D.; MARTIN, H. J. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3. ed. no prelo. Stanford: Stanford University; Boulder: University of Colorado at Boulder, 2018. Disponível em: https://web.stanford.edu/~jurafsky/slp3/ed3bookaug20_2024.pdf. Acesso em: 17 set. 2024.

KAHNEMAN, D. **Rápido e devagar**: duas formas de pensar. Rio de Janeiro: Objetiva, 2012.

KAMPIĆ, M.; BABIĆ, M. B. Sentiment analysis of President Trump's tweets: from winning the election to the fight against COVID-19. **Communication Management Review**, v. 6, n. 2, p. 90-111, 2021. Preliminary communication. Disponível em: http://www.commreview.hr/wp-content/uploads/2022/01/CMR2_vol6_2021_Kampic-Bagic-Bebic.pdf. Acesso em: 8 dez. 2024. DOI: 10.22522/cmr20210272.

KATONA, G. **Psychological economics**. New York: Elsevier Scientific Publishing Company, 1975.

LEVENE, H. Robust tests for equality of variances. In: OLKIN, Ingram; HOTELLING, Harold; et al. (eds.). **Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling**. Stanford University Press, p. 278–292, 1960.

LEVENE'S TEST: THE ASSESSMENT FOR EQUALITY OF VARIANCES. Disponível em: <https://medium.com/@kyawsawhtoon/levenes-test-the-assessment-for-equality-of-variances-94503b695a57>. Acesso em: 04 ago. 2024.

LOUGHRAN, T.; McDONALD, B. Textual Analysis in Finance. **Annual Review of Financial Economics**, v. 12, p. 357-375, 2020. Disponível em: <https://www.annualreviews.org/docserver/fulltext/financial/12/1/annurev-financial-012820-032249.pdf?expires=1728303748&id=id&accname=guest&checksum=2C98BBAE74ECCA0E3A21BC38AC3B4F43>. Acesso em: 26 set. 2024. <https://doi.org/10.1146/annurev-financial-012820-032249>.

MARTÍN, E. D. V.; VALENTÍN, L. D. L. F. Sentiment analysis methods for politics and hate speech contents in Spanish language: a systematic review. **IEEE Latin America Transactions**, v. 21, n. 3, p. 408–418, 2023. Disponível em: <https://latamt.ieee9.org/index.php/transactions/article/view/7268>.

MCCLOSKEY, D. N. **If You Are So Smart: The Narrative of Economic Expertise**. Chicago: University of Chicago Press, 1990.

MELEK, N. Ç.; VIGFUSSON, R. J. The evolving link between oil prices and U.S. consumer spending. Kansas City: **Federal Reserve Bank of Kansas City**, p. 41-55, 2021. Disponível em: <https://www.kansascityfed.org/Economic%20Review/documents/7910/1q21ercakirmelekvigfussou.pdf>. Acesso em: 6 ago. 2024.

MIN, B.; ROZONoyer, B.; QIU, H.; ZAMANIAN, A.; MACBRIDE, J. ExcavatorCovid: Extracting Events and Relations from Text Corpora for Temporal and Causal Analysis for COVID-19. 2021. Disponível em: <https://arxiv.org/pdf/2105.01819>. Acesso em: 10 set. 2024.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Skewness. Disponível em: <https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/skewness.htm>. Acesso em: 29 set. 2024.

NBC NEWS. (2024). 1 IN 5 U.S. ADULTS GET THEIR NEWS FROM SOCIAL MEDIA INFLUENCERS, ACCORDING TO PEW REPORT. NBC NEWS. Disponível em: <https://www.nbcnews.com/tech/news-influencers-social-media-pew-report-rcna179786>. Acesso em: 8 dez. 2024.

OIL DEMAND GROWING AT A SLOWER PACE AS POST-COVID REBOUND RUNS ITS COURSE. Disponível em: <https://www.iea.org/commentaries/oil-demand-growing-at-a-slower-pace-as-post-covid-rebound-runs-its-course>. Acesso em: 4 jun. 2024.

PITALE, S. K. Topic Modelling and Theme Discovery on Aylien News Articles during COVID-19. 2020. 55 f. Tese de Mestrado em Data Analytics, Dublin Business School. Disponível em: <https://esource.dbs.ie/server/api/core/bitstreams/7ba2b48a-6541-438b-97e0-935209a2090b/content>. Acesso em: 10 set. 2024.

QUINTAS, B. C. M.; FONA, L. E. M.; LIMA, L. R. O acirramento da guerra de preços do petróleo entre a Rússia e a OPEP em 2020 e seus efeitos na geopolítica e economia russa. **O Cosmopolítico**, v. 7, n. 1, p. 120-139, jun. 2020. Disponível em: <https://periodicos.uff.br/ocosmopolitico/article/view/53841/31678>. Acesso em: 15 jul. 2024.

ROOS, M.; RECCIUS, M. Narratives in economics. Ruhr-University Bochum, 2022. Disponível em: <https://arxiv.org/pdf/2109.02331>. Acesso em: 10 dez. 2024.

ROVETTA, A. Raiders of the lost correlation: a guide on using Pearson and Spearman coefficients to detect hidden correlations in medical sciences. **Cureus**, v. 12, n. 11, p 1-7, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7779167/pdf/cureus-0012-00000011794.pdf>. Acesso em: 23 set. 2024. DOI: 10.7759/cureus.11794

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples), *Biometrika* v. 52, p. 591–611, 1965

SHAPIRO-WILK TEST FOR NORMALITY. [S. l.]: [s.n.], 2015. p. 1-12. Disponível em: <https://math.mit.edu/~rmd/465/shapiro.pdf>. Acesso em: 20 ago. 2024.

SHEN, Y.; ZHANG, P. K. Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT. [S. l.], 2024. Disponível em: <https://arxiv.org/pdf/2410.01987>. DOI: <https://doi.org/10.48550/arXiv.2410.01987>.

SHILLER, R. J. **Narrative economics**: How stories go viral & drive major economics events. Princeton: Princeton University Press, 2019.

SON, J. A Study on Aspect-oriented Summarization using Transformer. 2023. Tese (Mestrado em Engenharia) — Interdisciplinary Program in Bioengineering, The Graduate School, Seoul National University, Seoul, 2023. Disponível em: <https://space.snu.ac.kr/bitstream/10371/193412/1/000000174619.pdf>. Acesso em: 11 set. 2024.

TAHERDOOST, H.; MADANCHIAN, M. Artificial intelligence and sentiment analysis: a review in competitive research. **Computers**, v. 12, n. 2, p. 37, 2023. Disponível em: <https://doi.org/10.3390/computers12020037>. Acesso em: 10 set. 2024

UNIVERSITY OF MICHIGAN, Institute for Social Research. Index of Consumer Sentiment. Disponível em: <https://data.sca.isr.umich.edu/fetchdoc.php?docid=24770>. Acesso em: 8 ago. 2024.

UNIVERSITY OF MICHIGAN. Surveys of Consumers: Index Calculations. Disponível em: <https://data.sca.isr.umich.edu/fetchdoc.php?docid=24770>. Acesso em: 03 out. 2024.

U.S. BUREAU OF LABOR STATISTICS. Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average [CPILFESL]. Disponível em: <https://fred.stlouisfed.org/series/CPILFESL>. Acesso em: 2 out. 2024.

U.S. BUREAU OF LABOR STATISTICS. Consumer Price Index: 2020 in review. 15 jan. 2021. Disponível em: <https://www.bls.gov/opub/ted/2021/consumer-price-index-2020-in-review.htm>. Acesso em: 22 set. 2024.

U.S. ENERGY INFORMATION ADMINISTRATION. Crude oil prices increased in 2021 as global crude oil demand outpaced supply. 4 jan. 2022. Disponível em: <https://www.eia.gov/todayinenergy/detail.php?id=50738>. Acesso em: 22 set. 2024.

U.S. ENERGY INFORMATION ADMINISTRATION. Petroleum & other liquids: U.S. gasoline prices. Disponível em: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=emm_epm0_pte_nus_dpg&f=m. Acesso em: 27 set. 2024.

VASILIEV, Y. **Natural Language Processing with Python and SpaCy**. San Francisco: No Starch Press, 2020.

WÄRNERYD, K. The life and work of George Katona. **Journal of Economic Psychology**, v. 2, n. 1, p. 1-31, 1982. Disponível em: [https://doi.org/10.1016/0167-4870\(82\)90008-3](https://doi.org/10.1016/0167-4870(82)90008-3). Acesso em: 24 mar. 2024.

WANG, E.; ANDUKURI, C.; DASARI, S. Paper Trading From Sentiment Analysis on Twitter and Reddit Posts. Stanford CS224N Custom Project. **Stanford CS224N Natural Language Processing with Deep Learning**, p. 1-8, 2023. Disponível em: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1234/final-reports/final-report-169488138.pdf>. Acesso em: 26 set. 2024.

WEN, Z.; POWELL, G.; CHAFI, I.; BUCKERIDGE, D. L.; LI, Y. Inferring global-scale temporal latent topics from news reports to predict public health interventions for COVID-19. **Patterns**, v. 3, n. 5, p. 1-14, 2022. Disponível em: <https://doi.org/10.1016/j.patter.2022.100435>. Acesso em: 10 mar. 2024.

WORLD ECONOMIC FORUM. (2018). More Americans are going to social media instead of newspapers for their news. World Economic Forum. Disponível em: <https://www.weforum.org/stories/2018/12/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>. Acesso em: 08 dez. 2024.

ZHANG, B.; YANG, H.; LIU, X.-Y. Instruct-FinGPT: Financial Sentiment Analysis by Instruction Tuning of General-Purpose Large Language Models. [S. l.], 2023. Disponível em: <https://arxiv.org/pdf/2306.12659>.

ZHANG, W.; GUI, L.; PROCTER, R.; HE, Y. NewsQuote: A Dataset Built on Quote Extraction and Attribution for Expert Recommendation in Fact-Checking. arXiv, [S.l.], 2023. Disponível em: <https://arxiv.org/pdf/2305.04825>. Acesso em: 15 set. 2024.