

**UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS – CCSA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS – PPGCC
DOUTORADO EM CIÊNCIAS CONTÁBEIS**

**A INFLUÊNCIA DA GOVERNANÇA CORPORATIVA NA INCERTEZA
EXPRESSA NO RELATÓRIO DA ADMINISTRAÇÃO**

**JOÃO PESSOA - PB
AGOSTO/2025**

PATRÍCIA PEREIRA CASTRO

**A INFLUÊNCIA DA GOVERNANÇA CORPORATIVA NA INCERTEZA
EXPRESSA NO RELATÓRIO DA ADMINISTRAÇÃO**

Tese apresentada ao Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal da Paraíba como parte dos requisitos necessários à obtenção do título de Doutora em Ciências Contábeis.

Linha de Pesquisa: Informação Contábil para Usuários Externos

Orientador: Prof. Dr. Orleans Silva Martins

**JOÃO PESSOA – PB
AGOSTO/2025**

**Catalogação na publicação
Seção de Catalogação e Classificação**

C355i Castro, Patrícia Pereira.

A influência da governança corporativa na incerteza expressa no relatório da administração / Patrícia Pereira Castro. - João Pessoa, 2025.
164 f. : il.

Orientação: Orleans Silva Martins.
Tese (Doutorado) - UFPB/CCSA.

1. Contabilidade - Análise textual. 2. Sentimento linguístico. 3. Incerteza textual. 4. Governança corporativa. 5. American Depository Receipts (ADRs). I. Martins, Orleans Silva. II. Título.

UFPB/BC

CDU 657(043)

ATA DE DEFESA DE TESE DE DOUTORADO**DEFESA DE TESE Nº 47**

Ata de Sessão Pública de Defesa de Tese de Doutorado da aluna Patrícia Pereira Castro, do Curso de Doutorado do Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal da Paraíba.

Linha de Pesquisa: Informação Contábil para Usuários Externos

Ao primeiro dia do mês de agosto do ano de dois mil e vinte e cinco, às quatorze horas, reuniu-se, na Plataforma Google Meet, através do link: <https://meet.google.com/fxm-gv kf-rya>, a Banca Examinadora, composta pelos professores doutores: Orleans Silva Martins (Presidente da Banca Examinadora – PPGCC/UFPB), Luiz Renato Régis de Oliveira Lima (Membro Interno – PPGCC/UFPB), Wenner Glaucio Lopes Lucena (Membro Interno – PPGCC/UFPB), Lucas Lúcio Godeiro (Membro Externo – UFERSA) e Wagner Antônio Marques (Membro Externo - UFES), para julgar a tese intitulada: “**A Influência da Governança Corporativa na Incerteza Expressa no Relatório da Administração**”, de autoria da aluna Patrícia Pereira Castro, orientada pelo Prof. Dr. Orleans Silva Martins. Dando início aos trabalhos, o Presidente da Banca Examinadora, explicou aos presentes a finalidade da sessão pública, e passou a palavra à doutoranda para que fizesse a apresentação de sua tese. Após a apresentação do trabalho, a banca examinadora fez arguições à discente, que as respondeu. Em seguida, o presidente da banca examinadora, acatou as observações e a Banca Examinadora se reuniu reservadamente, para que fosse feito o julgamento do trabalho, e após decisão então atribuiu à aluna o conceito:

- () Aprovado(a)
() Insuficiente
() Reprovado (a)

Comentários da Banca Examinadora

A banca recomendou ajustes de melhoria no trabalho, os quais foram encaminhados à discente com cópia a seu orientador, incluindo a melhoria na fundamentação e na exploração do uso do ADR como proxy de melhor governança corporativa e do método diff-in-diff na análise dos eventos de interesse.

Proclamados os resultados, o Presidente da Banca Examinadora encerrou os trabalhos referentes à Defesa de Tese da discente. João Pessoa, 01 de agosto de 2025.

Prof. Dr. Orleans Silva Martins
Presidente da Banca Examinadora – PPGCC/UFPB

Prof. Dr. Luiz Renato Régis de Oliveira Lima
Membro Interno – PPGCC/UFPB

Prof. Dr. Wenner Glaucio Lopes Lucena
Membro Interno – PPGCC/UFPB

Documento assinado digitalmente
 LUCAS LUCIO GODEIRO
Data: 05/08/2025 14:54:04-0300
Verifique em <https://validar.itd.gov.br>

Prof. Dr. Lucas Lúcio Godeiro
Membro Externo - UFERSA

Prof. Dr. Vagner Antônio Marques
Membro Externo- UFES

Documento assinado digitalmente
 Patricia Pereira Castro
Data: 05/08/2025 17:02:02-0300
Verifique em <https://validar.itd.gov.br>

Patrícia Pereira Castro
Discente

Documento original assinado eletronicamente, conforme MP 2200-2/2001, art. 10, § 2º, por:

VAGNER ANTÔNIO MARQUES

CIDADÃO

assinado em 05/08/2025 11:11:16 -03:00



INFORMAÇÕES DO DOCUMENTO

Documento capturado em 05/08/2025 11:11:16 (HORÁRIO DE BRASÍLIA - UTC-3)
por VAGNER ANTÔNIO MARQUES (CIDADÃO)

Valor Legal: ORIGINAL | Natureza: DOCUMENTO NATO-DIGITAL

A disponibilidade do documento pode ser conferida pelo link: <https://e-docs.es.gov.br/d/2025-PC0FC1>

Emitido em 01/08/2025

ATA N° 47/2025 - CCSA-PPGCC (11.00.52.03)
(Nº do Documento: 47)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 06/08/2025 14:21)
WENNER GLAUCIO LOPES LUCENA
COORDENADOR CURS/POS-GRADUACAO
1526402

(Assinado digitalmente em 06/08/2025 11:11)
ORLEANS SILVA MARTINS
PROFESSOR DO MAGISTERIO SUPERIOR
1775506

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número: **47**, ano: **2025**, documento (espécie): **ATA**, data de emissão: **06/08/2025** e o código de verificação: **6bac8b9fa8**

Emitido em 01/08/2025

ATA Nº 47/2025 - CCSA-PPGCC (11.00.52.03)
(Nº do Documento: 47)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 08/08/2025 10:28)
LUIZ RENATO REGIS DE OLIVEIRA LIMA

PROFESSOR DO MAGISTERIO SUPERIOR

1788474

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número:
47, ano: 2025, documento (espécie): ATA, data de emissão: 07/08/2025 e o código de verificação: 2b23d71878

Dedico esta tese ao Tom, meu ponto de equilíbrio em meio à incerteza.

AGRADECIMENTOS

Antes de qualquer palavra, agradeço a Deus por guiar meus passos com infinita bondade, abrindo caminhos, alinhando o universo e permitindo que tudo acontecesse no seu tempo. A jornada foi longa, por vezes solitária e árdua, mas jamais faltou fé de que as coisas aconteceriam como deveriam.

Agradeço à minha criança interior, por sua resiliência e coragem. Filha de uma empregada doméstica (hoje costureira) e de um caminhoneiro (*in memoriam*), ela sempre soube o motivo de estarmos aqui. Em todos os momentos em que pensei em desistir, ela me lembrava com firmeza: “Olhe aonde já chegamos. Lembre-se de tudo o que superamos na vida. Não desista agora.”

À minha mãe, que me deu a vida e me ensinou a jamais esquecer de onde vim e para onde irei (do pó viemos e ao pó voltaremos). Sempre dizia: “Se você não quiser lavar o banheiro dos outros, estude!” E assim fiz. Pela educação, mudei meu destino. E uma das maiores razões pelas quais persisti neste doutorado foi o desejo de dar a ela uma vida mais digna e tranquila.

Ao meu irmão mais novo, pela compreensão silenciosa diante de ausências que me doeram profundamente. No segundo ano do doutorado, seu enteado e meu sobrinho de coração, partiu por suicídio, e mesmo diante dessa dor irreparável, eu não pude estar fisicamente presente. E quando a vida se renovou, no nascimento do meu sobrinho mais novo, também não pude estar ao seu lado. A tese caminhava, mas a vida acontecia e muitas vezes longe de mim. Obrigada por entender que, embora ausente, meu coração esteve com vocês em cada um desses momentos.

À psicóloga, Ana Karla Albuquerque, por estar comigo nas madrugadas da alma. Sua escuta acolhedora e respeitosa foi âncora nos dias em que tudo parecia ruir. Mais do que aliviar a ansiedade, ajudou-me a sempre analisar o que está sob meu controle e não carregar a culpa pelas ausências que a vida acadêmica impôs. Sua presença silenciosa foi parte da estrutura invisível que me manteve de pé.

À Universidade Federal de Rondônia (UNIR), pela concessão do afastamento para realização deste doutorado, e em especial à Prof.^a Dr.^a Iluska Braga, minha chefe de departamento à época, por sua sensibilidade e apoio genuíno. Aos demais professores do DACCNT-PVH que aprovaram esse afastamento, meu sincero reconhecimento. À Universidade Federal da Paraíba (UFPB), que me acolheu nesta etapa tão desafiadora.

Ao meu orientador Prof. Dr. Orleans Martins, que assumiu a orientação em um momento delicado e conduziu com firmeza, clareza e humanidade. Com discernimento e objetividade, transformou urgência em ação. Praticamente, em menos de um mês, construímos um projeto, a qualificação ocorreu no prazo e chegamos até aqui.

Aos professores da banca, agradeço pelas contribuições fundamentais: Prof. Dr. Orleans Silva Martins - Presidente; Prof. Dr. Luiz Renato Lima - Membro Interno; Prof. Dr. Wenner Glaucio Lopes Lucena - Membro interno; Prof. Dr. Lucas Lucio Godeiro (UFERSA) e Prof. Dr. Wagner Antônio Marques (UFES). Em especial ao Prof. Dr. Luiz Renato, por ir além da crítica e me apresentar caminhos. Uma de suas sugestões está guardada para um futuro próximo; a outra foi abraçada nesta tese. Sua generosidade ao enviar materiais, gravar áudio, esclarecer dúvidas e até me ligar quando eu mais precisava, revelou, mesmo sem obrigação alguma, um gesto de humanidade que jamais esquecerei.

Gostaria de expressar minha profunda gratidão a todo o corpo docente do PPGCC, por sua paciência, humanidade e compromisso com a formação de pesquisadores. Cada professor(a), à sua maneira, contribuiu para a construção do caminho que percorri. A seguir deixo registrada, a essência do que cada um representou para mim:

- Prof. Dr. Aldo Callado – *Estruturação*. Sua disciplina, foi essencial para a base conceitual e técnica da tese e contribuíram para uma evolução gradual da tese.
- Prof. Dr. Andson Aguiar (USP) – *Referência*. Por me mostrar, com sua postura atenta, humana e sensível, o tipo de docente que desejo ser. Seu compromisso com o aprendizado, evidenciado nos *feedbacks* pontuais a cada atividade e cuidadosos, marcou profundamente meu caminho. Finalizamos a disciplina com um projeto-piloto rodado e uma inspiração viva para a docência.
- Prof. Dr. Dimas Barreto – *Disciplina*. Seu rigor e dedicação me ensinaram mais que Teoria da Contabilidade: me mostraram o valor de estar plenamente presente em sala de aula.
- Prof. Dr. Josedilton Diniz – *Estímulo*. Por promover espaços de diálogo entre linhas distintas, permitindo que mesmo não matriculados em sua disciplina participassem de discussões enriquecedoras, como o bate-papo sobre experimentos, que despertou ainda mais minha curiosidade sobre o tema.

- Prof. Dr. Lauro Vinício – *Provocação*. Em sua disciplina nasceram meus primeiros estudos aplicados em legibilidade e tom, que viriam a se tornar a base desta tese.
- Prof. Dr. Luiz Renato – *Excelênci*a. Suas aulas foram um marco: foi nelas que perdi o medo do R. Tudo o que aprendíamos era aplicado, interpretado e apresentado, em uma didática prática que nos fazia dominar o conteúdo com segurança. Sua generosidade, clareza e humildade são exemplos do que quero levar para minha docência.
- Prof.^a Dr.^a Márcia Reis – *Doçura*. Foi a única professora que se despediu de mim pessoalmente, com um abraço e um olhar de cuidado quase materno, no último evento presencial de que participei na UFPB. Posteriormente, em uma reunião para discussão sobre um artigo, ensinou que o verdadeiro resultado não estava nos dados, mas no quanto havíamos crescido durante as férias.
- Prof. Dr. Márcio Machado – *Reflexões*. Provocou reflexões fundamentais sobre o fazer científico e os desafios da trajetória acadêmica. Seus textos e questionamentos deixaram marcas importantes sobre a tensão entre pesquisar e produzir.
- Prof. Dr. Orleans Martins – *Formação*. Mostrou que conhecimento prático, paixão e exigência constroem o aprendizado de qualidade. Transformou o conteúdo em aprendizado aplicado.
- Prof. Dr. Paulo Amilton – *Leveza*. Sua forma serena de ensinar e dialogar foi um respiro constante na caminhada.
- Prof. Dr. Robério França – *Sabedoria*. Pelas orientações generosas oferecidas nos momentos mais inesperados e por ter sido curador atento em um dos meus artigos, desenvolvido em coautoria com uma de suas orientandas.
- Prof.^a Dr.^a Rossana Guerra – *Desconstrução*. Suas provocações epistemológicas me ajudaram a enxergar estruturas que eu mesma não via, especialmente ao me conduzir à leitura de Bell Hooks. E, claro, por despertar em mim a curiosidade sobre dissonância cognitiva e prevenção à fraude (um dia sairá da “gaveta”).
- Prof. Dr. Wenner Lucena – *Paizão*. Meu primeiro professor no PPGCC e nosso coordenador, sempre com a escuta ativa e paciente. Também foi quem mais nos impulsionou à pesquisa, e graças a esse incentivo, juntos conquistamos um prêmio de melhor artigo no Congresso da USP e ali iniciei no *Machine Learning*.

E por fim, mas não menos importante, meu carinho e gratidão às funcionárias da secretaria do PPGCC, Wilma e Cecília. Sem vocês, nosso programa não teria a mesma alma. Em especial, a Wilma, com seu jeito acolhedor e humano, esteve presente em momentos e diálogos delicados. Discreta, escutava com empatia, oferecendo não apenas apoio técnico, mas um acolhimento sincero.

Aos colegas e amigos de jornada, aos que dividiram leituras, planilhas e angústias, meu carinho eterno. Em especial, à Alice Rainho, Clara Monise, Diego Siqueira, João Marcelo Macedo, Kalina Kelly Leite e Paulo Matos. Aos meus amigos e familiares que renunciaram à minha presença por diversas vezes, minha gratidão por não me deserdarem.

Aos meus nobres amigos da CAE-CFC, minha gratidão por me acolherem sempre com um abraço quentinho e escuta paciente, em nossas reuniões, em especial ao Prof. Dr. José Antônio França, que por diversas vezes me aconselhou e até mesmo se colocou à disposição para ser meu coorientador, para que eu pudesse escrever sobre perícia, minha verdadeira paixão.

Ao contador Rômulo Teotônio, ex-presidente do CRCPB, pela oportunidade de coordenar a comissão CRC PB Jovem no período em que morei em João Pessoa/PB, o que, além de me manter no trabalho voluntário, me permitiu ter uma válvula de escape saudável do doutorado. E, claro aos meus liderados que de mente aberta se permitiram ser liderados por alguém que veio lá do Mato Grosso do Sul.

E por fim, ao meu gato, companheiro silencioso de dias e madrugadas. Mesmo sem entender nada, esteve ali, nos momentos em que vibrei, chorei, me irritei ou silenciei. Chegou a adoecer durante nossa estadia em Jampa; cheguei a achar que iria perdê-lo. Obrigada por me lembrar, mesmo em silêncio, que o amor está também na presença e que, às vezes, os pedidos por atenção foram pausas necessárias para que eu lembresse de respirar.

“Não duvide de nada, pois qualquer coisa pode acontecer e o negócio mesmo é entender as coisas quando elas acontecem.”

Adaptado de Alexandre Werneck,
A Desculpa

RESUMO

Objetivo: Esta tese tem por objetivo investigar como a governança corporativa, por meio da listagem de empresas brasileiras via *American Depository Receipts* (ADRs) no mercado norte-americano, influencia a incerteza textual dos relatórios da administração. Este estudo analisa conjuntamente a legibilidade, o tom textual e o impacto da governança sobre a opacidade informacional, considerando o papel disciplinador de ambientes regulatórios mais exigentes.

Metodologia: A pesquisa emprega uma estratégia quantitativa com delineamento quase-experimental, articulando o modelo *Difference-in-Differences* (DiD) escalonado à técnica de *Propensity Score Matching* (PSM), de modo a atenuar potenciais vieses de seleção amostral. A amostra foi construída com base em relatórios da administração publicados por empresas não financeiras listadas na B3 (2010-2024), complementados por dados da Economatica e Refinitiv Eikon. Os indicados textuais (Legibilidade, *Length*, *Common words*, *Financial words*, tom e incerteza textual) foram extraídos por algoritmos de Processamento de Linguagem Natural (PLN) e *machine learning*. A variável de governança corporativa foi representada pela listagem das empresas por meio de *American Depository Receipts* (ADRs). O modelo final inclui variáveis explicativas textuais e controles relacionados às características das empresas e ao ambiente econômico. Testes de robustez, como placebo *test*, foram aplicados para verificar validade interna.

Resultados: Os achados não confirmam as três hipóteses formuladas, mas oferecem evidências relevantes. Primeiramente, observou-se que relatórios da administração mais legíveis estão associados a menores níveis de incerteza textual, em linha com o esperado, embora sem significância estatística. Em segundo lugar, verificou-se que o uso de tom negativo quando combinado com relatórios mais claros, reduz a incerteza textual, revelando que a legibilidade modera o impacto do tom negativo e permite que o tom sinalize transparência, quando usado adequadamente.

Por fim, não se confirma que a listagem de empresas brasileiras via ADRs esteja associada a uma redução estatisticamente significativa da incerteza textual expressa nos relatórios da administração. Ainda assim, os resultados apontam que padrões mais rigorosos de governança e exigências de disclosure podem exercer influência positiva sobre a qualidade linguística dos relatórios narrativos, especialmente quando combinados com práticas de comunicação claras e consistentes.

Originalidade: A originalidade do estudo reside em três pilares: (i) integração inédita entre governança corporativa, legibilidade, tom e incerteza textual; (ii) uso combinado de PLN e econometria causal (PSM e DiD) no contexto brasileiro; e (iii) exploração dos ADRs como mecanismo estrutural de governança em mercados emergentes, revelando seus efeitos sobre atributos linguísticos da comunicação corporativa.

Contribuições: Esta tese oferece três contribuições principais. A primeira é de natureza teórica, ao integrar de forma inédita os construtos de governança corporativa, legibilidade e tom textual em um modelo de efeito estimado sob suposições sobre a formação da incerteza informacional nos relatórios da administração. Os resultados aprofundam os fundamentos das teorias da Agência, Sinalização e Divulgação. Essa abordagem amplia a compreensão de como mecanismos formais de governança e estratégias narrativas interagem na formação da opacidade informacional em mercados emergentes.

A segunda contribuição é metodológica, ao incorporar ao contexto brasileiro o uso combinado de técnicas de Processamento de Linguagem Natural (PLN) e *Machine Learning*, e métodos econométricos robustos (PSM e DiD). Essa abordagem permite automatizar a análise de grandes volumes de texto com precisão replicável e elevada credibilidade inferencial. A terceira contribuição é prática e institucional, ao fornecer evidências e métricas que podem apoiar reguladores, empresas e investidores no aprimoramento das práticas de divulgação financeira. Ao mensurar de forma objetiva dimensões como clareza textual, tom e incerteza linguística, o estudo apresenta insumos para o desenvolvimento de políticas públicas, o aperfeiçoamento de padrões de governança e a incorporação de variáveis qualitativas nos modelos de avaliação de risco.

Palavras-chave: análise textual; sentimento linguístico; incerteza textual; governança corporativa; american depositary receipts (ADRs).

LISTA DE TABELAS

Tabela 1 Sistematização de Alguns Estudos com o Índice Flesch.	48
Tabela 2 Quantitativo de Arquivos PDF Ajustados Manualmente por Ano.	69
Tabela 3 Resultado da Regressão DiD – Teste Placebo (datas falsas de adoção). ..	78
Tabela 4 Resultado da Regressão DiD – Teste de Estresse – Exclusão de Coortes Sensíveis.	79
Tabela 5 Resultado da Regressão DiD – Teste Placebo – Janela alternativa para t0.	81
Tabela 6 Relação entre os objetivos, hipóteses e modelos por função.	82
Tabela 7 Variáveis incluídas no PSM e no modelo DiD.	83
Tabela 8 Variáveis de Controle.	94
Tabela 9 Número de Observações por Variável.	95
Tabela 10 Fundamentação Metodológica para Especificações Parcimoniosas em Painéis Desbalanceados.	96
Tabela 11 Distribuição das Empresas Tratadas por Ano de Anúncio e Adoção dos ADRs.	99
Tabela 12 Amostra: Empresas e Observações no Painel (2010–2024).	100
Tabela 13 Estatísticas de Balanceamento Antes do Pareamento (PSM).	102
Tabela 14 Estatísticas de Balanceamento Após do Pareamento (PSM).	103
Tabela 15 Resultado da Regressão (H1).	106
Tabela 16 Resultado da Regressão (H2).	110
Tabela 17 Resultado da Regressão DiD (H3).	114
Tabela 18 Resultado da Regressão DiD – Teste de Robustez – Incerteza Alternativa (EPU).	120
Tabela 19 Exemplo de Categorização de Relatório da Administração.	155

LISTA DE QUADROS

Quadro 1 <i>Variável Dependente da Pesquisa</i>	84
Quadro 2 <i>Constructos e palavras-chave para busca bibliográfica</i>	132
Quadro 3 <i>Strings de busca por base prioritária e relação de busca</i>	133
Quadro 4 <i>Resumo da RSL</i>	135

LISTA DE FIGURAS

Figura 1 Diagrama de Venn.....	26
Figura 2 Diagrama de Efeito Estimado.....	29
Figura 3 Desenho e hipóteses da pesquisa.....	63
Figura 4 Design da Pesquisa e Enquadramento Metodológico.....	67
Figura 5 Empresas brasileiras com ADRs por ano de emissão (1993–2024)	73
Figura 6 Evolução temporal da incerteza textual por grupo.	77
Figura 7 Comparativo dos índices Flesch Original e Adaptado.....	86
Figura 8 Legibilidade média – escolaridade.	87
Figura 9 Legibilidade média – estrutura textual.....	88
Figura 10 Balanceamento das covariáveis antes e depois do pareamento.....	101
Figura 11 Distribuição dos escores de propensão antes e depois do pareamento.	102
Figura 12 Evolução da legibilidade média dos RAs por grupo de empresas.....	105

LISTA DE EQUAÇÕES

Equação (1) Modelo 1 – Legibilidade e Incerteza Textual (H1)	74
Equação (2) Modelo 2 – Tom Negativo e Incerteza Textual (H2).....	75
Equação (3) Modelo 3 – Efeito Estimado da Governança via ADRs (H3)	76
Equação (4) Incerteza	83
Equação (5) Legibilidade (índice flesch adaptado).....	85
Equação (6) Length Index	89
Equação (7) Common Words Index	89
Equação (8) Financial Words Index.....	90
Equação (9) Tom	91
Equação (10) Tom negativo	91

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

ADRs	<i>American Depository Receipts</i>
ARI	<i>Automated Readability Index</i>
B3	Brasil, Bolsa, Balcão
CD&A	<i>Compensation Discussion and Analysis</i>
CEO	<i>Chief Executive Officer</i>
CVM	Comissão de Valores Mobiliários
DiD	<i>Difference-in-Differences</i>
DFP	Demonstrações Financeiras Padronizadas
ESG	<i>Environmental, Social, and Governance</i>
EPU	<i>Economic Policy Uncertainty</i>
EUA	Estados Unidos da América
FGV	Fundação Getúlio Vargas
ICC	Índice de Confiança do Consumidor
IFRS	<i>International Financial Reporting Standards</i>
IPO	<i>Initial Public Offering</i>
LN	Logaritmo Natural
MD&A	<i>Management's Discussion and Analysis</i>
OCR	<i>Optical Character Recognition</i>
PDF	Portable Document Format
PNL	Processamento de Linguagem Natural
PSM	<i>Propensity Score Matching</i>
PWA	<i>Plain Writing Act</i>
RA	Relatório da Administração
ROA	Retorno sobre os Ativos
RSL	Revisão Sistemática da Literatura
SEC	<i>Securities and Exchange Commission</i>
SMOG	<i>Simple Measure of Gobbledygook</i>
SOX	<i>Sarbanes-Oxley Act</i>
TXT	Arquivo de Texto ou Documento de Texto
VIF	<i>Variance Inflation Factor</i>
VMP	Valor de Mercado sobre o Patrimônio

SUMÁRIO

1.	INTRODUÇÃO	22
2.	FUNDAMENTAÇÃO TEÓRICA E DESENVOLVIMENTO DAS HIPÓTESES .	35
2.1	Teoria da agência	35
2.2	Teoria da sinalização	39
2.3	Legibilidade	45
2.4	Tom (Sentimento Textual)	52
2.5	Governança corporativa	58
2.6	Paradoxos da divulgação: entre agência, sinalização e transparência	63
3.	METODOLOGIA.....	67
3.1	<i>Design</i> da pesquisa e enquadramento	67
3.2	Procedimentos de coleta e tratamento dos dados.....	67
3.3	<i>Propensity score matching</i> (PSM)	71
3.3.1	Delineamento da amostra.....	72
3.3.2	Escolha das covariáveis para o <i>propensity score matching</i> (PSM)	73
3.4	Modelos econométricos da pesquisa.....	74
3.4.1	Modelo 1 – legibilidade e incerteza textual (H1).....	74
3.4.2	Modelo 2 –tom negativo e incerteza textual (H2)	75
3.4.3	Modelo 3 – efeito estimado da governança via ADRs (H3 – PSM + DiD).....	75
3.5	Variável dependente	83
3.6	Variável de tratamento.....	84
3.7	Variáveis explicativas e de controle	85
3.7.1	Legibilidade	85
3.7.1.1	Índice flesch adaptado para o português.....	85
3.7.1.2	Comprimento do documento (<i>length</i>)	89
3.7.1.3	Proporção de palavras comuns (<i>common words index</i>).....	89
3.7.1.4	Proporção de palavras financeiras (<i>financial words index</i>).....	90
3.7.2	Tom Negativo.....	90
3.7.3	Variáveis de controle	91
3.7.4	Variáveis Governança corporativa: considerações metodológicas e restrições empíricas.....	94
3.8	Limitações da pesquisa	97
4.	ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	99
4.1	Descrição da amostra e aplicação do PSM	99
4.2	Análise das hipóteses.....	104
4.2.1	Relação entre legibilidade e incerteza (H1).....	104

4.2.2	Relação entre Tom Negativo e Incerteza (H2).....	109
4.2.3	Efeito da Governança via ADRs sobre a Incerteza (H3).....	113
4.3	Considerações Finais da Análise das Hipóteses	118
4.4	Teste de Robustez.....	119
5.	CONSIDERAÇÕES FINAIS.....	122
	REFERÊNCIAS.....	126
	Apêndice A – Protocolo de Revisão Sistemática da Literatura (RSL).....	132
	Apêndice B – <i>Script R</i> Extração do RA’s e Conversão “.pdf” para “.txt”.....	136
	Apêndice C – <i>Script R</i> para calcular características textuais – Linguagem R	138
	Apêndice D – <i>Script</i> em R para PSM e DiD	142
	Apêndice E – Lista das Palavras Positivas	147
	Apêndice F – Listas das Palavras Negativas	149
	Apêndice G – Listas das Palavras Incerteza.....	154
	Apêndice H – Exemplificação Cálculos de Legibilidade, Tom e Incerteza.....	155

1. INTRODUÇÃO

No mercado de capitais, a qualidade da informação divulgada é fundamental para a redução da assimetria informacional e para o funcionamento eficiente dos mercados. Relatórios corporativos mais legíveis permitem que os investidores formem expectativas mais precisas, ao reduzir a incerteza informacional e a opacidade sobre os fundamentos econômicos e financeiros da empresa (Healy & Palepu, 2001; Jensen & Meckling, 1976). Além da apresentação dos números, a narrativa dos Relatórios da Administração desempenha papel crucial na comunicação entre gestores e investidores (Aerts, 2005). A forma como essas informações são estruturadas, em termos de legibilidade, tom e complexidade, influencia diretamente a interpretação das informações, o custo de capital e a percepção de risco (Li, 2008; Hassan et al., 2022).

Neste estudo, a legibilidade se refere à facilidade de compreensão dos relatórios financeiros e foi mensurada a partir do índice Flesch adaptado para o português (Martins et al., 1996), complementada por métricas adicionais, como o comprimento do documento (*Length*), a proporção de palavras comuns (*Common words*) e a proporção de palavras financeiras (*Financial words*). O tom, por sua vez, representa a conotação emocional da linguagem utilizada nos relatórios, podendo ser positivo, negativo ou neutro e foi calculado conforme a metodologia de Henry (2008), com base no dicionário de Loughran e McDonald (2011), identificando a predominância de termos positivos e negativos na comunicação corporativa.

O comprimento (*length*) do documento foi medido pelo logaritmo natural do número total de palavras (Abernathy et al., 2019; Li, 2008; Loughran & McDonald, 2014). E, a métrica de *common words* representa a proporção média de palavras comuns em relação ao total de palavras, enquanto *financial words* captura a proporção de termos financeiros, conforme o dicionário do Prof. Campbell Harvey (Abernathy et al., 2019).

Os gestores utilizam os relatórios financeiros e outras divulgações corporativas para comunicar o desempenho e a governança da empresa aos investidores e demais *stakeholders*. Esses instrumentos são essenciais para o funcionamento eficiente do mercado de capitais, pois afetam a alocação de recursos e podem influenciar a incerteza informacional e a percepção de risco por parte dos investidores (Healy & Palepu, 2001).

Para este estudo, a incerteza informacional foi captada por meio da incerteza textual, uma métrica linguística que reflete a opacidade comunicacional presente nos relatórios da administração. Essa incerteza é operacionalizada conforme o dicionário desenvolvido por (Loughran & McDonald, 2011, 2014), que quantifica a frequência de termos associados à ambiguidade, imprecisão e contingência, transformando a análise textual em uma ferramenta objetiva para mensurar o risco informacional percebido pelos investidores.

Quando os gestores estruturam esses relatórios de forma clara, eles reduzem a assimetria de informações, conflito de agências e permitem que os investidores formem expectativas mais precisas sobre o futuro das empresas (Healy & Palepu, 2001; Jensen & Meckling, 1976). Além de apresentar os resultados financeiros, os gestores utilizam as narrativas dos relatórios da administração para contextualizar esses números, oferecendo uma visão mais ampla da evolução do negócio e da situação financeira (Aerts, 2005). Assim, a maneira como os gestores redigem os relatórios reflete a clareza e a transparência na comunicação da empresa (Dau et al., 2024). Logo, os gestores devem assegurar que esses relatórios sejam acessíveis e compreensíveis para os investidores, analistas e demais *stakeholders* (Aerts, 2005).

A escolha das palavras e a estrutura dos relatórios não são apenas questões estilísticas, mas fatores que impactam diretamente as condições financeiras da empresa. A literatura demonstra que a complexidade textual e o tom dos relatórios financeiros influenciam a interpretação das informações e afetam o nível de incerteza percebido pelo mercado (Hassan et al., 2022; Li, 2008). Assim, relatórios financeiros menos legíveis ou com tom ambíguo podem aumentar o custo de captação de recursos, pois a dificuldade na interpretação das informações aumenta a percepção de risco, devido à possibilidade de decisões ineficientes de alocação de capital, elevando as exigências para concessão de crédito (Ertugrul et al., 2017).

Empresas podem adotar estratégias narrativas distintas, seja para fornecer uma visão detalhada e técnica das operações ou para mascarar desempenhos insatisfatórios. Relatórios mais complexos e de difícil leitura tendem a gerar maior incerteza informacional, pois dificultam a interpretação dos dados e ampliam a seção de risco percebido pelos investidores minimizar a percepção de problemas, o que pode afetar diretamente a incerteza percebida pelos investidores (Aerts, 2005; Elliott et al., 2024; Li, 2008; Tan et al., 2014). Pois, a transparência e a qualidade da comunicação financeira moldam as percepções imediatas dos investidores e

credores, sinalizando o nível de risco e a qualidade da gestão corporativa (Ertugrul et al., 2017; Henry, 2008; Li et al., 2019; Reichmann, 2023).

Além dos fatores textuais, o ambiente regulatório e as práticas de governança corporativa impactam a qualidade e a transparência das divulgações financeiras. A governança corporativa desempenha um papel essencial na forma como as empresas comunicam informações ao mercado. Quando estruturada com conselhos de administração mais robustos e apoiada por práticas regulatórias mais rigorosas, ela favorece uma divulgação mais clara e transparente. Isso não apenas aprimora a qualidade das informações financeiras, como também contribui para diminuir a incerteza percebida pelos investidores (Hussainey et al., 2022; Shleifer & Vishny, 1997).

Nesse sentido, um fator relevante para a potencial melhoria da qualidade textual dos seus relatórios da administração é a adoção regulatória de listagem no mercado de ações norte-americano, via emissão de *American Depository Receipts* (ADRs), que são certificados representativos de ações de empresas estrangeiras. Isso porque as empresas que negociam ADRs nos Estados Unidos precisam atender aos padrões rigorosos da *Securities and Exchange Commission* (SEC), que impõe regras mais estritas de governança e transparência, refletindo os efeitos da globalização financeira sobre a redução dos custos de capital e do risco corporativo (Stulz, 2022). Isso exige que empresas estrangeiras cumpram padrões mais rigorosos de governança e transparência, exigidos pela SEC e pela Sarbanes-Oxley Act (SOX), como a implementação de comitês independentes de auditoria, de políticas formais de ética e de mecanismos robustos de controle interno (Sayari & Marcum, 2018).

Os ADRs foram criados para facilitar o acesso de investidores norte-americanos a ativos de empresas de outros países, permitindo a negociação em dólares e sob o ambiente regulatório norte-americano. Assim, a emissão de ADRs, especialmente nos níveis 2 e 3, estabelece uma relação formal entre a empresa estrangeira e o ambiente institucional dos Estados Unidos, sujeitando-a a um arcabouço regulatório mais rigoroso em termos de governança e divulgação (Sayari & Marcum, 2018).

Nesse contexto, a globalização financeira reduziu o custo de capital ao permitir maior diversificação dos investidores e fortalecer a governança corporativa (Stulz, 2022). Em mercados emergentes, a integração aos mercados globais, com a remoção de restrições à entrada de investidores estrangeiros, mostrou-se determinante para

reduzir o custo de capital, ampliar o acesso ao financiamento e melhorar a precificação dos ativos (Stulz, 2022). Ainda, empresas de mercados emergentes que adotam padrões de governança similares aos dos Estados Unidos, como ocorre nos programas de ADRs de nível 2 e 3, apresentam uma redução no risco informacional. O primeiro risco, mensurado pela volatilidade dos retornos das ações, decorre da dificuldade de diversificar riscos em mercados segmentados e da maior exposição a choques macroeconômicos locais, conforme discutido na literatura sobre integração financeira (Stulz, 2022). O segundo risco está associado à assimetria informacional e à baixa qualidade das divulgações típicas de ambientes com governança frágil, sendo mitigado pela adoção de padrões internacionais de governança e divulgação, como aqueles exigidos na listagem via ADRs (Sayari & Marcum, 2018).

No entanto, Stulz (2022) destaca que apenas abrir o mercado, isto é, permitir a entrada de investidores estrangeiros e o acesso das empresas locais aos mercados de capitais internacionais, não é suficiente para assegurar os benefícios da integração. Na ausência de governança corporativa eficaz, esses benefícios podem ser comprometidos, como ilustram casos documentados na literatura, a exemplo da Malásia, onde a liberalização não foi acompanhada de melhorias institucionais adequadas. Essa lógica também se aplica a mercados emergentes como o Brasil, embora a adoção de melhores práticas mais robustas de governança, como adesão ao Novo Mercado da B3, tenha representado avanços institucionais relevantes, a qualidade da informação divulgada continua sendo um fator determinante para reduzir a incerteza (Sayari & Marcum, 2018; Stulz, 2022).

Diante disso, a adoção regulatória/estratégica de listagem como ADRs pode estar associada à produção de relatórios mais legíveis e menos ambíguos, o que tende a contribuir para reduzir a incerteza informacional e o risco percebido pelos investidores.

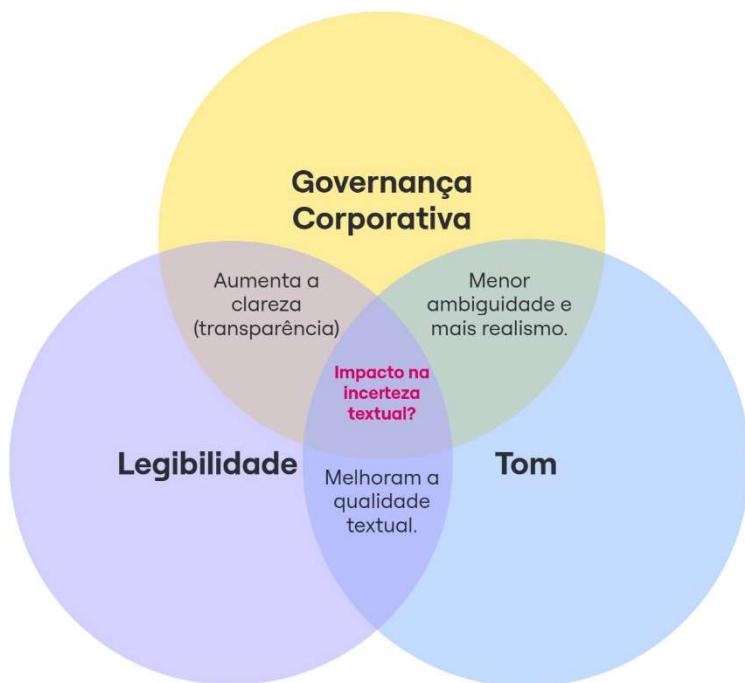
Estudos anteriores apontam que, diante dos benefícios que a literatura atribui à melhoria da clareza dos relatórios financeiros (Aerts, 2005; Healy & Palepu, 2001), a adoção regulatória de listagem via ADRs, é frequentemente apontada como um mecanismo institucional que pode contribuir para reduzir a incerteza informacional, ao exigir padrões regulatórios e de governança mais rigorosos, como aqueles impostos pela SEC e pela SOX (Sayari & Marcum, 2018; Stulz, 2022). Contudo, enquanto a literatura confirma que relatórios bem estruturados tendem a diminuir a assimetria de informações, há também evidências de que a complexidade textual e o tom que os

gestores adotam pode aumentar a percepção de risco (Hassan et al., 2022; Li, 2008; Rajabalizadeh, 2023).

Embora estudos tenham analisado isoladamente a legibilidade, o tom dos relatórios e as práticas de governança corporativa, sendo que esta última, quando robusta, tende a promover divulgações mais transparentes (Shleifer & Vishny, 1997) a interação entre esses fatores, não foi evidenciada na literatura revisada. Por isso, identificar lacunas na literatura existente e formular questões de pesquisa específicas são estratégias eficazes para contribuir para o avanço do conhecimento científico (Sandberg & Alvesson, 2011).

Para ilustrar a fragmentação supracitada, a Figura 1 apresenta um diagrama de Venn que sintetiza as interseções entre legibilidade, tom dos relatórios e governança corporativa. O diagrama evidencia que, embora cada aspecto tenha sido estudado separadamente, a análise integrada dos seus efeitos sobre a incerteza textual ainda não foi evidenciada.

Figura 1
Diagrama de Venn.



Fonte: Elaboração própria (2025).

Portanto, mesmo considerando as evidências de que a adoção de padrões internacionais, como os exigidos para empresas que negociam ADRs, podem

melhorar a qualidade dos relatórios, persiste uma lacuna quanto à compreensão de como a combinação desses elementos afeta a incerteza informacional. Assim, este estudo busca responder à seguinte questão de pesquisa: **O melhor nível de governança corporativa, evidenciado pela listagem de American Depository Receipts (ADRs), influencia a incerteza informacional expressa na legibilidade e no tom dos relatórios da administração de empresas brasileiras listadas na B3?**

Diante da lacuna identificada na literatura sobre o impacto da governança corporativa na incerteza textual dos relatórios da administração, o presente estudo busca aprofundar a compreensão dessa relação. Especificamente, pretende-se investigar como diferentes dimensões da qualidade textual, como a legibilidade e o tom afetam a incerteza nas divulgações corporativas e de que forma a governança influencia essa dinâmica.

Para responder à questão de pesquisa, o objetivo geral é **investigar como a governança corporativa, por meio da listagem de ADRs, influencia a incerteza informacional expressa nos relatórios da administração, considerando a análise conjunta da legibilidade e do tom adotado nas divulgações.**

Os seguintes objetivos específicos foram definidos com base no objetivo geral:

- a) Analisar a associação entre a legibilidade dos relatórios da administração e a incerteza textual.
- b) Analisar a associação entre o tom dos relatórios da administração e a incerteza textual.
- c) Estimar o efeito da listagem de empresas brasileiras nos Estados Unidos via ADRs sobre a incerteza textual.

Ao pesquisar esses objetivos, esta tese parte da premissa de que a governança corporativa desempenha um papel crucial na mitigação dos problemas de agência e na redução da assimetria informacional entre gestores e investidores, garantindo que as decisões estratégicas sejam alinhadas aos interesses dos investidores (Healy & Palepu, 2001; Jensen & Meckling, 1976; Shleifer & Vishny, 1997).

No âmbito da comunicação corporativa, a qualidade da informação que as empresas divulgam influencia diretamente a decisão dos investidores, pois uma comunicação clara reduz os custos de processamento da informação e a incerteza percebida (Verrecchia, 2001), bem como afeta a formação dos preços dos ativos e a eficiência dos mercados (Fama, 1970), tornando a legibilidade e o tom das informações elementos cruciais para reduzir a incerteza informacional.

A literatura sugere que a legibilidade e o tom dos relatórios financeiros afetam a percepção dos investidores, e os gestores podem utilizá-los intencionalmente. Relatórios menos legíveis e com tom ambíguo podem aumentar a incerteza informacional, dificultando a interpretação dos dados financeiros, elevando a percepção de risco por parte do mercado (Dalwai et al., 2021; Hassan et al., 2022; Li, 2008) e são frequentemente associados a tentativas de ocultação de informações ou de gerenciamento de impressões por parte da administração (Elliott et al., 2024; Rajabalizadeh, 2023).

Nesse sentido, a governança corporativa atua como um mecanismo disciplinador, exercendo um papel essencial na redução da assimetria informacional e melhoria da transparência das divulgações financeiras (Shleifer & Vishny, 1997). Estudos indicam que mecanismos de governança mais rígidos reduzem práticas oportunistas e melhoram a clareza das informações que as empresas divulgam, tornando os relatórios financeiros mais acessíveis e compreensíveis para os investidores (Ginesti et al., 2017, 2018).

No contexto global, empresas brasileiras listadas nos Estados Unidos por meio de ADRs estão sujeitas a padrões regulatórios mais rigorosos, como aqueles estabelecidos pela SEC e pela SOX. Os reguladores impõem melhores práticas de governança corporativa e transparência, resultando em relatórios financeiros mais claros e objetivos, reduzindo a incerteza informacional (Sayari & Marcum, 2018). Dessa forma, a adoção regulatória de listagem via ADR é adotada na literatura como uma *proxy* válida de governança corporativa aprimorada, representando um alinhamento mais estreito com as boas práticas internacionais.

Além disso, a globalização dos mercados financeiros impõe padrões mais elevados de governança, levando empresas que acessam o mercado norte-americano a melhorar a qualidade de suas divulgações para atender às expectativas dos investidores internacionais (Stulz, 2022).

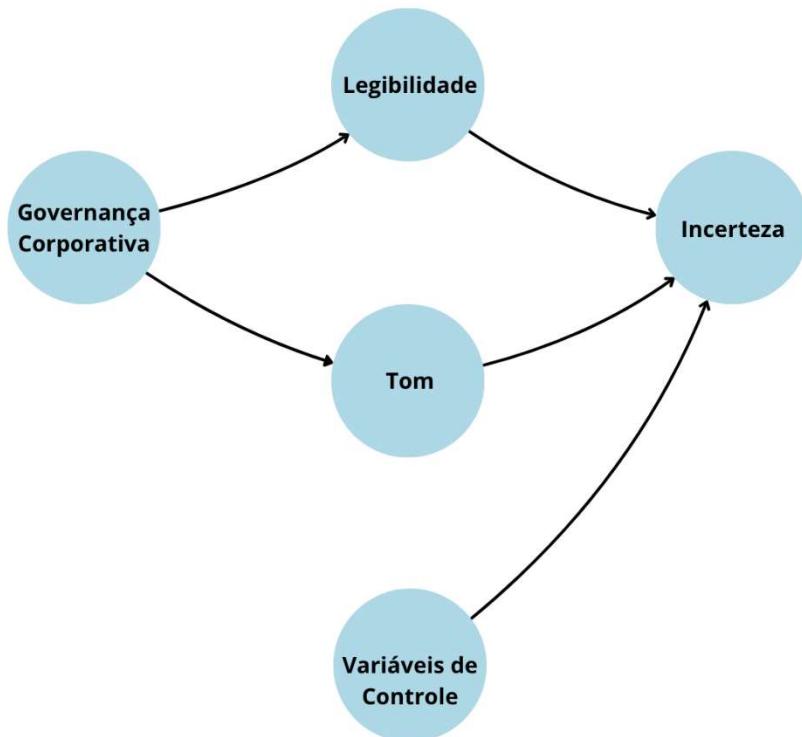
Dessa forma, este estudo sustenta a tese de que **a listagem de empresas brasileiras nos Estados Unidos via American Depository Receipts (ADRs) melhora a legibilidade e está associada ao uso mais adequado do tom negativo nos relatórios da administração, contribuindo para reduzir a incerteza textual, em razão da convergência de padrões de governança mais rigorosos.**

A Figura 2 apresenta o diagrama de causalidade que representa a tese defendida neste estudo. Nele, a variável Governança Corporativa, operacionalizada

por meio da listagem em ADRs, exerce influência direta sobre duas variáveis textuais: Legibilidade e Tom dos relatórios da administração. Ambas, por sua vez, estão associadas diretamente à Incerteza Textual, variável dependente central deste trabalho. Adicionalmente, o diagrama contempla Variáveis de Controle, que capturam características específicas das firmas, dos textos e do ambiente econômico, as quais também exercem influência direta sobre a Incerteza Textual, mitigando possíveis vieses decorrentes da omissão de variáveis relevantes.

Figura 2

Diagrama de Efeito Estimado.



Nota: Elaboração própria (2025).

Sendo assim, este estudo utilizou a listagem das empresas em ADRs como proxy de tratamento na análise do impacto da governança corporativa sobre a incerteza textual, permitindo avaliar se a regulação norte-americana influencia a qualidade das informações que as empresas divulgam.

Os relatórios da administração foram escolhidos como objeto de análise por serem documentos obrigatórios, padronizados e de ampla abrangência, que desempenham um papel central na comunicação entre empresas e investidores. Esses relatórios reúnem informações prospectivas, qualitativas e narrativas sobre o

desempenho, os riscos e as perspectivas futuras das empresas, indo além da simples apresentação dos números financeiros (Aerts, 2005; Dau et al., 2024; Hassan et al., 2022).

Na literatura internacional, os relatórios da administração são reconhecidos como peças fundamentais na redução da assimetria informacional, funcionando como ferramentas de transparência, sinalização e gestão de impressões (Dau et al., 2024; Henry, 2008; Li, 2008; Tan et al., 2014). Além disso, a natureza textual desses relatórios permite a aplicação objetiva de métricas linguísticas, como legibilidade, tom e incerteza, as quais têm sido amplamente utilizadas na literatura para avaliar como as características da comunicação corporativa afetam o comportamento dos investidores, o custo de capital e a percepção de risco (Hassan et al., 2022; Li, 2008; Rajabalizadeh, 2023).

Este estudo analisa a relação entre a governança corporativa e a incerteza textual nos relatórios da administração, utilizando métricas linguísticas e métodos quantitativos amplamente consolidados na literatura. Operacionalizou-se a governança corporativa pela listagem das empresas brasileiras no mercado norte-americano, por meio de ADRs, e mensurou-se a incerteza textual com índices derivados da análise dos relatórios da administração, com base em técnicas de análise textual.

Adicionalmente, foram consideradas variáveis textuais (legibilidade, tom e comprimento do relatório), bem como variáveis de controle associadas às características das firmas e do ambiente econômico. Para assegurar robustez na identificação do efeito associado, aplicou-se os métodos de PSM e DiD, que permitem estimar o impacto da adoção de padrões mais rigorosos de governança sobre a incerteza textual.

Compreender o papel da governança corporativa na redução da assimetria informacional e na proteção dos interesses dos investidores se torna fundamental. Estudos indicam que sua atuação pode amplificar os efeitos positivos das informações corporativas na precificação das ações (Tahat et al., 2021), moderar a relação entre despesas de capital e sustentabilidade financeira (Ammer & Salim, 2024), além de reduzir a tendência de gestão de resultados (Ishak et al., 2018). Adicionalmente, entender de que forma a governança influencia a clareza, o tom dos relatórios corporativos e a incerteza é essencial para aprimorar a transparência e a qualidade informacional do mercado financeiro.

Este estudo oferece uma contribuição relevante ao analisar como a exposição a ambientes de governança mais rigorosos, como o mercado norte-americano via ADRs, impacta as características textuais dos relatórios da administração e, consequentemente, a incerteza percebida pelos investidores. A pesquisa aborda diretamente três dimensões críticas da comunicação corporativa: legibilidade, tonalidade e incerteza textual.

Além disso, o tema é particularmente oportuno diante das mudanças recentes no ambiente regulatório brasileiro, que passaram a exigir maior clareza e qualidade informacional, como evidenciado pelas Resoluções CVM nº 59/2021 (CVM, 2021) e nº 87/2022 (CVM, 2022), e Projeto de Lei nº 2925/2023 (Brasil, 2023). Tais avanços refletem a crescente importância da comunicação transparente e padronizada no fortalecimento do mercado de capitais brasileiro.

Do ponto de vista prático, este estudo possui potencial de impacto direto em múltiplos públicos interessados. Reguladores, como a CVM, podem se beneficiar dos achados ao desenvolver diretrizes que estimulem a clareza, a objetividade e a acessibilidade das informações textuais presentes nos relatórios da administração. Empresas podem aprimorar suas práticas de divulgação das informações, alinhando a comunicação aos princípios de governança e às expectativas de investidores, especialmente os minoritários. Adicionalmente, investidores e analistas podem aperfeiçoar seus modelos de avaliação de risco informacional, ao compreenderem como a qualidade textual dos relatórios reflete práticas de governança e padrões regulatórios.

A relevância do tema também se conecta à crescente preocupação global com a comunicação clara e acessível no mercado financeiro (Biehl et al., 2024). Evidências empíricas, como a implementação da *Plain Writing Act* (PWA) nos Estados Unidos, demonstram que relatórios mais claros e legíveis reduzem a volatilidade dos retornos e melhoram a capacidade dos investidores de monitorar as empresas (Cho et al., 2024).

Do ponto de vista teórico, este estudo avança a literatura ao integrar quatro construtos, governança corporativa, legibilidade e tom, para entender seus efeitos sobre a incerteza textual. Embora estudos anteriores tenham analisado essas variáveis de forma isolada, não se identificou, na literatura, investigações que combinem esses elementos em um modelo de efeito estimado aplicado ao contexto de empresas brasileiras listadas nos Estados Unidos via ADRs.

Adicionalmente, a pesquisa oferece uma contribuição metodológica inédita no contexto brasileiro ao empregar técnicas de processamento de linguagem natural e *Machine Learning* para análise textual. Automatizar a coleta, a leitura e a análise dos relatórios da administração, com critérios linguísticos bem definidos e replicáveis, representam um avanço importante em relação às antigas análises feitas de forma manual e, muitas vezes, subjetiva.

A combinação dos métodos de PSM e DiD contribui para fortalecer a identificação do efeito estimado ao balancear covariáveis pré-tratamento observáveis, e em seguida, estimar o efeito da governança sobre a incerteza textual. Dessa forma, a pesquisa não apenas descreve padrões linguísticos, mas estima o impacto da governança corporativa, via ADRs, sobre a qualidade textual e a incerteza dos relatórios corporativos.

A originalidade desta pesquisa reside na combinação de três dimensões complementares: (i) a integração de construtos que, até então, eram analisados de forma isolada na literatura; (ii) a inovação metodológica aplicada ao contexto brasileiro, tanto no tratamento dos dados textuais quanto na robustez da inferência do efeito estimado; e (iii) a exploração de um contexto institucional particularmente relevante, caracterizado pelos desafios dos mercados emergentes e pela adoção de mecanismos de governança internacionalizada via ADRs.

O primeiro pilar da originalidade se ancora na integração teórica de construtos relacionados à governança corporativa, legibilidade e tom textual, e seus efeitos conjuntos sobre a incerteza informacional. Embora a literatura internacional tenha avançado significativamente na análise isolada desses construtos, com estudos focados na legibilidade (Li, 2008), no tom (Henry, 2008; Loughran & McDonald, 2011) ou na governança corporativa (Dodge et al., 2004; Stulz, 2022), não se observava, até então, uma abordagem que investigasse de forma conjunta como esses elementos interagem na formação da qualidade informacional dos relatórios da administração e na redução da incerteza textual, sobretudo no contexto de mercados emergentes. Esta pesquisa, portanto, não se limita à mera identificação de uma lacuna (*gap-spotting*), como discutem Sandberg e Alvesson (2011), mas avança na direção de uma problematização mais profunda dos modelos existentes, ao demonstrar que a compreensão isolada desses fatores é insuficiente para explicar a dinâmica da incerteza textual.

O segundo eixo de originalidade é de natureza metodológica. A pesquisa incorpora, de forma inédita no contexto brasileiro, o uso combinado de Processamento de Linguagem Natural (PLN), *Machine Learning*, e técnicas econométricas robustas, com destaque para a integração dos métodos *Propensity Score Matching* (PSM) e *Difference-in-Differences* (DiD). Esta abordagem permite não apenas automatizar a coleta, classificação e mensuração dos relatórios da administração, com critérios linguísticos consistentes, como também assegura inferências causais robustas. A utilização do Índice de Legibilidade de Flesch adaptado para o português (Martins et al., 1996) assegura que as especificidades linguísticas do idioma sejam corretamente capturadas, superando as limitações dos modelos desenvolvidos para a língua inglesa. A validação empírica dessa adaptação, aplicada especificamente a documentos técnicos e corporativos, reforça a acurácia dos resultados. Soma-se a isso o uso integrado de métodos econométricos de pareamento e avaliação de impacto (PSM e DiD), que conferem elevada credibilidade de efeito estimado aos resultados, o que representa um avanço metodológico significativo na análise de comunicação financeira no Brasil.

O terceiro eixo reside no próprio contexto de aplicação, que, ao focar em empresas brasileiras listadas no mercado norte-americano via ADRs, permite compreender como a adoção de padrões internacionais de governança impacta a qualidade da comunicação textual corporativa. Este contexto é especialmente relevante para mercados emergentes, onde predominam maiores níveis de assimetria informacional, menor *enforcement* regulatório e fragilidades institucionais. A listagem via ADRs funciona, nesse sentido, como um mecanismo robusto de “sinalização estrutural” (Coffee, 2002; Doidge et al., 2004; Stulz, 2022), impondo práticas de governança e divulgação mais rigorosas, cujos efeitos sobre os atributos linguísticos dos relatórios são aqui investigados de forma inédita.

De forma alinhada aos objetivos estabelecidos, os resultados empíricos não confirmam integralmente as três hipóteses centrais desta tese, mas oferecem evidências relevantes sobre os fatores que influenciam a incerteza textual nos relatórios da administração. No caso da H1, constatou-se que relatórios da administração mais legíveis estão associados a menores níveis de incerteza textual, embora não tenha sido estatisticamente significativa. Esse resultado reforça, ainda que de forma não significante estatisticamente a importância da clareza na comunicação corporativa como mecanismo de redução da opacidade informacional.

Quando a H2, verificou-se que o tom negativo, quando combinado com relatórios mais claros reduzem a incerteza, indicando que a legibilidade atua como moderadora, entre o tom negativo e a incerteza textual, sugerindo que esse tipo de tonalidade pode funcionar como sinalizador crível da situação da empresa.

Por fim, a H3 não foi confirmada de forma robusta, embora alguns períodos após o anúncio ou a listagem via ADRs apresentem redução significativa da incerteza textual. Esses achados sugerem que o efeito disciplinador de ambientes institucionais mais exigentes pode existir, mas tende a ser condicionado por fatores como o reduzido número de empresas tratadas e a dispersão temporal entre eventos de listagem, o que limita a detecção de efeitos persistentes no horizonte analisado.

2. FUNDAMENTAÇÃO TEÓRICA E DESENVOLVIMENTO DAS HIPÓTESES

2.1 Teoria da agência

A teoria da agência desempenha um papel central neste estudo, ao fornecer uma estrutura para compreender como os mecanismos da governança corporativa mitigam os conflitos de interesse entre gestores (agentes) e acionistas (principais) (Jensen & Meckling, 1976). Esses conflitos decorrem, sobretudo, da assimetria informacional, na qual os gestores detêm mais informações do que os investidores, podendo adotar comportamentos oportunistas (Jensen & Meckling, 1976).

Neste contexto, mecanismos de governança, como auditorias independentes, conselhos de administração ativos e relatórios financeiros transparentes, são fundamentais para reduzir os custos de agência (Jensen & Meckling, 1976). Assim, além de disciplinar os gestores, esses instrumentos visam mitigar a assimetria informacional, um dos principais fatores geradores de custos de agência (Jensen & Meckling, 1976), contribuindo, portanto, para ampliar a qualidade das informações disponibilizadas ao mercado (Hassanein & Albitar, 2024; Healy & Palepu, 2001).

Jensen e Meckling (1976) enfatizam que o monitoramento eficaz dos gestores, combinado com a implementação de incentivos adequados, é crucial para alinhar os interesses dos agentes com os dos principais. Esse alinhamento melhora a qualidade da informação divulgada, na medida em que os agentes passam a ter menor incentivo para ocultar ou distorcer informações, contribuindo para a redução dos custos de agência (Jensen & Meckling, 1976).

Além dos mecanismos internos de governança, a literatura destaca que empresas podem adotar estratégias institucionais externas para mitigar a assimetria informacional. Ao se submeter voluntariamente às exigências regulatórias de mercados mais desenvolvidos, como ocorre na listagem via ADRs, as empresas estabelecem um mecanismo de *bonding* (vinculação externa) capaz de fortalecer a proteção aos investidores, reduzir custos de agência e aumentar a credibilidade da informação divulgada ao mercado (Coffee, 2002). Nesse sentido, a governança deixa de ser restrita a estruturas internas e passa a incorporar elementos institucionais e jurídicos externos, que reforçam os sistemas de monitoramento e disciplinamento dos agentes econômicos (Coffee, 2002).

Embora originalmente associada à análise de estruturas contratuais e mecanismos formais de controle, a Teoria da Agência também oferece um referencial

para compreender como os sistemas de informação, funcionam como instrumentos de mitigação da assimetria informacional e de redução dos custos de monitoramento (Eisenhardt, 1989). Segundo Eisenhardt (1989), a estrutura da agência é aplicável em uma variedade de configurações, abrangendo desde questões de nível macro até fenômenos como gestão de impressões, uso estratégico da linguagem estratégica, opacidade informacional e outras expressões de interesse próprio.

Uma das proposições fundamentais da teoria é que a existência de sistemas de informação eficazes atua como mecanismo restritivo ao comportamento oportunista dos agentes, ao permitir que os principais monitorem e verifiquem suas ações (Eisenhardt, 1989). Sob essa perspectiva, os relatórios corporativos são reconhecidos como componentes essenciais dos sistemas de informação utilizados para mitigar os problemas decorrentes da assimetria informacional entre gestores (agentes) e investidores (principais) e elementos centrais dos sistemas de governança, pois ampliam a capacidade dos principais de observar, avaliar e disciplinar os agentes (Eisenhardt, 1989).

Complementarmente, Lajili (2015) amplia esse entendimento ao demonstrar que os sistemas de governança não devem ser concebidos apenas como instrumentos de controle formal, mas como arranjos capazes de alinhar capital humano, ativos organizacionais e informações. A autora argumenta que atributos como especificidade, complementaridade e incerteza do capital humano exercem influência determinante sobre a arquitetura dos mecanismos de governança (Lajili, 2015).

Destaca-se que a comunicação corporativa, especialmente por meio dos relatórios financeiros e narrativos, é um dos principais instrumentos para mitigar tanto a assimetria informacional quanto os custos de agência (Healy & Palepu, 2001). Quando dotada de clareza, objetividade e credibilidade, ela desempenha papel central na redução do risco informacional e do risco de agência, ao limitar tanto os problemas de seleção adversa quanto os incentivos ao comportamento oportunista dos gestores (Healy & Palepu, 2001).

Consequentemente, isso impacta diretamente a eficiência dos mercados de capitais, ao melhorar a qualidade da informação disponível para os investidores (Healy & Palepu, 2001). Ou seja, a eficácia dos mecanismos de governança depende, em grande medida, da qualidade das informações disponibilizadas (Eisenhardt, 1989).

Portanto, a teoria da agência é fundamental para entender como práticas de governança impactam características textuais, como legibilidade, tom e incerteza, na medida em que a assimetria informacional surge exatamente do fato de que os agentes detêm mais informações do que os investidores (Hassanein & Albitar, 2024; Hodgdon & Hughes, 2016; Jensen & Meckling, 1976). Reduzir essa assimetria não depende apenas de quantidade de informações, mas da clareza, precisão e objetividade na forma como essas informações são comunicadas (Hassanein & Albitar, 2024; Hodgdon & Hughes, 2016).

Empresas com melhores práticas de governança e auditoria produzem relatórios menos genéricos e mais claros, específicos e informativos, reduzindo significativamente a assimetria informacional, pois eliminam jargões, ambiguidades e redundâncias, facilitando a interpretação pelos usuários externos (Hodgdon & Hughes, 2016). Contudo, o tom incerto nas comunicações corporativas exerce efeito significativo sobre o desconto de preço (*underpricing*) no *Initial Public Offering* (IPO), funcionando como um mecanismo de sinalização da qualidade e da percepção de risco por parte dos investidores (González et al., 2019).

Além disso, o excesso de informações irrelevantes gera sobrecarga cognitiva, reduz a capacidade dos investidores de interpretar o conteúdo e, consequentemente, eleva a percepção de risco e incerteza e compromete a utilidade das informações (Hassanein & Albitar, 2024). Corroborando essa perspectiva, Carlini et al. (2020) demonstram que o tom e o grau de certeza de notícias relacionadas à governança impactam significativamente as reações dos mercados acionários. Especificamente, notícias com tom negativo e elevado grau de certeza geram perdas anormais mais severas, reforçando que a qualidade da comunicação desempenha papel disciplinador relevante na mitigação da assimetria informacional (Carlini et al., 2020).

Adicionalmente, a legibilidade dos relatórios financeiros contribui para mitigar a assimetria, garantindo a monitorização dos investidores, bem como que os investidores tenham acesso a informações precisas e relevantes para tomar decisões informadas (Cho et al., 2024 Dau et al., 2024). Da mesma forma, o tom exerce papel disciplinador relevante, afetando a percepção de risco e a avaliação dos investidores (Carlini et al., 2020; González et al., 2019). Adicionalmente, exigências regulatórias mais rigorosas, somadas à atuação de investidores institucionais e analistas, exercem pressão sobre os gestores, incentivando a adoção de práticas de divulgação de

informações mais transparentes, que mitigam a assimetria informacional e reduzem os custos de agência (Coffee, 2002; Healy & Palepu, 2001).

Contudo, a lógica da Teoria da Agência pressupõe que, na presença de assimetrias informacionais e de incentivos desalinhados, os gestores podem recorrer a estratégias comunicacionais oportunistas (Eisenhardt, 1989; Jensen & Meckling, 1976). Ou seja, gestores podem recorrer à adoção de um tom mais positivo ou ambíguo nos relatórios, desconectado dos fundamentos operacionais, como estratégia para suavizar percepções negativas no mercado (Pouryousof et al., 2023).

De acordo com a teoria da agência, a governança corporativa alinha melhor os interesses dos gestores e acionistas, o que, por sua vez, melhora o desempenho da empresa (Grove et al., 2011). Nesse contexto, a eficiência empresarial depende fortemente do alinhamento de interesses entre acionistas e gestores, com o conselho corporativo realizando atividades que obrigam os gestores a perseguirem o interesse dos acionistas (Waheed & Malik, 2019).

Diante desse contexto, evidencia-se que a governança corporativa, ao mitigar os problemas decorrentes da assimetria informacional, não apenas disciplina os agentes, mas também aprimora a qualidade da comunicação corporativa (Coffee, 2002; Eisenhardt, 1989; Jensen & Meckling, 1976). Nesse sentido, os relatórios da administração deixam de serem meros instrumentos formais de compliance e passam a desempenhar um papel estratégico na redução dos custos de agência, na medida em que tornam a informação mais clara, precisa e útil para os investidores (Healy & Palepu, 2001; Hodgdon & Hughes, 2016).

Contudo, além de funcionarem como mecanismos de controle e monitoramento, as características textuais desses relatórios, especialmente sua legibilidade, tom e nível de incerteza, também operam como sinais informacionais relevantes, capazes de modular percepções sobre risco, credibilidade e qualidade (Carlini et al., 2020; González et al., 2019).

Assim, a Teoria da Agência estabelece as bases para compreender como práticas de governança impactam a qualidade da informação. No entanto, entender de que forma os agentes utilizam a comunicação para reduzir incertezas percebidas pelo mercado exige a articulação conceitual com a Teoria da Sinalização (Spence, 1973). Segundo essa teoria, os agentes utilizam sinais, como características textuais dos relatórios, para transmitir informações sobre a qualidade, a governança e o

desempenho da firma, mitigando os efeitos da assimetria informacional (Dau et al., 2024; Pouryousof et al., 2023).

2.2 Teoria da sinalização

Além da teoria da agência, a teoria da sinalização, proposta por Spence (1973), oferece um arcabouço complementar para compreender como as empresas buscam reduzir os efeitos da assimetria informacional. Connelly et al. (2011) complementam essa perspectiva, fornecendo uma síntese abrangente dos principais construtos da dessa teoria e destacando o papel crítico da assimetria informacional no processo de sinalização. Essa teoria parte do princípio de que, em mercados caracterizados por informações assimétricas, os agentes mais bem informados, no caso, os gestores, utilizam sinais observáveis para transmitir informações críveis sobre a qualidade, o desempenho e as perspectivas futuras da empresa aos investidores (Spence, 1973).

Nesse contexto, os relatórios narrativos, especialmente os relatórios da administração, deixam de ser apenas instrumentos formais de *compliance* e passam a operar como veículos estratégicos de sinalização (Dau et al., 2024; González et al., 2019; Li, 2008). Características textuais como a legibilidade, o tom e o nível de incerteza textual funcionam como sinais não financeiros capazes de modular as percepções dos investidores sobre a qualidade da governança, a solidez financeira e a credibilidade das informações divulgadas (Dau et al., 2024; Li, 2008; Loughran & McDonald, 2014). Esses sinais contribuem para a redução da assimetria informacional porque a sua credibilidade é reforçada pelo custo de transmitir informações falsas ou enganosas (Connelly et al., 2011).

A legibilidade dos relatos corporativos constitui um mecanismo de sinalização textual relevante no âmbito da Teoria da Sinalização (Spence, 1973). A clareza textual funciona como um sinal observável da qualidade das informações divulgadas e da transparência informacional, permitindo aos investidores inferirem, com maior precisão, a qualidade dos fundamentos econômicos e das decisões gerenciais da empresa (Li, 2008; Loughran & McDonald, 2014).

Relatórios excessivamente complexos ou opacos estão associados a maior assimetria informacional, menor qualidade dos resultados, maior percepção de risco e, consequentemente, maiores custos de capital (Li, 2008; Loughran & McDonald, 2014). Esse entendimento dialoga diretamente com a Teoria da Sinalização (Spence,

1973), na medida em que a legibilidade opera como um sinal não financeiro capaz de reduzir a opacidade informacional percebida pelos investidores. A complexidade textual compromete a credibilidade do sinal, na medida em que dificulta o processamento das informações e amplia a percepção de incerteza (Connelly et al., 2011).

Esse efeito se conecta ao arcabouço teórico do *disclosure*, segundo o qual a divulgação de informações claras e compreensíveis reduz os custos de processamento de informações, mitiga a seleção adversa e melhora a eficiência das decisões dos investidores (Dye, 1985; Healy & Palepu, 2001 Verrecchia, 2001).

Adicionalmente, a legibilidade dos relatos carrega sinais sobre as decisões estratégicas e de investimento das empresas, influenciando as expectativas dos investidores quanto à qualidade dos fundamentos econômicos e ao desempenho futuro (Dau et al., 2024; Li, 2008). No contexto dos mercados emergentes, nos quais a assimetria informacional tende a ser mais acentuada, a legibilidade dos relatos assume um papel relevante como mecanismo de sinalização (González et al., 2019).

Portanto, textos mais claros e acessíveis são percebidos como indicativos de práticas gerenciais transparentes e de maior qualidade informacional, enquanto relatórios excessivamente complexos ou opacos são interpretados como sinal de risco, opacidade e, possivelmente, de tentativa de mascaramento (Dau et al., 2024; Li, 2008; Loughran & McDonald, 2014).

Por outro lado, a adoção de linguagem excessivamente técnica, densa ou ambígua pode ser percebida como sinal adverso, associado a práticas de ocultamento ou de gestão oportunística da informação (Li, 2008). Dessa forma, a comunicação textual, ao refletir escolhas estratégicas dos gestores, transcende sua função de simples reporte, assumindo papel central na mitigação da opacidade informacional e na construção da reputação corporativa (Li, 2008).

Assim como a legibilidade, o tom dos relatos corporativos constitui um mecanismo de sinalização textual relevante no âmbito da Teoria da Sinalização (Spence, 1973). A literatura demonstra que o mercado interpreta o tom como um sinal não financeiro da credibilidade, da robustez das práticas de governança e da qualidade dos fundamentos econômicos da empresa (Carlini et al., 2020; Huang et al., 2014; Pouryousof et al., 2023).

No contexto da América Latina, González et al., 2019 evidenciam que o tom adotado nos prospectos de IPO funciona como um sinal autônomo da incerteza

informacional percebida pelo mercado, afetando diretamente o nível de *underpricing* das ofertas públicas iniciais. Ainda que estruturas formais de governança atuem na mitigação da assimetria, os investidores atribuem peso significativo aos sinais transmitidos pela linguagem dos documentos, especialmente no que se refere ao tom e à clareza textual (González et al., 2019).

Estudos anteriores mostram que o mercado reage inicialmente a tons anormalmente otimistas como indicativo de melhores perspectivas futuras (Huang et al., 2014). Contudo, quando esse otimismo não se sustenta nos fundamentos econômicos, ocorrem correções subsequentes nos preços das ações, sinalizando que o tom é percebido como um elemento reputacional relevante, embora suscetível a revisão (Huang et al., 2014). Por outro lado, a adoção de tom negativo, especialmente em contextos relacionados à governança, transmite sinais de aumento na percepção de risco informacional, impactando significativamente os retornos das ações (Carlini et al., 2020).

Discrepâncias entre o tom adotado nos relatos e os resultados econômicos, particularmente quando há tom otimista desconectado da realidade operacional, são interpretadas negativamente pelo mercado, elevando a percepção de risco e sinalizando opacidade informacional (Pouryousof et al., 2023). Em contrapartida, a consistência entre discurso e desempenho é percebida como um sinal crível de transparência, credibilidade e alinhamento com práticas de governança efetiva (Pouryousof et al., 2023).

Esses achados, em conjunto, fortalecem a compreensão de que o tom dos relatos corporativos não constitui uma escolha discursiva meramente retórica, mas um mecanismo informacional observável, capaz de modular a percepção dos investidores sobre risco, governança e credibilidade da firma (Carlini et al., 2020; González et al., 2019; Huang et al., 2014; Pouryousof et al., 2023). No âmbito da Teoria da Sinalização (Spence, 1973), o tom funciona como um canal reputacional complementar aos mecanismos estruturais de governança, especialmente em ambientes caracterizados por elevada assimetria informacional.

A incerteza textual expressa nos relatos corporativos funciona como um sinal não financeiro capaz de refletir a qualidade informacional e da robustez das práticas de governança. Níveis mais elevados de incerteza na comunicação são associados a maior opacidade informacional, aumento da percepção de risco e menor qualidade

dos fundamentos econômicos e das práticas corporativas (Li, 2008; González et al., 2019; Pourousof et al., 2023).

Incerteza textual elevada está negativamente relacionada à persistência dos resultados, indicando maior assimetria informacional e menor previsibilidade dos fluxos futuros (Li, 2008). No contexto latino-americano, a presença de incerteza nos prospectos de IPO sinaliza risco informacional, elevando o nível de *underpricing* das ofertas públicas iniciais, o que reforça que o mercado precifica ativamente os sinais linguísticos (González et al., 2019).

Esse efeito confirma que, embora estruturas formais de governança sejam relevantes, o mercado atribui peso significativo aos atributos narrativos, particularmente ao grau de certeza ou incerteza transmitido, como proxy da credibilidade, da opacidade e da qualidade das informações (González et al., 2019). A incerteza textual, portanto, opera como mecanismo complementar aos sinais estruturais de governança, sobretudo em ambientes caracterizados por elevada assimetria informacional (Li, 2008; Pourousof et al., 2023; Spence, 1973).

A consistência intertemporal desse atributo reforça seu papel como sinal reputacional. Desalinhamentos recorrentes entre a incerteza textual e os fundamentos econômicos, especialmente quando acompanhados de tom excessivamente otimista, são interpretados pelo mercado como sinal de opacidade informacional e potencial mascaramento dos resultados (Pourousof et al., 2023).

Por outro lado, quando a comunicação apresenta níveis controlados e coerentes de incerteza, compatíveis com a realidade operacional e com o ambiente macroeconômico, o mercado interpreta esse padrão como sinal de transparência, credibilidade e alinhamento com práticas de governança efetiva (González et al., 2019; Li, 2008; Pourousof et al., 2023).

Assim como a legibilidade e o tom, a incerteza textual constitui um mecanismo informacional observável e precificado pelos agentes de mercado. Esse atributo reforça a lógica da Teoria da Sinalização (Spence, 1973), funcionando como proxy da qualidade informacional, da robustez das práticas de governança e da credibilidade da firma. Em mercados caracterizados por elevados níveis de assimetria informacional, como os da América Latina, a qualidade da divulgação corporativa é interpretada como reflexo direto da credibilidade e da robustez das práticas corporativas (González et al., 2019).

Contudo, além desses sinais textuais, os investidores também observam atributos estruturais, como a adoção de padrões de governança mais rigorosos e a listagem em mercados internacionais. Esses elementos funcionam como sinais adicionais de comprometimento com a transparência, a proteção dos investidores e a mitigação da opacidade informacional (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018).

A lógica da Teoria da Sinalização, portanto, se materializa tanto na comunicação textual quanto nas escolhas institucionais das empresas. Enquanto a clareza, o tom, a incerteza e a consistência discursiva operam como sinais informacionais de curto prazo, decisões como a adesão a regimes de governança mais rigorosos e a listagem em mercados estrangeiros representam sinais estruturais de longo prazo, observáveis e precificados pelos agentes de mercado.

A decisão de listar ações em mercados mais desenvolvidos, por meio de ADRs, funciona como um mecanismo robusto de sinalização estrutural. De um lado, essa estratégia reforça as práticas de governança formal ao submeter a empresa a exigências mais rigorosas de divulgação das informações, fiscalização regulatória (*enforcement*) e proteção aos investidores (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018; Stulz, 2022). Embora os estudos sobre características textuais não analisem diretamente os efeitos da listagem em ADRs sobre o conteúdo narrativo, é possível inferir, com base na literatura de disclosure (Dye, 1985; Healy & Palepu, 2001; Verrecchia, 2001), que a adesão a padrões informacionais mais exigentes tende a favorecer práticas de comunicação mais claras, transparentes e alinhadas às expectativas de investidores globais.

Esse mecanismo se ancora na lógica da Teoria da Sinalização (Spence, 1973), ao combinar escolhas institucionais observáveis, como a adesão a mercados com *enforcement* (fiscalização regulatória) mais rigoroso, com melhorias na qualidade informacional percebida. A decisão pela listagem internacional não apenas indica comprometimento com padrões mais elevados de governança, como também reduz a assimetria informacional e fortalece a reputação corporativa.

Esse efeito se conecta diretamente ao arcabouço teórico de *disclosure* (divulgação corporativa) (Healy & Palepu, 2001; Verrecchia, 2001), segundo o qual a divulgação de informações claras, precisas e acessíveis reduz os custos de processamento informacional, mitiga os efeitos da seleção adversa e melhora a eficiência das decisões dos investidores. A partir da integração dos pressupostos da

Teoria da Sinalização (Spence, 1973) e dos modelos teóricos de divulgação corporativa (Dye, 1985; Healy & Palepu, 2001; Verrecchia, 2001), bem como das evidências empíricas sobre o papel dos sinais textuais (González et al., 2019; Li, 2008; Loughran & McDonald, 2014; Pouryousof et al., 2023) e dos sinais institucionais (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018), conclui-se que esses mecanismos operam de maneira complementar na construção da credibilidade informacional e na redução da percepção de risco.

A Teoria da Sinalização (Spence, 1973) oferece um arcabouço robusto para compreender como, em ambientes marcados por elevada assimetria informacional, as empresas recorrem simultaneamente a mecanismos textuais, representados pela clareza, tom, incerteza e consistência narrativa e a mecanismos estruturais, como a adoção de padrões de governança mais rigorosos e a listagem internacional via ADRs (Coffee, 2002; Doidge et al., 2004).

A decisão de acessar mercados estrangeiros, particularmente por meio de ADRs, não se restringe a uma estratégia de captação de recursos, mas opera como um sinal observável de comprometimento com práticas de governança robustas, transparência e proteção aos investidores (Doidge et al., 2004). Assim, observa-se que a combinação de mecanismos institucionais, como a adesão a mercados com *enforcement* (fiscalização regulatória) mais rigoroso, via ADRs e práticas comunicacionais de maior qualidade exerce um papel ainda mais relevante em mercados emergentes, como o brasileiro, caracterizados por maiores fragilidades institucionais e elevados níveis de assimetria informacional (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018).

No âmbito da Teoria da Sinalização, a legibilidade, o tom, a incerteza textual e a consistência narrativa operam como proxies da credibilidade informacional, da robustez das práticas de governança e da percepção de risco informacional (González et al., 2019; Li, 2008; Loughran & McDonald, 2014; Pouryousof et al., 2023). Esse conjunto de evidências consolida a compreensão de que a comunicação narrativa transcende sua função meramente instrumental, atuando como mecanismo reputacional capaz de modular a percepção de risco, influenciar a precificação dos ativos e fortalecer a credibilidade da firma perante os *stakeholders*, especialmente em ambientes regulatórios mais exigentes, como aqueles decorrentes da listagem em mercados estrangeiros (Coffee, 2002; Dau et al., 2024; Doidge et al., 2004; González

et al., 2019; Li, 2008; Loughran & McDonald, 2014; Pouryousof et al., 2023; Sayari & Marcum, 2018).

2.3 Legibilidade

Neste estudo, legibilidade se refere à facilidade ou dificuldade com que um texto pode ser lido e compreendido pelo público-alvo (Aldahray, 2024; Flesch, 1948; Hassan et al., 2022; Jayasree & Shette, 2021; Yusuf & Jordan, 2017). Para Yusuf e Jordan (2017) a legibilidade garante que os leitores compreendam o material, processem as informações em uma velocidade adequada e o considerem interessante, sendo que, fatores como a média de comprimento das frases, o número de palavras e a complexidade gramatical influenciam essa facilidade.

De forma semelhante, Jayasree e Shette (2021) destacam que textos com frases longas, palavras desnecessárias e jargões legais e financeiros tornam a leitura mais difícil. Portanto, a legibilidade é necessária para a compreensão dos relatórios e uma legibilidade inadequada amplia a opacidade informacional, aumenta os custos de processamento de informações e eleva a percepção de risco, impactando negativamente a avaliação econômica da empresa pelos investidores e credores (Ertugrul et al., 2017; Saravanan et al., 2024). Dessa forma, relatórios mais claros e acessíveis sinalizam práticas contábeis transparentes, enquanto textos excessivamente complexos dificultam a compreensão de informações relevantes (Ben-Amar & Belgacem, 2018).

Neste sentido, Huang et al. (2016) focam no movimento "*Plain Language Movement*", que busca simplificar a linguagem, evitando termos obscuros, vocabulário excessivamente técnico e construções de frases complicadas, facilitando assim a compreensão dos relatórios financeiros (Huang et al., 2016). Assim, a legibilidade dos relatórios anuais é uma característica importante que determina a qualidade da comunicação entre uma empresa e seus *stakeholders*, além de facilitar a capacidade dos investidores de processar e analisar informações. Isso resulta em um melhor desempenho da empresa e em menores custos de agência (Dalwai et al., 2021). Ou seja, é um atributo importante da informação textual e tem sido examinada extensivamente em vários campos (Luo et al., 2018).

A mensuração da legibilidade textual fundamenta-se em trabalhos pioneiros, como o Flesch (1948), que desenvolveu o *Flesch Reading Ease*, uma métrica que

quantifica a facilidade de leitura com base no comprimento médio das frases (em palavras) e no número médio de sílabas por palavra. O escore gerado apresenta interpretação inversa à dificuldade de leitura: quanto mais elevado, maior a facilidade de compreensão do texto.

Klare (1974-1975), em uma extensa revisão sobre legibilidade, reforça que a eficácia e popularidade dessas métricas decorrem de sua simplicidade operacional, adaptabilidade a diferentes tipos de textos e robustez empírica. Klare (1974-1975) ressalta que, apesar das limitações na captura de aspectos semânticos e discursivos mais subjetivos, como clareza conceitual e coerência argumentativa, tais índices permanecem como ferramentas válidas e consistentes para avaliar a complexidade estrutural dos textos.

O uso dessas métricas transbordou o campo educacional, consolidando-se na literatura contábil e financeira. Estudos como os de Hrasky et al. (2009), Huang et al. (2016), Pajuste et al. (2021) e Toerien & du Toit (2024) demonstram que o Índice Flesch e métricas análogas são amplamente aplicados na análise de relatórios anuais, cartas aos acionistas, seções de análise e discussão da administração (MD&A) e notas explicativas. Esses trabalhos destacam que a legibilidade, mensurada por tais fórmulas, permite identificar práticas de ofuscação, avaliar padrões de divulgação das informações e até mesmo inferir relações com desempenho corporativo e práticas de governança.

Contudo, uma limitação metodológica relevante refere-se à aplicação direta de fórmulas desenvolvidas para a língua inglesa em textos escritos em outros idiomas, como o português. De acordo com Martins et al. (1996), essa restrição decorre de diferenças linguísticas significativas, especialmente no que tange à morfossintaxe. Em média, a língua portuguesa possui 2,2 sílabas por palavra, em contraste com 1,8 no inglês, além de apresentar estruturas sintáticas mais complexas. Como resultado, o uso não ajustado da fórmula de Flesch em textos em português leva à subestimação sistemática da legibilidade.

Martins et al. (1996) realizaram um estudo empírico utilizando textos paralelos, isto é, textos originalmente escritos em inglês e suas respectivas traduções para o português e identificaram que, mantendo-se constante o conteúdo semântico, os textos em português apresentaram uma redução média de entre 35 e 42 pontos no escore de legibilidade. Para mitigar esse viés, os autores propuseram uma adaptação da fórmula, baseada na adição de um fator de correção aos escores obtidos. Embora

o estudo utilize a expressão "modificação da fórmula", a adaptação consiste operacionalmente na adição de uma constante (+42) aos escores obtidos, sem alteração dos coeficientes estruturais da equação original.

Dessa forma, o Índice Flesch Adaptado para o Português, conforme proposto por Martins et al. (1996), oferece uma métrica mais aderente às especificidades linguísticas do português brasileiro, mantendo a simplicidade operacional e a robustez teórica da versão original. A utilização do Índice de Flesch em estudos sobre legibilidade de relatórios corporativos encontra respaldo sólido na literatura internacional. Desde os trabalhos pioneiros de Pashalian e Crissy (1950), que avaliaram a legibilidade dos relatórios anuais corporativos norte-americanos utilizando a fórmula de Flesch, há evidências consistentes de que tais documentos tendem a apresentar elevada complexidade textual, dificultando a compreensão por parte de acionistas e *stakeholders*.

Pesquisas mais recentes ampliaram essa abordagem para diferentes contextos e estruturas narrativas dos relatórios financeiros. Huang et al. (2016), por exemplo, analisaram a legibilidade de três seções distintas: i) *Compensation Discussion and Analysis* (CD&A), ii) *Management Discussion and Analysis* (MD&A) e iii) *Letter to Shareholders* (carta aos acionistas). Huang et al. (2016) constataram que, embora o CD&A tenha se tornado mais legível após a implementação da iniciativa *Plain English* pela SEC, ele permanece menos acessível do que a carta aos acionistas, mas mais legível que o MD&A.

Ginesti et al. (2017) demonstraram que o grau de conexões nos conselhos de administração de empresas italianas impacta positivamente a legibilidade dos MD&As, indicando que mecanismos de governança corporativa influenciam a qualidade da comunicação financeira. No contexto dos mercados emergentes, Pajuste et al. (2021) aplicaram as métricas de Flesch e Fog para avaliar a legibilidade dos MD&As de empresas nos Estados Bálticos, reforçando a aplicabilidade da metodologia em diferentes ambientes institucionais. Saravanan et al. (2024) corroboraram esses achados ao verificar que a adoção das IFRS no Reino Unido resultou em relatórios anuais significativamente mais longos e complexos, sendo as notas explicativas e o MD&A as seções que mais contribuíram para esse aumento de complexidade textual.

Ben-Amar e Belgacem (2018), por sua vez, ao investigar o impacto das práticas de responsabilidade social corporativa (CSR) sobre a complexidade textual dos MD&As, concluiu que empresas mais engajadas em CSR tendem a elaborar

documentos mais longos e menos legíveis. Esse achado sugere o uso potencial de divulgação de informações sociais como instrumento de ofuscação, reforçando o papel da legibilidade como *proxy* para práticas de transparência e governança.

Esses estudos confirmam que o Índice de Flesch, embora originalmente desenvolvido para o idioma inglês e posteriormente adaptado para o português (Martins et al., 1996), possui ampla aceitação na literatura contemporânea aplicada à análise de narrativas financeiras, tanto em seções estruturadas como MD&A e notas explicativas quanto em comunicações discursivas, como cartas aos acionistas.

Conforme evidenciado na literatura, diversos autores têm aplicado o Índice de Flesch, e métricas análogas, para analisar a legibilidade de diferentes seções dos relatórios corporativos, como *Management Discussion and Analysis* (MD&A), *Compensation Discussion and Analysis* (CD&A), notas explicativas e, em alguns casos, cartas aos acionistas, corroborando sua aplicabilidade em contextos financeiros e empresariais (Ben-Amar & Belgacem, 2018; Ginesti et al., 2017; Huang et al., 2016; Pajuste et al., 2021; Pashalian & Crissy, 1950; Saravanan et al., 2024).

A **Tabela 1** apresenta uma sistematização de alguns dos estudos que aplicaram o Índice de Flesch, para mensurar a legibilidade de narrativas financeiras e relatórios corporativos. Esses trabalhos ilustram a ampla adoção da métrica em diferentes contextos, incluindo relatórios anuais, cartas aos acionistas, seções de análise da administração (MD&A), notas explicativas e outros documentos de divulgação corporativa. Embora a tabela não esgote a literatura existente, ela evidencia a recorrência do uso do índice como ferramenta válida e amplamente reconhecida para avaliar a complexidade textual e, consequentemente, a qualidade da comunicação financeira.

Tabela 1

Sistematização de Alguns Estudos com o Índice Flesch.

Autor(es)	Objetivo	Seção Avaliada	Métrica	Principais Resultados
Flesch (1948)	Desenvolver um índice objetivo de legibilidade aplicável a textos diversos.	Textos gerais, incluindo relatórios	Flesch Reading Ease	Criação da fórmula Flesch. Validação empírica para diferentes tipos de documentos, incluindo materiais corporativos e governamentais.
Pashalian & Crissy (1950)	Avaliar a legibilidade de	Relatórios Anuais	Flesch Reading Ease	Identificaram que os relatórios eram excessivamente difíceis de ler. Recomendam a

Tabela 1*Sistematização de Alguns Estudos com o Índice Flesch.*

Autor(es)	Objetivo	Seção Avaliada	Métrica	Principais Resultados
Hrasky et al. (2009)	relatórios anuais corporativos.			adoção de práticas para simplificar a linguagem corporativa.
	Investigar a relação entre desempenho corporativo e a legibilidade das cartas aos acionistas.	Carta aos Acionistas	Flesch Reading Ease	Empresas com pior desempenho tendem a produzir textos mais complexos e menos legíveis, sugerindo uso de ofuscação.
Huang et al. (2016)	Avaliar a legibilidade do CD&A, MD&A e carta aos acionistas nos EUA.	CD&A, MD&A, Carta aos Acionistas	Flesch, Fog e Plain English	A legibilidade melhorou após a <i>Plain English Initiative</i> , mas o CD&A permaneceu menos legível que a carta aos acionistas e mais legível que o MD&A.
Ginesti et al. (2017)	Analizar o impacto das conexões no conselho na legibilidade dos MD&As em empresas italianas.	MD&A	Flesch Reading Ease	Conselhos com maior interconectividade promovem relatórios mais legíveis, sugerindo governança mais eficiente na comunicação.
Pajuste et al. (2021)	Analizar a relação entre complexidade textual e earnings management no mercado báltico.	MD&A	Flesch e Fog	<i>Earnings management</i> se associa à complexidade textual apenas nas empresas mais líquidas. Não há relação significativa na amostra completa.
Saravanan et al. (2024)	Examinar o impacto da adoção das IFRS na legibilidade dos relatórios anuais de empresas do Reino Unido.	MD&A, Notas Explicativas, Relatórios Anuais	Flesch e Fog	Adoção das IFRS levou a relatórios mais longos e menos legíveis. As notas explicativas foram as seções que mais contribuíram para essa redução de legibilidade.
Toerien & du Toit (2024)	Avaliar se mudanças nas normas contábeis (IAS 39 e IFRS 9) melhoraram a legibilidade das divulgações de risco.	Notas Explicativas e Divulgações de Risco	Flesch e Fog	As divulgações se tornaram mais longas e menos legíveis após as mudanças nas normas, contrariando a expectativa de melhora na clareza da comunicação.
Martins et al. (1996)	Adaptar a fórmula de Flesch para o idioma português.	Textos técnicos em português	Flesch (Adaptado)	Demonstraram que o Flesch subestima a legibilidade em português. Propuseram um ajuste que adiciona +42 pontos para corrigir o viés morfossintático da língua portuguesa.

Fonte: Dados da Pesquisa, 2025.

Além da métrica principal baseada no Índice de Flesch, a literatura recomenda o uso de variáveis complementares capazes de capturar dimensões adicionais da

complexidade textual e da sobrecarga cognitiva associada à leitura de documentos corporativos. Entre essas variáveis, destaca-se o comprimento do documento (*Length*), operacionalizado como o número total de palavras, uma *proxy* amplamente aceita para mensurar a carga informacional e os custos de processamento associados (Li, 2008). Documentos excessivamente longos podem representar barreiras cognitivas relevantes, dificultando a compreensão e, portanto, reduzindo a efetividade da comunicação financeira.

Outra métrica frequentemente utilizada é o *Common Words Index*, que avalia a proporção de palavras comuns em relação ao total de palavras no texto. Estudos como Abernathy et al. (2019) evidenciam que uma maior frequência de palavras de uso comum está associada a maior acessibilidade textual, funcionando como um indicativo de textos mais fáceis de ler e compreender. Por fim, a densidade de termos técnicos do jargão financeiro pode ser mensurada por meio de índices construídos a partir de dicionários especializados, como o *Financial Words Index*, que quantifica a frequência de termos financeiros no texto. Essa métrica é fundamentada na lógica dos léxicos especializados desenvolvidos por Loughran & McDonald (2011) e foi operacionalizada, por exemplo, por Abernathy et al. (2019), com base no *Campbell Harvey's Finance Glossary*, permitindo aferir o grau de tecnicidade das narrativas corporativas.

A literatura de divulgação corporativa sugere que a qualidade textual das informações narrativas exerce um papel central na mitigação das assimetrias informacionais, no fortalecimento da transparência e na eficiência dos mercados (Courtis, 2004; Healy & Palepu, 2001). Nesse contexto, a legibilidade dos relatórios, entendida como a facilidade com que os *stakeholders* conseguem processar, interpretar e utilizar as informações divulgadas, emerge como um atributo fundamental da comunicação corporativa eficaz (Ginesti et al., 2017).

Estudos anteriores documentam que comunicações corporativas excessivamente complexas, prolixas ou tecnicamente densas aumentam os custos de processamento da informação, ampliam a opacidade textual e reforçam a percepção de risco, sobretudo em mercados caracterizados por menor *enforcement* regulatório e baixa literacia financeira (Ertugrul et al., 2017; Li, 2008; Saravanan et al., 2024). Nesse sentido, a complexidade textual dificulta a compreensão e atua como mecanismo oportunista de ofuscação, permitindo que gestores ocultem informações adversas em meio a narrativas redundantes ou excessivamente técnicas (Bar-Hod et al., 2021).

De acordo com a Teoria da Agência (Jensen & Meckling, 1976), a qualidade das divulgações é determinante para alinhar interesses entre gestores e investidores, especialmente porque relatórios narrativos, como o relatório da administração funcionam como instrumentos de monitoramento dos agentes e de mitigação dos custos de agência (Healy & Palepu, 2001; Luo et al., 2018). Nesse arcabouço, a legibilidade não se restringe a um atributo linguístico formal, mas desempenha um papel econômico relevante, ao reduzir a assimetria informacional e aumentar a eficácia dos contratos implícitos entre gestores e *stakeholders*.

Adicionalmente, sob a ótica da Teoria da Sinalização (Spence, 1973), a adoção de práticas comunicacionais claras e acessíveis atua como um sinal crível de comprometimento com a transparência, a governança e a responsabilidade corporativa. Empresas que priorizam a legibilidade em seus relatórios transmitem ao mercado uma sinalização de qualidade, reduzindo a incerteza percebida e, consequentemente, o risco associado aos seus ativos (Liu, 2021).

Os resultados empíricos da literatura corroboram essa relação. Evidências mostram que níveis mais elevados de legibilidade estão associados a menores custos de agência, à maior eficácia das divulgações e à redução da opacidade informacional (Dalwai et al., 2021; Ginesti et al., 2017; Luo et al., 2018). Por outro lado, relatórios excessivamente complexos, seja pela linguagem técnica, estruturas sintáticas rebuscadas ou informações desnecessárias, elevam a opacidade, dificultam a interpretação de riscos de mercado e aumentam a percepção de incerteza por parte dos investidores (Ertugrul et al., 2017; Saravanan et al., 2024).

No contexto específico dos mercados emergentes, esses efeitos tendem a ser potencializados, dado o ambiente institucional caracterizado por fragilidades regulatórias (González et al., 2021) e menores níveis de literacia financeira e baixa rigidez na fiscalização regulatória (Saravanan et al., 2024). Nesse ambiente, a clareza textual melhora o processamento das informações, contribui para reduzir a opacidade e fortalece a confiança dos investidores, sobretudo em mercados emergentes caracterizados por fragilidades regulatórias e baixa literacia financeira (Dalwai et al., 2021; Ertugrul et al., 2017; González et al., 2021; Healy & Palepu, 2001; Jensen & Meckling, 1976; Luo et al., 2018; Saravanan et al., 2024) e da Teoria da Sinalização (Dalwai et al., 2021; Ginesti et al., 2017; Spence, 1973), e ancorada em uma ampla base de evidências empíricas (Ertugrul et al., 2017; Li, 2008; González et al., 2019; Saravanan et al., 2024), sustenta-se que maiores níveis de legibilidade contribuem

para a redução da incerteza textual nas divulgações narrativas corporativas, especialmente nos relatórios da administração, ao tornarem as informações mais acessíveis, transparentes e menos sujeitas à ambiguidade (Bar-Hod et al., 2021).

Diante desse arcabouço, espera-se que a legibilidade dos relatórios da administração atue como um mecanismo eficaz de mitigação da incerteza textual, na medida em que uma comunicação mais clara, objetiva e bem estruturada reduz a ambiguidade, melhora o entendimento sobre riscos e estratégias e, consequentemente, fortalece a transparência informacional (Dalwai et al., 2021; Ginesti et al., 2017; Luo et al., 2018). Assim, sustenta-se que quanto maior o nível de legibilidade, menor será a incerteza textual expressa nas divulgações narrativas das empresas, sendo formulada a seguinte hipótese:

H₁: Existe uma associação negativa entre a legibilidade dos relatórios da administração e a incerteza textual.

2.4 Tom (Sentimento Textual)

Neste estudo, o tom se refere às formas linguísticas positivas ou negativas utilizadas para apresentar informações aos usuários externos dos relatórios (Hassan et al., 2022). De acordo com Henry (2008, p. 376), o tom é definido como “o afeto ou sentimento transmitido por uma comunicação”, representando, portanto, o efeito da comunicação, destacando seu papel promocional nos comunicados de imprensa sobre lucros, na medida em que termos positivos induzem os investidores a interpretar os resultados em termos de aumentos relativos a pontos de referência (Henry, 2008).

Nesse sentido, o uso de palavras como “diminui”, “deteriorado” e “declínio” caracteriza o tom negativo, enquanto expressões como “aumentou” e “melhorou” definem o tom positivo, que, apesar de corretas, atinge o propósito mais promocional ao selecionar um aspecto positivo do desempenho ou fazer uma comparação positiva explícita ou usar palavras avaliativas explicitamente positivas (Henry, 2008).

O tom, especialmente quando negativo, textual transmite informações e funciona como um mecanismo estratégico de gestão das percepções dos *stakeholders*, especialmente em ambientes marcados por assimetria informacional (Hassan et al., 2022; Henry, 2008; Pouryousof et al., 2022). Contudo, esse efeito não decorre unicamente da presença de palavras com carga negativa, mas da qualidade,

clareza, legibilidade e, sobretudo, da aderência contextual do tom às condições econômicas da firma (Elliott et al., 2024; Hassan et al., 2022; Pouryousof et al., 2022).

Quando os gestores expressam o tom negativo de forma clara, objetiva e não ambígua, transformam esse recurso linguístico em um sinal de prudência, realismo e transparência, atendendo aos critérios de observabilidade, credibilidade e custo de imitação, conforme preceitua a Teoria da Sinalização (Connelly et al., 2011; Spence, 1973).

A Teoria da Agência (Jensen & Meckling, 1976), reconhece que os mercados financeiros são estruturados sob a presença de assimetria informacional e conflitos de interesse entre gestores (agentes) e investidores (principais). Nesse contexto, os gestores detêm informações privadas sobre a condição operacional, os riscos e as perspectivas futuras da empresa, que não estão imediatamente disponíveis aos *stakeholders* externos (Jensen & Meckling, 1976).

Em ambientes caracterizados por assimetria informacional, como os mercados financeiros, os agentes mais bem informados, como os gestores, recorrem a sinais observáveis, como o tom adotado nos relatórios para transmitir informações críveis sobre suas condições operacionais, seus riscos e suas expectativas futuras da organização (Connelly et al., 2011; Spence, 1973). Nessa lógica, o tom textual deixa de ser apenas uma característica linguística e passa a exercer uma função econômica relevante, atuando como um indicador da qualidade da informação, da credibilidade da comunicação e do grau de alinhamento das mensagens divulgadas com os fundamentos operacionais e financeiros da empresa (Pouryousof et al., 2022; Wang et al., 2024).

No contexto das divulgações financeiras, um tom negativo, quando adotado de maneira clara, objetiva e aderente às condições econômicas da empresa, preenche os requisitos de sinal eficaz, funcionando como um mecanismo eficaz de sinalização, capaz de mitigar a opacidade informacional e reduzir a percepção de incerteza textual entre os *stakeholders* (Elliott et al., 2024; Hassan et al., 2022; Pouryousof et al., 2022). Nessa perspectiva, o tom textual se torna uma extensão do processo de sinalização, funcionando como um indicativo da qualidade, da credibilidade e do alinhamento das informações divulgadas aos fundamentos econômicos da empresa (Pouryousof et al., 2022; Wang et al., 2024).

Contudo, em cenários caracterizados por risco, instabilidade operacional ou desempenho adverso, os investidores tendem a abandonar heurísticas

simplificadoras e passam a processar as informações de maneira mais criteriosa e analítica, atribuindo maior peso às comunicações percebidas como claras, objetivas e realistas (Carlini et al., 2020; Elliott et al., 2024). Visto que as más notícias, especialmente quando comunicadas com alto grau de clareza e certeza linguística, recebem maior atenção, passam por análises mais rigorosas e exercem influência mais significativa nos julgamentos dos investidores (Carlini et al., 2020; Elliott et al., 2024).

Nesse contexto, os gestores que estruturam suas divulgações com tom claro, realista e aderente aos riscos enfrentados pela organização contribuem diretamente para reduzir a assimetria informacional, fortalecer a credibilidade da divulgação de informações e mitigar a percepção de incerteza textual (Elliott et al., 2024; Ertugrul et al., 2017; Hassan et al., 2022; Pouryousof et al., 2022). As comunicações com tom mais claro e realista, particularmente quando associadas a eventos adversos ou notícias negativas, geram reações significativas no mercado, reforçando seu papel como mecanismo de transmissão de informações e de percepção dos riscos operacionais, financeiros e econômicos enfrentados pela firma (Carlini et al., 2020).

Nessa linha, quando confrontados com más notícias, os investidores abandonam o uso de heurísticas simplificadoras e passam a processar as informações de maneira mais criteriosa (Elliott et al., 2024). Essa atenção redobrada aos detalhes leva a uma maior credibilidade das comunicações percebidas como claras, objetivas e realistas. Esse efeito cognitivo é maximizado quando o tom negativo é apresentado de forma direta, reduzindo ambiguidade e facilitando o alinhamento das expectativas dos *stakeholders* (Carlini et al., 2020; Elliott et al., 2024). Notícias negativas comunicadas com alto grau de certeza linguística são mais efetivamente absorvidas pelo mercado, atribuindo maior valor informacional à clareza do discurso em contextos adversos (Carlini et al., 2020).

Além do tom, a legibilidade é um fator crucial. Em cenários de alto risco financeiro, gestores adotam tom negativo combinado a alta legibilidade como estratégia para clarificar as causas dos riscos enfrentados pela organização (Hassan et al., 2022). Essa abordagem facilita o entendimento por parte dos investidores e contribui para a redução da percepção de incerteza, sinalizando comprometimento com a transparência informacional (Hassan et al., 2022).

Em situações caracterizadas por desempenho adverso, instabilidade operacional ou riscos macroeconômicos, comunicações estruturadas com menor

ambiguidade de tom, maior clareza e elevada legibilidade atendem de maneira mais eficaz às expectativas informacionais dos investidores. Esse padrão comunicacional contribui diretamente para mitigar a assimetria informacional e reduzir a percepção de incerteza sobre os riscos operacionais, financeiros e estratégicos enfrentados pela firma (Ertugrul et al., 2017; Hassanein & Albitar, 2024; Saravanan et al., 2024).

Sendo assim, a comunicação de eventos negativos, quando realizada com alto grau de certeza linguística, isto é, com menor ambiguidade textual, tem impacto mais efetivo no mercado, evidenciando que a clareza e objetividade da mensagem determinam a eficácia da transmissão informacional (Carlini et al., 2020). O tom textual presente nas divulgações corporativas afeta significativamente o comportamento dos analistas, impactando tanto a precisão de suas previsões sobre resultados financeiros futuros quanto suas decisões de acompanhamento da empresa (Zeng et al., 2023).

Por fim, estudos como os de Kang et al. (2018) e Li et al. (2019) evidenciam que as informações qualitativas extraídas das seções narrativas dos relatórios, especialmente o tom textual, possuem valor informacional incremental, alinhando-se aos dados numéricos tradicionais na sinalização do desempenho operacional e financeiro da firma. O tom das seções narrativas dos relatórios reflete o desempenho operacional e financeiro das empresas, funcionando como um complemento informacional relevante aos dados quantitativos (Kang et al., 2018).

Quando há alinhamento entre dados textuais e numéricos, reforça-se que o tom cumpre uma função descritiva e atua como um mecanismo linguístico capaz de reduzir assimetria informacional, à medida que oferece aos investidores sinais adicionais sobre a real condição econômica da firma, mitigando, portanto, a percepção de opacidade e incerteza nas divulgações (Kang et al., 2018). Sendo assim, o tom textual desempenha um papel informacional incremental, ao antecipar resultados futuros e refletir aspectos qualitativos que não estão plenamente capturados nos dados contábeis tradicionais (Li et al., 2019).

Ou seja, o tom influencia significativamente a percepção dos investidores, molda suas expectativas sobre o desempenho da firma e impacta suas reações no mercado, sobretudo em contextos de maior incerteza ou quando as informações quantitativas são insuficientes ou ambíguas (Li et al., 2019). Apesar dos benefícios associados ao uso de tom negativo e linguagem clara, a literatura alerta que determinadas práticas discursivas podem, inversamente (excesso de informações, linguagem ambígua, prolixas ou excessivamente complexas), comprometer a qualidade

informacional, elevando a opacidade e a percepção de risco entre os *stakeholders* (Bar-Hod et al., 2021; Ertugrul et al., 2017; Hassanein & Albitar, 2024; Saravanan et al., 2024).

A hipótese de que níveis mais elevados de tom negativo reduzem a incerteza assume que esse tom é genuíno e informativo, e não um instrumento de gestão de impressão ou ofuscação (Ayuningtyas & Harymawan, 2021; Li, 2008; Rajabalizadeh, 2023; Reichmann, 2023). Gestores podem usar a complexidade textual e o tom para ofuscar informações desfavoráveis (Ayuningtyas & Harymawan, 2021; Li, 2008; Dalwai et al., 2021; Rajabalizadeh, 2023).

Reichmann (2023) define o conceito de “*tone management*” (gerenciamento de tom) como a parcela do tom que não é explicada pelos fundamentos econômicos da empresa, representando, portanto, um componente discricionário do discurso (Reichmann, 2023). Esse gerenciamento de tom está positivamente associado ao risco de crash nos preços das ações, especialmente quando os relatórios são mais legíveis (Reichmann, 2023). Isso revela um efeito paradoxal: quanto maior a clareza textual, maior também o impacto do tom manipulado na percepção de incerteza, uma vez que a legibilidade amplia a capacidade dos investidores de identificar incoerências, omissões ou sinais de falta de credibilidade na comunicação.

Nesse contexto, Reichmann (2023) demonstra que a ausência de um tom negativo genuíno, isto é, quando os gestores omitem, de forma estratégica, sinais sobre riscos ou desafios futuros (*less downward tone management*) eleva a opacidade informacional, e contribui diretamente para o aumento da incerteza e do risco de colapso nos preços. O equilíbrio no volume de informações divulgadas, evitando tanto a omissão quanto o excesso, é fundamental para garantir que a divulgação cumpra sua função informacional de forma eficiente, conforme demonstram os efeitos não lineares da quantidade de informações sobre o valor da empresa (Hassanein & Albitar, 2024).

A divulgação até um nível ótimo contribui para reduzir a incerteza informacional, mas quando esse limite é ultrapassado, o excesso de informações, muitas vezes prolixas ou genéricas, gera sobrecarga cognitiva e deterioração da eficácia comunicacional (Hassanein & Albitar, 2024). Este entendimento é reforçado por evidências de que o excesso de complexidade e baixa legibilidade comprometem a qualidade informacional, aumentando a percepção de risco (Ertugrul et al., 2017; Saravanan et al., 2024).

O fenômeno dos “fatos enterrados” (“*buried facts*”) aponta que informações críticas podem ser soterradas sob discursos prolixos, genéricos ou excessivamente positivos, dificultando a extração de *insights* relevantes e elevando a percepção de risco e de opacidade informacional (Bar-Hod et al., 2021; Hassanein & Albitar, 2024).

Independentemente do desempenho financeiro real, o tom dos comunicados de imprensa influencia significativamente a percepção dos investidores (Henry, 2008). No entanto, quando o tom se torna excessivamente otimista, dissociado dos fundamentos econômicos, ocorre uma deterioração da credibilidade, pois os investidores tendem a perceber esse tom como menos crível, especialmente quando as notícias são boas (Elliott et al., 2024).

Essa percepção de artificialidade no tom textual está associada à inconsistência discursiva, frequentemente observada quando o tom adotado pelos gestores não reflete de maneira fidedigna os fundamentos operacionais ou os riscos enfrentados pela empresa (Pouryousof et al., 2022). Contudo, quando há omissão de riscos ou uso excessivo de linguagem positiva, vaga ou excessivamente técnica, isso é interpretado pelo mercado como uma tentativa de ofuscação informacional, sobretudo em períodos de crise, comprometendo a percepção de transparência e aumentando a incerteza (Pouryousof et al., 2022).

Consequentemente, relatórios com tom ambíguo, combinado à baixa legibilidade, estão fortemente associados a maiores custos de empréstimo, pois a ambiguidade textual amplia a percepção de risco e a incerteza entre os credores (Ertugrul et al., 2017). Por outro lado, quando o tom adotado nas divulgações é mais cauteloso, claro e coerente com o contexto operacional da firma, as informações tendem a ser processadas de forma mais sistemática pelos investidores, contribuindo para aumentar a credibilidade da comunicação (Elliott et al., 2024).

Diante do arcabouço teórico e empírico apresentado, consolida-se que o tom negativo, quando adotado de forma clara, objetiva e alinhada às condições operacionais da empresa, atua como um sinal linguístico de prudência, realismo e aderência aos fundamentos econômicos, contribuindo diretamente para a mitigação da assimetria informacional (Elliott et al., 2024; Hassan et al., 2022).

Conforme prevê a Teoria da Sinalização (Connelly et al., 2011 ; Spence, 1973, e à luz dos achados de Hassan et al. (2022), Elliott et al. (2024) e Carlini et al. (2020), defende-se que, em ambientes marcados por elevada assimetria informacional, como mercados emergentes, cenários de crise ou instabilidade operacional, a adoção de

tom negativo exerce uma função informacional central na redução da incerteza, contribuindo para tornar as divulgações narrativas mais transparentes, informativas e alinhadas às expectativas dos *stakeholders*. Esse padrão discursivo fortalece a credibilidade das informações narrativas, contribui para alinhar as expectativas dos *stakeholders* e mitiga tanto a opacidade quanto a percepção de incerteza textual (Bar-Hod et al., 2021; Carlini et al., 2020; Connelly et al., 2011; Elliott et al., 2024; Ertugrul et al., 2017; Hassan et al., 2022; Spence, 1973). Assim, formula-se a seguinte hipótese:

H₂: Existe uma associação negativa entre a interação de legibilidade com o tom negativo dos relatórios da administração e a incerteza textual.

2.5 Governança corporativa

Empresas que acessam o mercado norte-americano estão sujeitas a padrões regulatórios e de governança significativamente mais rigorosos, impostos pela SEC e pela SOX (Sayari & Marcum, 2018; Stulz, 2022). A listagem via ADRs funciona como um mecanismo externo de governança, capaz de reduzir os custos de agência e mitigar a assimetria informacional entre gestores e investidores (Coffee, 2002).

Este estudo parte do pressuposto, ancorado na Teoria da Agência e na Teoria da Sinalização, de que a governança corporativa exerce papel central na melhoria da qualidade da informação divulgada, pois reforça a transparência, aprimora a legibilidade dos relatórios, disciplina o uso do tom e contribui para reduzir a incerteza informacional (Ayuningtyas & Harymawan, 2021; Carlini et al., 2020; Ertugrul et al., 2017; Hassan et al., 2022; Pourousof et al., 2022). Assim, a listagem via ADRs representa, além de um processo de internacionalização, um indicador robusto de governança corporativa, utilizado neste estudo para mensurar padrões mais elevados de governança entre empresas brasileiras listadas na B3.

Esse entendimento está diretamente alinhado ao papel informacional da divulgação corporativa, que se mostra essencial para a eficiência dos mercados de capitais, na medida em que contribui para mitigar a assimetria de informações e os conflitos de agência entre gestores e investidores externos (Healy & Palepu, 2001). Neste contexto, a governança corporativa atua como um mecanismo disciplinador, capaz de alinhar interesses, reduzir custos de agência e fortalecer a credibilidade dos

reportes financeiros, aprimorando a qualidade da informação divulgada (Healy & Palepu, 2001).

Governança corporativa é definida como um conjunto de mecanismos internos e externos pelos quais os fornecedores de financiamento para as empresas se asseguram de obter um retorno sobre seu investimento e como eles controlam os gerentes (Shleifer & Vishny, 1997). Uma governança corporativa eficaz contribui para alinhar os interesses entre acionistas e gestores, mitigando custos de agência e reduzindo incentivos para comportamentos oportunistas (Gompers et al., 2003; Shleifer & Vishny, 1997).

A Teoria da Agência, proposta por Jensen e Meckling (1976), estabelece que a separação entre propriedade e controle gera problemas de agência, originados, principalmente, pela assimetria informacional, custos de monitoramento, estruturas de incentivo e custos residuais. Nesse contexto, a governança corporativa consiste em mecanismos por meio dos quais os fornecedores de capital controlam os gestores, asseguram retorno sobre seu investimento e limitam comportamentos oportunistas, contribuindo para o alinhamento de interesses e a mitigação da assimetria informacional (Shleifer & Vishny, 1997).

Complementarmente, a Teoria da Sinalização (Spence, 1973), explica que, em ambientes marcados por assimetrias informacionais, os agentes podem recorrer a sinais observáveis, críveis e economicamente custosos para transmitir informações sobre características não observáveis, como qualidade, comprometimento ou credibilidade. No contexto corporativo, práticas como a divulgação voluntária de riscos ou o alinhamento com padrões mais rigorosos de governança operam como sinais que reduzem a assimetria informacional e a percepção de incerteza pelos investidores (Spence, 1973).

Além disso, conforme postulado pela Teoria da Agência, os conflitos entre agentes e principais podem ser mitigados por meio de mecanismos de governança eficazes, que atuam como instrumentos de monitoramento e alinhamento de interesses (Jensen & Meckling, 1976). Adicionalmente, conselhos de administração mais bem conectados, do ponto de vista social, profissional e informacional, tendem a melhorar a legibilidade e a transparência dos relatórios narrativos, como a seção de Discussão e Análise da Administração (MD&A) (Ginesti et al., 2017). Isso ocorre porque o compartilhamento de informações e o enriquecimento do capital humano dos

diretores elevam a qualidade dos reportes corporativos e reduzem práticas de ofuscação textual por parte das gerências (Ginesti et al., 2017).

Nesse sentido, uma governança robusta, composta por conselhos mais independentes, contribui para a mitigação de assimetrias informacionais (González et al., 2019) e, quando associada a comitês de auditoria eficazes e à adoção de auditorias externas de alta qualidade, eleva a qualidade dos relatórios financeiros e narrativos, aprimorando a clareza, a relevância e a utilidade das informações disponibilizadas aos investidores (Hodgdon & Hughes, 2016).

A adoção de práticas de governança mais robustas, como conselhos independentes, auditorias de alta qualidade e comitês eficazes, está associada à redução de informações genéricas ou irrelevantes, à maior clareza dos reportes e à produção de informações mais pertinentes, o que fortalece a credibilidade e a utilidade dos relatórios para os investidores (Hodgdon & Hughes, 2016; Ivashkovskaya & Adamu, 2021). Assim, a governança corporativa não apenas protege os interesses dos acionistas, como também opera como um sinal observável de compromisso com a transparência, a integridade da informação e a redução da incerteza (Coffee, 2002; Dodge et al., 2004; Spence, 1973).

Empresas com estruturas de governança mais alinhadas aos interesses dos acionistas, caracterizadas por maior proteção aos direitos dos acionistas, apresentam desempenho operacional superior, refletido em maiores margens de lucro, retorno sobre o patrimônio (ROE) e crescimento de vendas (Gompers et al., 2003). Esse efeito decorre da atuação dos mecanismos de governança na mitigação dos custos de agência, alinhando os interesses dos gestores aos dos acionistas (Gompers et al., 2003). Assim, uma governança corporativa eficaz, que protege os direitos dos acionistas, alinha melhor os interesses dos gestores aos dos acionistas, diminuindo decisões que favorecem gestores em detrimento dos acionistas, como projetos de investimento subótimos ou uso indevido de recursos (Gompers et al., 2003).

No mercado de capitais, a decisão de listar ações em mercados mais regulados, como os Estados Unidos, por meio de ADRs, opera como um mecanismo de sinalização estrutural (Coffee, 2002; Dodge et al., 2004; Sayari & Marcum, 2018). Essa estratégia impõe às empresas o cumprimento de padrões regulatórios mais rigorosos, como aqueles definidos pela SEC e pela SOX, os quais fortalecem os mecanismos de governança e disciplinam práticas de divulgação das informações (Sayari & Marcum, 2018). A listagem internacional não se restringe à busca por maior

liquidez, mas representa um compromisso observável com padrões mais elevados de governança, transparência e proteção ao investidor (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018).

Sendo assim, a adoção de padrões institucionais mais rigorosos, como aqueles associados à listagem internacional via ADRs, fortalece os mecanismos de governança, disciplina as práticas de divulgação e amplia a proteção dos investidores (Coffee, 2002; Doidge et al., 2004; Sayari & Marcum, 2018). Este fortalecimento institucional, promovido pelo ato de uma empresa, voluntariamente se submeter a um ambiente regulatório mais rigoroso associado à listagem internacional, impõe padrões mais rigorosos de governança e divulgação, o que tende a repercutir na qualidade da comunicação, que passam a refletir maior clareza informacional, alinhando-se às necessidades dos investidores (Coffee, 2002).

Esse efeito decorre do fortalecimento dos mecanismos de governança, que disciplinam as práticas de divulgação das informações e alinham os interesses entre gestores e investidores, reduzindo assim a assimetria informacional e, por extensão, a incerteza textual, especialmente em mercados emergentes, como o Brasil (Brunelli et al., 2021; González et al., 2021; Hodgdon & Hughes, 2016).

Empresas de mercados emergentes recorrem à listagem em mercados como os Estados Unidos para sinalizar comprometimento com padrões mais elevados de governança e transparência. Ao se submeterem às regras da SEC e às exigências da SOX, essas empresas mitigam as limitações institucionais de seus países de origem (Coffee, 2002). Essa decisão melhora o acesso a mercados de capitais e impõe disciplina sobre a divulgação das informações, tornando os relatórios menos opacos e mais alinhados às expectativas dos investidores globais (Coffee, 2002).

Nesta perspectiva, empresas de mercados emergentes que se listam nos Estados Unidos por meio de ADRs experimentam uma melhora significativa na governança e uma maior valoração de mercado (Doidge et al., 2004), o que pode levar a um menor custo de capital e maior transparência. Contudo, o efeito vai além do acesso à liquidez: manifesta-se na adoção de padrões de divulgações mais rigorosos, incluindo maior detalhamento, menor uso de linguagem genérica e relatórios mais informativos, o que contribui para a redução da assimetria informacional e fortalecimento da credibilidade (Sayari & Marcum, 2018; Spence, 1973).

Além disso, empresas de mercados emergentes, listadas nos EUA via ADRs, adotam práticas de divulgação mais rigorosas e extensas, em conformidade com os

padrões regulatórios da SEC e da SOX, o que fortalece os mecanismos de governança e contribui para mitigar riscos associados à assimetria informacional (Sayari & Marcum, 2018). Reduzindo a assimetria informacional e sinalizando ao mercado um compromisso institucional com padrões internacionais de transparência (Sayari & Marcum, 2018).

No contexto das práticas de governança, avanços regulatórios estão associados à melhoria da utilidade das divulgações, especialmente no que se refere às informações sobre riscos e continuidade operacional, aumentando a percepção dos investidores e a relevância econômica dessas informações (Brunelli et al., 2021). As mudanças nos padrões regulatórios na Itália resultaram em divulgações financeiras mais claras e relevantes, especialmente no que se refere aos riscos, o que contribuiu para aumentar a utilidade dessas informações e aprimorar a capacidade dos investidores de avaliar a continuidade operacional (Brunelli et al., 2021).

No contexto latino-americano, verifica-se que empresas que adotam práticas comunicativas mais claras e utilizam um tom mais adequado em seus relatórios apresentam melhor desempenho financeiro e maior valoração de mercado (González et al., 2021). Evidências mostram que, na América Latina, empresas que produzem relatórios mais claros, com menor carga de incerteza e linguagem mais acessível, apresentam melhor desempenho financeiro, refletido em maiores níveis de ROE e maior valoração de mercado (González et al., 2021). Além disso, se constatou que empresas que adotam práticas comunicativas de maior qualidade reduzem significativamente a assimetria informacional (González et al., 2021).

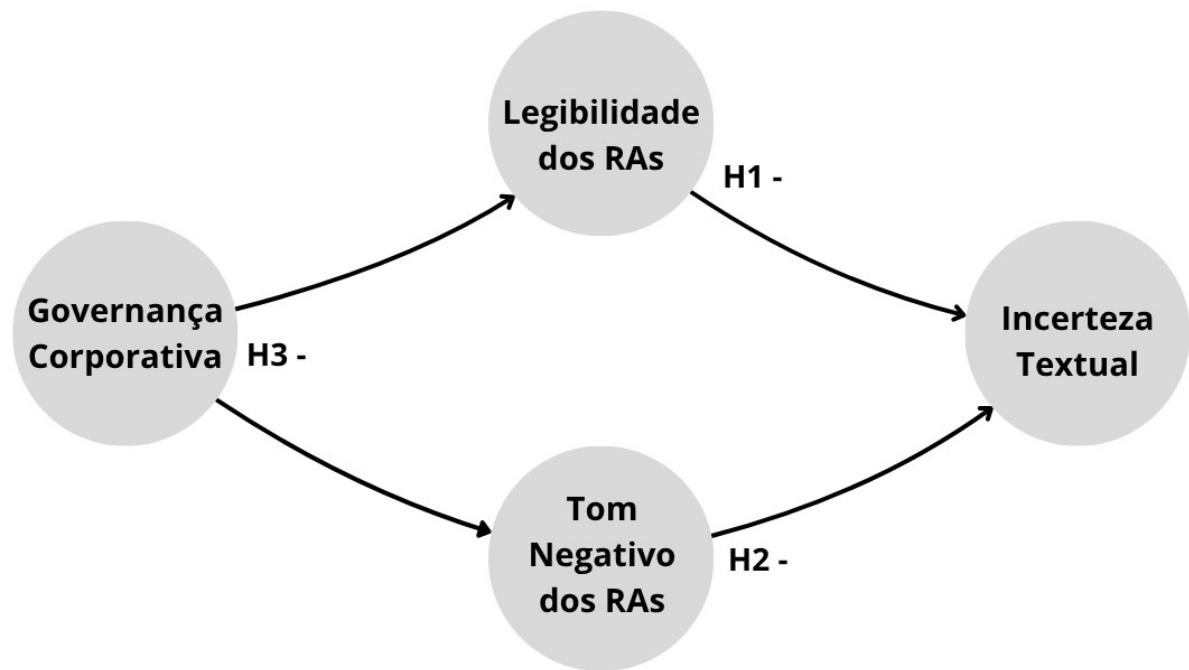
Portanto, à luz da Teoria da Agência e da Teoria da Sinalização, e considerando as evidências empíricas discutidas, conclui-se que a listagem internacional, além de aprimorar os mecanismos formais de governança, repercute na qualidade dos relatos corporativos, especialmente na legibilidade, no tom e na redução da incerteza textual. Assim, formula-se a seguinte hipótese:

H₃: Empresas brasileiras listadas via ADRs, ao adotarem padrões mais rigorosos de governança corporativa, apresentam maior legibilidade e uso mais disciplinado do tom em seus relatórios da administração, o que reduz a incerteza textual.

O desenho da pesquisa e as respectivas hipóteses levantadas estão representadas na **Figura 3**.

Figura 3

Desenho e hipóteses da pesquisa.



Nota: A Figura representa o modelo teórico da pesquisa, no qual a governança corporativa influencia diretamente a legibilidade e o tom negativo dos relatórios da administração, os quais, por sua vez, afetam a incerteza textual. As hipóteses H1, H2 e H3 estão indicadas conforme as relações esperadas no modelo.

Fonte: Elaboração própria (2025).

2.6 Paradoxos da divulgação: entre agência, sinalização e transparência

A literatura sobre divulgação corporativa é majoritariamente ancorada em dois grandes arcabouços teóricos: a Teoria da Agência (Jensen & Meckling, 1976) e a Teoria da Sinalização (Connelly et al., 2011; Spence, 1973). Embora frequentemente tratadas como complementares, essas abordagens teóricas também revelam tensões importantes quando aplicadas ao estudo da comunicação textual corporativa, especialmente no que tange aos relatórios da administração e sua capacidade de reduzir assimetrias informacionais.

A Teoria da Agência parte do pressuposto de que a presença de assimetria informacional e interesses desalinhados entre gestores e acionistas gera incentivos para práticas oportunistas de divulgação (Healy & Palepu, 2001; Jensen & Meckling, 1976). Nesse contexto, a comunicação textual pode ser instrumentalizada como uma

ferramenta de ofuscação (*obfuscation*), seja pela prolixidade, ambiguidade ou pela manipulação do tom e da narrativa (Li, 2008; Loughran & McDonald, 2016; Reichmann, 2023).

Por outro lado, a Teoria da Sinalização pressupõe que, diante da assimetria informacional, gestores de empresas de alta qualidade têm incentivos para emitir sinais críveis, capazes de diferenciar sua firma das demais (Connelly et al., 2011; Spence, 1973). Neste arcabouço, características como maior legibilidade, tom prudente e comunicação clara funcionam como sinais observáveis de governança robusta, transparência e compromisso com a proteção dos investidores (Elliott et al., 2024; Henry, 2008; Pouryousof et al., 2022).

A tensão entre os dois arcabouços é evidente. A mesma característica linguística, como a legibilidade, pode, sob determinadas condições, ser interpretada tanto como um sinal legítimo de transparência quanto como uma estratégia de manipulação. Esse paradoxo se intensifica quando se considera que, em ambientes institucionais frágeis ou com *enforcement* regulatório limitado, como frequentemente ocorre em mercados emergentes, os incentivos para a adoção de práticas oportunistas se sobrepõem aos mecanismos de sinalização crível (La Porta et al., 1998; Stulz, 2022).

O conceito de *tone management* exemplifica de forma clara essa tensão. Estudos como Huang et al. (2014) e Reichmann (2023) demonstram que um tom excessivamente positivo nos relatórios, quando descolado dos fundamentos econômicos, pode não apenas falhar como sinal, mas atuar como uma forma de desinformação, antecipando riscos futuros, como o aumento da volatilidade ou até crash no preço das ações. Por outro lado, o uso de tom negativo, quando associado a uma comunicação clara e objetiva, pode funcionar como um sinal de prudência, conservadorismo e transparência, mitigando a opacidade informacional (Henry, 2008; Elliott et al., 2024).

Nesse cenário, a governança via ADRs emerge como um mitigador das tensões entre oportunismo e sinalização. O ambiente regulatório norte-americano, notadamente, o *enforcement* da SEC e os requisitos da SOX, impõe custos econômicos, reputacionais e legais elevados para práticas de divulgação oportunistas (Coffee, 2002; Hassanein & Albitar, 2024; Stulz, 2022). Assim, a listagem via ADRs funciona tanto como um mecanismo de *bonding* (vinculação externa) quanto de sinalização estrutural, disciplinando os gestores e induzindo uma comunicação textual

mais alinhada às expectativas dos investidores internacionais (González et al., 2021; Sayari & Marcum, 2018).

Embora a literatura frequentemente associe melhorias nas características textuais, como maior legibilidade e tom prudente, a aumentos na transparência, esse vínculo não é necessariamente automático. Existe um *trade-off* inerente entre transparência formal, representada por métricas objetivas derivadas da análise textual, e transparência substantiva, que diz respeito à efetiva utilidade informacional e ao conteúdo econômico transmitido (Bar-Hod et al., 2021; Hassanein & Albitar, 2024).

Por exemplo, aumentar a extensão dos relatórios (*length*) pode, em tese, ser interpretado como maior transparência formal, mas, na prática, pode gerar sobrecarga cognitiva, dificultando a extração de informações relevantes pelos usuários (Li, 2008; Loughran & McDonald, 2016). Reichmann (2023) chama atenção para um paradoxo interessante: quanto mais claro e fácil de entender é um texto, mais ele pode tanto reforçar uma comunicação honesta quanto potencializar os efeitos de um tom manipulador. Isso acontece porque, de um lado, a clareza ajuda os leitores a perceberem sinais informativos; de outro, também pode tornar mais eficientes certas estratégias de persuasão ou distorção da percepção. Por isso, embora medidas como legibilidade, tom e incerteza textual sejam ferramentas valiosas para avaliar a qualidade da comunicação, elas não estão livres de usos oportunistas, sobretudo em contextos em que há fragilidade institucional ou pouca rigidez na aplicação das regras.

Portanto, enquanto métricas como legibilidade, tom e incerteza textual oferecem aproximações objetivas da qualidade da comunicação, elas não são imunes a usos oportunistas, especialmente em ambientes institucionais frágeis ou com baixo *enforcement* regulatório.

Dessa forma, comprehende-se que a relação entre governança corporativa, características linguísticas dos relatórios e incerteza informacional não é linear, mas mediada por mecanismos institucionais que deslocam as empresas ao longo de um espectro que vai da manipulação oportunista à sinalização crível, e da divulgação de informações prolixa e opaca a divulgação de informação clara, informativa e disciplinada. Esse entendimento não apenas fundamenta teoricamente as hipóteses deste estudo, mas também contribui para o avanço da literatura sobre comunicação financeira, governança corporativa e análise textual em mercados emergentes.

Assim, as hipóteses aqui formuladas são fundamentadas na premissa de que a qualidade textual dos relatórios, medida por legibilidade, tom e incerteza, reflete não apenas escolhas estilísticas, mas sobretudo condicionantes institucionais, custos de agência e pressões de governança, internas e externas.

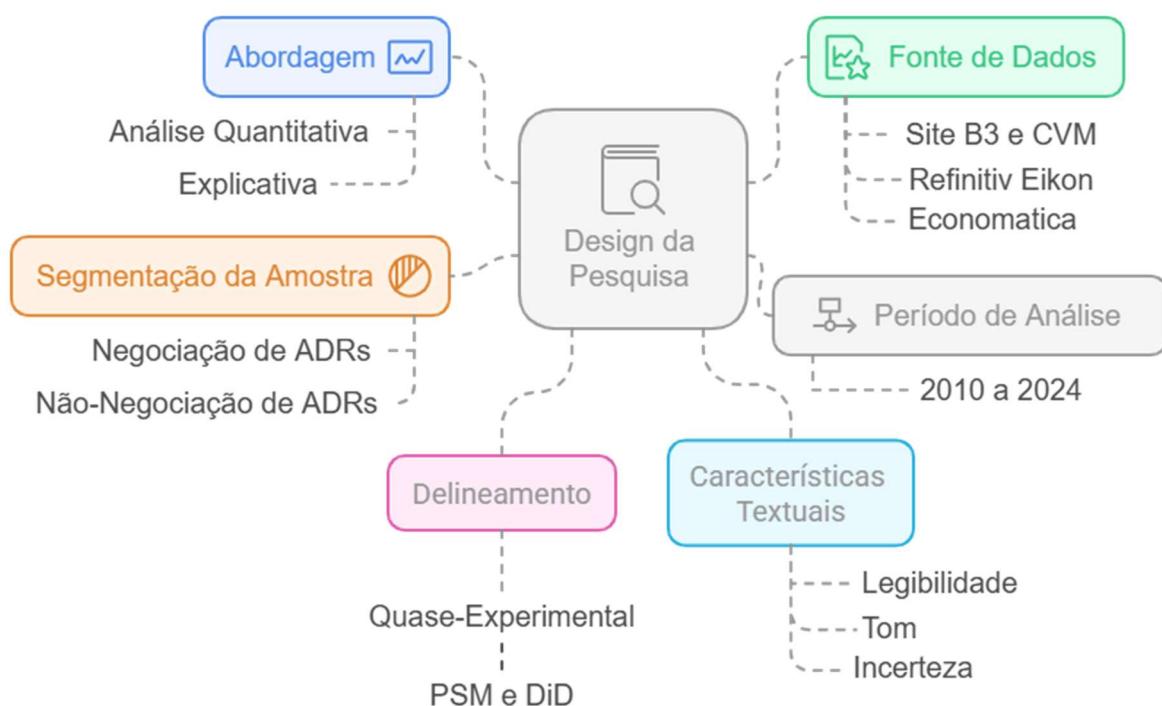
3. METODOLOGIA

3.1 Design da pesquisa e enquadramento

A **Figura 4** apresenta uma síntese visual do *design* metodológico adotado neste estudo, contemplando: a abordagem; as fontes de dados; o período analisado; o delineamento empírico; a segmentação da amostra; e as características textuais examinadas. O estudo se estrutura como uma pesquisa quantitativa e explicativa, com delineamento quase-experimental, fundamentado na aplicação combinada dos métodos PSM e DiD, visando à estimativa de efeitos causais da governança corporativa, listagem via ADRs, sobre a incerteza dos relatórios da administração.

Figura 4

Design da Pesquisa e Enquadramento Metodológico.



Nota: Elaboração própria (2025).

3.2 Procedimentos de coleta e tratamento dos dados

Este estudo tem como ponto de partida as empresas brasileiras não financeiras com ações listadas na B3 entre 2010 e 2024. A partir desse universo, foram definidos dois grupos distintos: o grupo de tratamento, composto por companhias que, ao longo desse período, negociaram ADRs no mercado norte-americano; e o grupo de controle, formado por aquelas que mantiveram suas operações exclusivamente no mercado

doméstico, sem qualquer emissão de ADRs. A identificação das empresas com ADRs foi realizada a partir da base pública da J.P. Morgan, principal depositária de programas brasileiros de ADRs.

A definição do período entre 2010 e 2024 considera a adoção obrigatória das normas internacionais de contabilidade (IFRS) no Brasil, implementada em 2010, e a disponibilidade pública das Demonstrações Financeiras Padronizadas (DFPs) em formato digital a partir desse mesmo ano, nos sites da B3 e CVM, permitindo a coleta estruturada dos dados. O ano de 2024 representa o último exercício social completo com dados disponíveis à época da coleta.

Os dados contábeis e financeiros foram coletados nas plataformas Economatica e Refinitiv Eikon. Já os relatórios da administração foram obtidos por meio do *download* dos arquivos compactados das Demonstrações Financeiras Padronizadas (DFPs), disponíveis nos sites da B3 e da CVM. Em seguida, foi realizado o processamento individualizado de cada arquivo “.pdf” para localizar e extrair a seção “Relatório da Administração”. Após a identificação, o conteúdo em formato “.pdf” foi convertido para “.txt”, utilizando *scripts* no *software R*, visando viabilizar a análise textual automatizada. Durante esse processo, foram removidos elementos não textuais, como quadros, imagens, cabeçalhos e rodapés, com o objetivo de preservar apenas o conteúdo narrativo principal do relatório.

Dificuldades técnicas associadas à conversão de determinados arquivos foram tratadas manualmente. Na primeira rodada, os arquivos com incompatibilidades de codificação, falhas de reconhecimento de texto ou problemas de legibilidade no OCR foram identificados e revisados individualmente com o auxílio do Adobe Acrobat Pro. Contudo, mesmo após o tratamento inicial, alguns arquivos ainda apresentaram dificuldades de extração textual. Aparentemente, parte do texto dessas seções está corrompida ou criptografada, o que faz com que a conversão para .txt produza caracteres ilegíveis ou não processáveis. Nesses casos, os arquivos são descartados na segunda rodada, durante o cálculo dos indicadores textuais. A **Tabela 2** apresenta o quantitativo de arquivos ajustados por ano na primeira rodada.

Tabela 2*Quantitativo de Arquivos PDF Ajustados Manualmente por Ano.*

Ano	Total de Relatórios Disponíveis ¹	Convertidos via Word ²	OCR aplicado ³	Total de arquivos reparados ⁴
2010	575	09	08	17
2011	575	08	05	13
2012	574	06	07	13
2013	568	06	06	12
2014	569	06	03	09
2015	559	01	03	04
2016	528	08	05	13
2017	552	09	06	15
2018	565	11	09	20
2019	577	-	11	11
2020	640	-	23	23
2021	631	-	22	22
2022	592	-	10	10
2023	592	05	04	09
2024	676	01	58	59
Total	8.773	70	180	250

Nota:

¹Refere-se ao total de relatórios disponíveis para *download* no site da B3 e da CVM, incluindo empresas financeiras e não financeiras.

²Conversão aplicada a arquivos originalmente em texto, mas cuja exportação para .txt resultou em símbolos, códigos ou caracteres ilegíveis, exigindo tratamento intermediário via Microsoft Word para recuperação do conteúdo para posterior conversão em PDF e .txt de forma automatizada.

³Tratamento aplicado a arquivos escaneados ou com elementos gráficos de fundo (como papel timbrado ou imagens), nos quais o texto original é convertido por meio de OCR (Reconhecimento Óptico de Caracteres) para viabilizar sua extração.

⁴Após a conversão para .txt, todos os arquivos foram verificados individualmente, de forma manual, para auditoria da integridade da extração textual. Os reparos indicados referem-se aos arquivos que demandaram intervenção para correção de erros de conversão.

Fonte: Dados da Pesquisa, 2025.

Depois de convertidos e limpos, os textos foram padronizados com alguns ajustes essenciais. Substituíram-se abreviações técnicas por marcadores neutros e foram retiradas palavras muito frequentes e pouco informativas, como “de”, “para”, “o”, “a”, a fim de reduzir interferências nas análises textuais automatizadas. Em seguida, foi aplicado um algoritmo de *stemming* (radicalização), por exemplo, “atingir” →

“ating”, “eficiência” → “efici”, “adulteração” → “adulter”, com o objetivo de padronizar flexões verbais e plurais. Essa etapa foi replicada tanto nos textos processados quanto nos dicionários de apoio, assegurando uniformidade na redução dos termos aos seus radicais

As variáveis contábeis e financeiras, utilizadas como controle foram coletadas nas bases Economatica e Refinitiv Eikon, garantindo a completude e precisão das informações. A escolha por utilizar as referidas bases se justifica pela necessidade de obter dados complementares que não são completamente cobertos por uma única fonte. Essas bases fornecem variáveis críticas para a realização das análises propostas:

- **Refinitiv Eikon:** Oferece dados relacionados à governança corporativa por meio de indicadores como *Governance Pillar Score*, *ESG Score*, *ESG Controversies Score*, *Board Size* (tamanho do conselho), *CEO Board Member* (CEO é também membro do conselho), *CEO Duality* (acúmulo da função de CEO e presidente do conselho) e *Female Board* (percentual de mulheres no conselho de administração). Também disponibiliza indicadores de mercado, como *Bid Price*, *Ask Price*, *Number of Employees* e *Number of Employees from CSR Reporting* (número de funcionários).
- **Economatica:** Fornece dados contábeis, financeiros e de mercado, sendo especialmente útil para acessar séries históricas padronizadas e calcular variáveis de controle. A partir dessa base, foram extraídas variáveis como:
 - **Dados cadastrais e estruturais:** Data do primeiro balanço, setor de atuação, segmento de listagem na B3, *status* de negociação (ativo/cancelado) e tipo de bolsa fonte.
 - **Indicadores financeiros e contábeis:** Ativo total, passivo circulante, passivo não circulante, dívida bruta total, patrimônio líquido, lucro líquido, receita, caixa e equivalentes de caixa, caixa operacional e ativo imobilizado.
 - **Variáveis de controle econômico-financeiro:** Tamanho (log do ativo total), alavancagem (dívida/ativos), retorno, volume financeiro negociado (Volume\$), dividendos pagos, valor de mercado, quantidade de ações e preço unitário das ações.
 - **Indicadores de risco de mercado:** Volatilidade e índice de Sharpe.

Essas fontes, utilizadas em conjunto, garantem a cobertura completa das variáveis necessárias para a análise deste estudo. A combinação das bases também permite cruzar e validar informações, assegurando maior confiabilidade nos dados.

As métricas de legibilidade, comprimento dos relatórios, tom, foram calculadas por meio de *scripts* automatizados no *software R*. Foram extraídas variáveis básicas como o número total de palavras, frases, sílabas, letras e palavras difíceis (com três ou mais sílabas), prevalência de palavras financeiras, palavras comuns, número de palavras positivas, negativas frequência e termos associados à ambiguidade.

Todos os dados extraídos foram estruturados em painel empresa-ano e exportados em arquivo no formato “.xlsx”, compondo a base quantitativa utilizada nos modelos estatísticos subsequentes. Além disso, a base de dados configura um painel desbalanceado, uma vez que as empresas analisadas não estão presentes em todos os anos do período, seja por terem sido constituídas após 2010, encerrado suas atividades antes de 2024 ou, ainda, por alterações no *status* de listagem. Este desbalanceamento é estrutural e natural em estudos longitudinais com empresas, refletindo a dinâmica temporal do mercado de capitais.

Com relação às variáveis utilizadas nos modelos estatísticos, especificamente na estimação do PSM e DiD, identificaram-se dados ausentes pontuais, decorrentes dessa mesma dinâmica temporal. Esse padrão de ausência não decorre de falhas na coleta ou processamento dos dados, mas da própria variação na idade das empresas e no tempo de permanência na base de análise. Para mitigar perdas amostrais e assegurar a completude dos dados nas variáveis essenciais para o PSM e dos modelos econôméticos subsequentes, foi adotado um procedimento de imputação pela mediana.

3.3 Propensity score matching (PSM)

A decisão de listar ADRs no mercado norte-americano representa uma escolha estratégica por parte de empresas brasileiras, e por se tratar de uma decisão voluntária, é plausível supor que as empresas que optam por esse tipo de listagem apresentem características sistematicamente diferentes daquelas que permanecem exclusivamente no mercado doméstico, o que compromete a comparabilidade direta entre os grupos (Coffee, 2002; Doidge et al., 2004).

Para mitigar esse viés, este estudo adota um delineamento quase-experimental baseado no método PSM, que busca equilibrar os grupos com base em covariáveis mensuradas antes do tratamento. O escore de propensão foi estimado por meio de um modelo logístico, tendo como variável dependente binária a presença de ADRs, assumindo valor 1 para empresas que listaram ADRs em qualquer momento do período analisado, e 0 para aquelas que nunca realizaram esse tipo de listagem. Reconhece-se, contudo, que o PSM não corrige vieses decorrentes de seleção em características não observáveis, razão pela qual a robustez da identificação é complementada com o teste de paralelismo via *leads* no *event-study*, que atua como diagnóstico central da estratégia de efeito estimado.

As covariáveis utilizadas refletem atributos financeiros, econômicos e textuais, incluindo métricas derivadas da análise textual dos relatórios da administração, como legibilidade, comprimento, tom, frequência de palavras comuns e termos financeiros. A seleção dessas variáveis considerou sua significância empírica, a disponibilidade nas bases utilizadas e a viabilidade estatística frente ao delineamento amostral adotado.

O pareamento foi aplicado apenas para balancear covariáveis pré-tratamento e realizado com base no método do pareamento ótimo (*optimal matching*), que minimiza globalmente as distâncias entre os escores dos pares. A qualidade do pareamento foi avaliada por meio da diferença padronizada de médias (*standardized mean difference*) para cada covariável, considerando-se valores inferiores a 0,1 como indicativos de desequilíbrio negligenciável.

Adicionalmente, plotou-se a visualização gráfica das diferenças antes e depois do pareamento por meio do gráfico `love.plot()`, com o objetivo de observar a melhoria no balanceamento das covariáveis após a aplicação do PSM. A amostra pareada obtida por meio do PSM foi empregada na análise DiD subsequente.

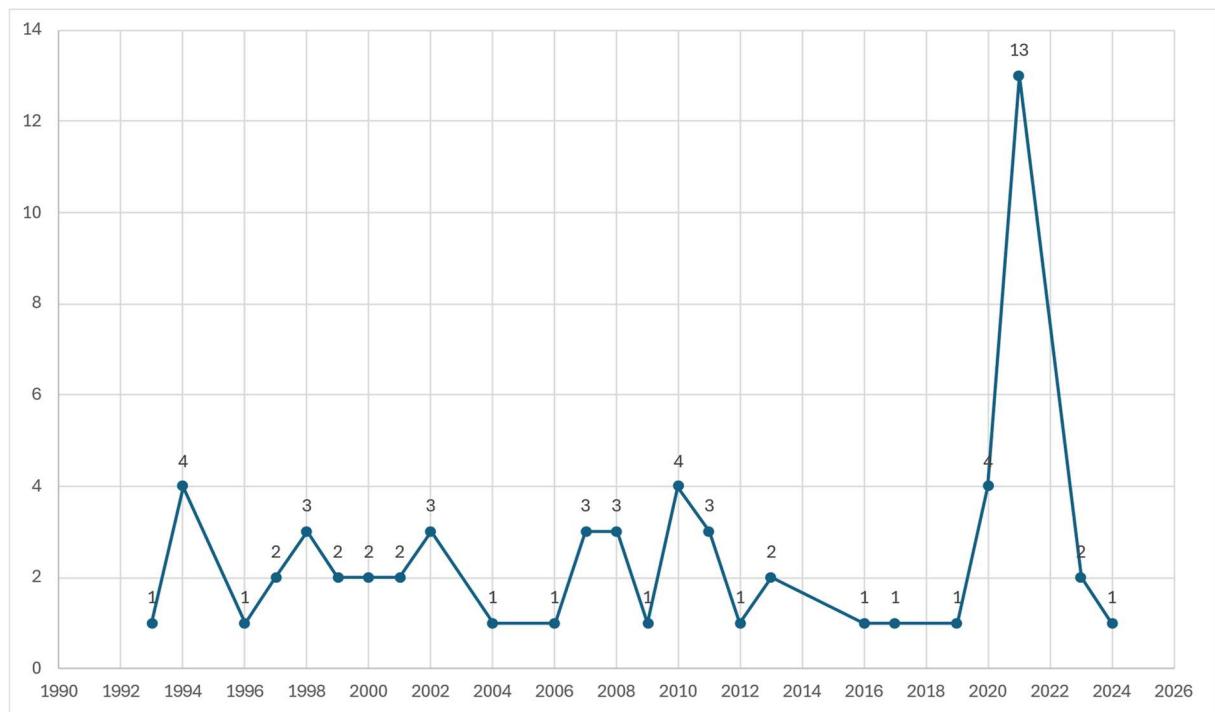
3.3.1 Delineamento da amostra

A amostra deste estudo é composta por empresas brasileiras não financeiras listadas na B3 entre 2010 e 2024, com informações estruturadas em formato painel desbalanceado. A construção do grupo de tratamento considerou todas as empresas que, ao longo desse período, emitiram ADRs no mercado norte-americano, conforme registros da base pública da J.P. Morgan.

A Figura 5 apresenta a distribuição anual da entrada de empresas brasileiras no mercado dos EUA via ADRs, entre 1993 e 2024. Observa-se que a entrada das empresas no mercado dos EUA ocorre de forma escalonada (*staggered adoption*), sendo que a primeira listagem ocorreu em 1993, conforme dados públicos da J.P. Morgan, e que há grande dispersão temporal entre as empresas tratadas. Essa heterogeneidade reflete decisões estratégicas individuais e, em alguns casos, um processo de preparação antecede que antecede a listagem oficial, o que abre a possibilidade de efeitos de antecipação, verificados nos testes de pré-tendências do *event-study*.

Figura 5

Empresas brasileiras com ADRs por ano de emissão (1993–2024).



Fonte: Dados da Pesquisa (2025).

Assim, considerando que o período analisado se refere a 2010 a 2024, foram excluídas da amostra as empresas cuja listagem ocorreu antes de 2010. Diante disto, após essa filtragem, a amostra resultante é composta por 32 empresas tratadas.

3.3.2 Escolha das covariáveis para o *propensity score matching* (PSM)

A seleção das covariáveis utilizadas no PSM considerou três critérios operacionais: (i) associação empírica com a variável de tratamento no período anterior

à intervenção; (ii) ausência de multicolinearidade; e (iii) viabilidade de inclusão conjunta no modelo, sem sobreposição com os desfechos analisados.

O primeiro procedimento consistiu na aplicação de testes de diferença de médias (*t-test*) entre os grupos com e sem ADRs, com base nos dados de 2010. Priorizou-se as variáveis que apresentaram diferenças estatisticamente significativas, por refletirem características com maior distinção entre os grupos.

Na etapa seguinte, a análise de correlação de Pearson permitiu identificar redundâncias estatísticas entre as variáveis remanescentes, especialmente entre proxies de tamanho (TAM, VM, Quantidade de Ações). Nessas situações, a seleção privilegiou a variável com menor sobreposição informacional.

A presença de multicolinearidade foi verificada por meio do cálculo dos Fatores de Inflação da Variância (VIF), considerando como critério valores abaixo de 5. As variáveis que atenderam a esse requisito foram incluídas no modelo de estimação do escore.

O escore de propensão foi estimado com base nas covariáveis observadas em 2010, primeiro ano do período de análise, a partir do escore estimado por meio de regressão logística (*logit*), com aplicação do método de pareamento ótimo (*optimal matching*), conforme implementado no pacote MatchIt do R. O critério adotado foi de pareamento um para um (1-to-1), sem reposição, em que cada empresa tratada foi pareada com uma empresa controle com escore semelhante. Não foi especificado *caliper*, uma vez que o algoritmo de otimização do método exclui automaticamente correspondências com distâncias excessivas.

3.4 Modelos econométricos da pesquisa

3.4.1 Modelo 1 – legibilidade e incerteza textual (H1)

Para testar a **Hipótese 1**, que propõe uma associação negativa entre a legibilidade dos relatórios da administração e a incerteza textual, foi especificado o seguinte modelo de regressão linear com efeitos fixos setoriais (Equação (1)):

$$Incerteza_{it} = \alpha + \beta_1 Legibilidade_{it} + \beta_2 ADR_i + X'_{it}\gamma + \delta_s + \varepsilon_{it} \quad (1)$$

Em que:

$Incerteza_{it}$ é o nível de incerteza textual do relatório da administração da

empresa i no tempo t ;

$Legibilidade_{it}$ é o índice de facilidade de leitura, valores mais altos indicam maior clareza textual;

ADR_i é a variável *dummy* que assume valor 1 para as empresas que listaram ADRs no mercado norte-americano durante o período analisado, e valor 0 para aquelas que nunca listaram;

$X'_{it}\gamma$ é o vetor de covariáveis de controle;

δ_s são os efeitos fixos por setor; e

ε é o termo de erro.

3.4.2 Modelo 2 –tom negativo e incerteza textual (H2)

A **Hipótese 2** propõe que a interação entre a legibilidade e o tom negativo dos relatórios da administração, contribui para a redução da incerteza textual. O modelo estimado é:

$$Incerteza_{it} = \alpha + \beta_1(Legibilidade_{it} \times TomNeg_{it}) + \beta_2ADR_i + X'_{it}\gamma + \delta_s + \varepsilon_{it} \quad (2)$$

Em que:

$TomNeg_{it}$ é a frequência relativa de termos negativos nos relatórios da empresa i no ano t .

As demais variáveis utilizadas neste modelo já foram descritas na subseção anterior (Modelo 1).

3.4.3 Modelo 3 – efeito estimado da governança via ADRs (H3 – PSM + DiD)

Este estudo adota uma estratégia econométrica combinada, integrando as técnicas de PSM e análise de eventos (*event study*) com efeitos fixos, com o objetivo de estimar a dinâmica da listagem de ADRs sobre a incerteza textual dos relatórios da administração. A emissão de ADRs por empresas brasileiras representa um evento institucional observável, frequentemente associada à adoção de padrões mais elevados de governança corporativa.

A abordagem de *event study* estende o modelo DiD tradicional ao substituir a interação única entre tratamento e período pós-evento por um conjunto de variáveis indicadoras para cada período relativo ao evento. Essa estrutura permite avaliar não

apenas o efeito médio do tratamento, mas também a sua evolução ao longo do tempo, assim como verificar a validade da hipótese de tendências paralelas por meio da inspeção dos coeficientes nos períodos pré-tratamento (*leads*).

Embora a abordagem DiD seja frequentemente aplicada em contextos de políticas públicas obrigatórias, a literatura reconhece sua validade em cenários de tratamentos voluntários, desde que três condições fundamentais sejam atendidas: (i) identificação clara do momento em que o tratamento ocorre; (ii) apenas parte da amostra seja afetada; e (iii) validade da hipótese de tendências paralelas nos períodos anteriores à intervenção (Angrist & Pischke, 2008). Por exemplo, Ferreira e Martins (2024), investigam como a adoção voluntária do relato integrado afeta a criação de valor para os acionistas de empresas de capital aberto, utilizando conjuntamente PSM e DiD.

Neste contexto, o grupo de tratamento é composto por empresas que passaram a negociar ADRs durante o período analisado, enquanto o grupo de controle inclui empresas que nunca adotaram essa prática. Foram estimados dois modelos separados: um utilizando como marco temporal o ano de anúncio da listagem, capturando potenciais efeitos antecipatórios; e outro utilizando o ano efetivo de negociação, capturando os efeitos pós-listagem. A variável de desfecho é a incerteza textual, mensurada por meio da frequência relativa de termos associados à ambiguidade e à imprecisão nos relatórios da administração.

Para testar a **Hipótese 3**, que investiga o efeito estimado da listagem de ADRs, proxy para maior governança, sobre a incerteza textual, utiliza-se um modelo DiD escalonado com pareamento prévio via PSM. A seguir, a especificação econométrica do modelo utilizado:

$$Incerteza_{it} = \alpha + \sum_{k \neq -1} \beta_k \times 1\{t - g_i = k\} + X'_{it}\gamma + \delta_s + \delta_t + \varepsilon_{it} \quad (3)$$

Em que:

$Incerteza_{it}$ é a medida de incerteza textual no relatório da administração da empresa i no ano t , calculada com base na frequência relativa de termos ambíguos e imprecisos;

g_i é o ano de tratamento (anúncio ou listagem efetiva) da empresa i ;

$1\{t - g_i = k\}$ é a variável *dummy* que assume valor 1 quando a observação está no período relativo k ao evento para empresas tratadas, e 0 caso contrário;

k varia de $-K_{lead}$ até K_{lag} , excluindo o período de referência (-1);

$X'_{it\gamma}$ é o vetor de variáveis de controle observadas por empresa e ano;

δ_s são efeitos fixos por setor de atividade econômica;

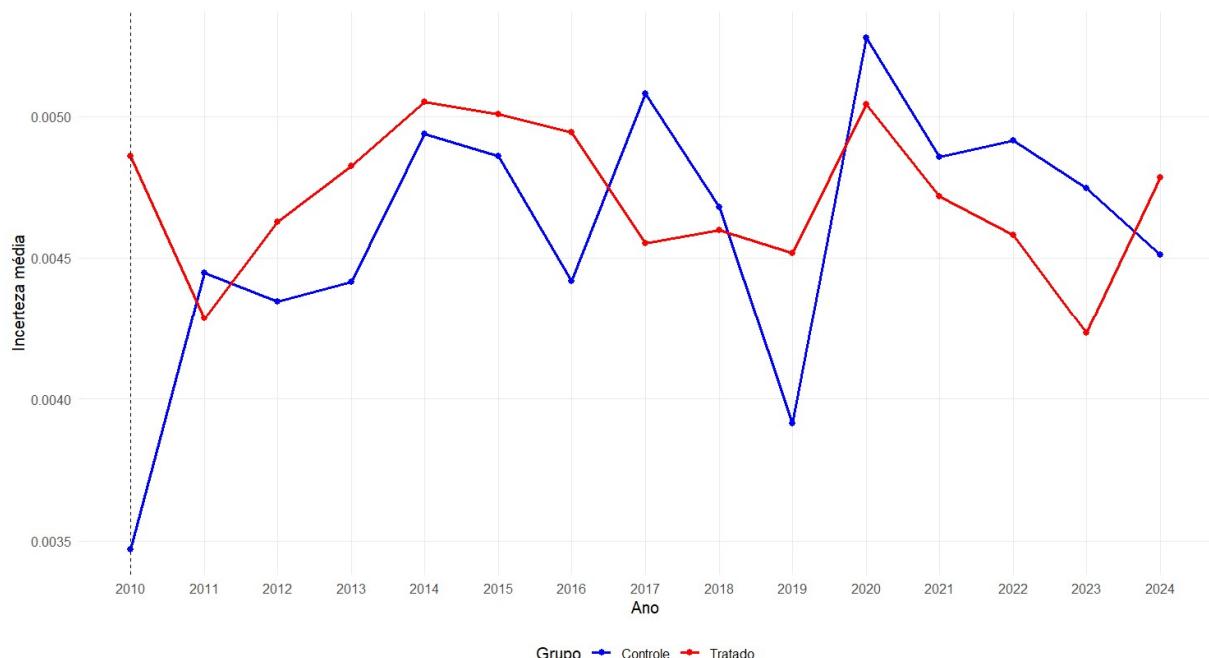
δ_t são efeitos fixos por ano; e

ε é o termo de erro.

Para garantir a robustez do modelo DiD, foi realizada uma verificação visual da suposição de tendências paralelas, a qual pressupõe que, na ausência do tratamento, os grupos tratado e controle teriam evoluído de forma semelhante ao longo do tempo. A **Figura 6** apresenta a média anual da incerteza textual por grupo. As trajetórias similares observadas no período anterior à listagem via ADRs reforçam a validade empírica do modelo e sustentam a aplicação do DiD na amostra analisada.

Figura 6

Evolução temporal da incerteza textual por grupo.



Fonte: Dados da Pesquisa (2025).

Para reforçar a validade do efeito estimado do estimador DiD, foi realizado um teste placebo (*placebo test*), atribuindo artificialmente a adoção do tratamento (listagem via ADRs) a períodos anteriores ao evento real, especificamente dois anos antes do efetivo ano de listagem. Os resultados da Tabela 3 mostraram significância estatística em apenas um ponto ($t = -12$), com coeficiente de 0,00152 ($p = 0,097$). Esse resultado pontual pode decorrer de flutuações idiosincráticas ou de choques específicos de baixa persistência, não caracterizando um padrão consistente de violação da premissa, dado que não há concentração de coeficientes significativos nem coerência na direção dos sinais nos períodos anteriores ao evento fictício. Assim, considera-se que o modelo DiD permanece válido sob a ótica desse teste de robustez.

Tabela 3

Resultado da Regressão DiD – Teste Placebo (datas falsas de adoção).

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
tempo_relativo:::-12	0,00152	0,00089	1,71140	0,09700 *
tempo_relativo:::-11	-0,00062	0,00074	-0,83389	0,41072
tempo_relativo:::-10	0,00096	0,00122	0,78123	0,44059
tempo_relativo:::-9	-0,00037	0,00066	-0,55212	0,58482
tempo_relativo:::-8	-0,00002	0,00079	-0,02391	0,98107
tempo_relativo:::-7	0,00015	0,00070	0,21169	0,83374
tempo_relativo:::-6	0,00028	0,00062	0,44293	0,66089
tempo_relativo:::-5	-0,00031	0,00058	-0,53161	0,59879
tempo_relativo:::-4	-0,00063	0,00056	-1,12656	0,26857
tempo_relativo:::-3	-0,00035	0,00042	-0,83842	0,40821
tempo_relativo:::-2	-0,00051	0,00034	-1,50011	0,14370
tempo_relativo:::0	-0,00021	0,00031	-0,68368	0,49926
tempo_relativo:::1	-0,00002	0,00044	-0,04508	0,96433
tempo_relativo:::2	0,00001	0,00032	0,04707	0,96276
tempo_relativo:::3	-0,00005	0,00033	-0,14903	0,88249
tempo_relativo:::4	-0,00012	0,00027	-0,45660	0,65114
tempo_relativo:::5	-0,00019	0,00033	-0,56423	0,57666
tempo_relativo:::6	-0,00057	0,00036	-1,56663	0,12735
tempo_relativo:::7	0,00075	0,00112	0,67168	0,50677
tempo_relativo:::8	-0,00049	0,00034	-1,46631	0,15263
tempo_relativo:::9	0,00066	0,00076	0,86219	0,39520
tempo_relativo:::10	0,00024	0,00039	0,60610	0,54886
tempo_relativo:::11	-0,00030	0,00038	-0,79465	0,43286
tempo_relativo:::12	0,00014	0,00031	0,45696	0,65088
tempo_relativo:::13	0,00009	0,00051	0,16604	0,86921
TomNeg	0,08479	0,02738	3,09730	0,00413 **
Length	0,00053	0,00014	3,84734	0,00056 ***
CommonWords	0,00107	0,00712	0,14971	0,88197
Legibilidade	-0,00001	0,00002	-0,69808	0,49033
FinancialWords	0,02233	0,01701	1,31280	0,19889
TAM	0,00000	0,00027	-0,00010	0,99992
TANGAT	0,00179	0,00102	1,74534	0,09083
VMP	-0,00002	0,00001	-2,06359	0,04751 *
EPU	-0,04529	0,05623	-0,80539	0,42673

R²

0,51417

Tabela 3*Resultado da Regressão DiD – Teste Placebo (datas falsas de adoção).*

R ² (efeito fixo)	0,22361
Observações	371

Legenda: *tempo_relativo* = indica a distância temporal em anos entre a observação e o momento da listagem via ADRs (t_0). Valores negativos representam períodos anteriores ao evento (*leads*), e valores positivos representam períodos posteriores (*lags*); *TomNeg* = proporção de palavras negativas no relatório da administração; *Length* = comprimento do relatório (log do número total de palavras); *CommonWords* = proporção de palavras comuns no texto; *Legibilidade* = grau de facilidade de leitura do texto; *FinancialWords* = proporção de termos financeiros no texto; *TAM* = tamanho da empresa (log do total de ativos); *TANGAT* = tangibilidade dos ativos (ativo imobilizado / total de ativos); *VMP* = razão entre valor de mercado e patrimônio líquido; *EPU* = índice de incerteza da política econômica.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

Além disso, para avaliar a robustez das estimativas do modelo DiD frente à composição amostral, foi conduzido um teste de estresse excluindo as empresas que listaram ADRs a partir de 2021. Essa escolha se justifica pelo fato de as coortes mais recentes representarem menor tempo de exposição pós-tratamento e, consequentemente, maior risco de distorção nas estimativas devido a efeitos de curto prazo ou choques específicos desses anos.

Os resultados da **Tabela 4** mostraram que os coeficientes para os períodos anteriores ao tratamento permaneceram estatisticamente insignificante, reforçando a plausibilidade da suposição de tendências paralelas. No período posterior ao evento (t_0), foram identificados alguns pontos com significância estatística a 10%, especificamente em $t = 6$, $t = 9$, $t = 12$ e $t = 13$ e um ponto com significância a 5% ($t = 4$). As magnitudes observadas foram pequenas (variando entre -0,0008 e -0,0013), sugerindo que se trata de flutuações pontuais, sem configuração de um padrão sistemático que indique viés estrutural nas estimativas.

Tabela 4*Resultado da Regressão DiD – Teste de Estresse – Exclusão de Coortes Sensíveis.*

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
tempo_relativo:::-10	-0,00030	0,00189	-0,15710	0,87713
tempo_relativo:::-9	0,00022	0,00123	0,17818	0,86082
tempo_relativo:::-8	-0,00135	0,00097	-1,39590	0,18181

Tabela 4

Resultado da Regressão DiD – Teste de Estresse – Exclusão de Coortes Sensíveis.

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
tempo_relativo:::-7	-0,00101	0,00094	-1,08072	0,29584
tempo_relativo:::-6	-0,00019	0,00096	-0,19715	0,84619
tempo_relativo:::-5	0,00062	0,00096	0,64674	0,52697
tempo_relativo:::-4	-0,00107	0,00101	-1,06520	0,30259
tempo_relativo:::-3	-0,00012	0,00097	-0,12259	0,90396
tempo_relativo:::-2	-0,00053	0,00050	-1,05914	0,30526
tempo_relativo:::0	-0,00055	0,00058	-0,96211	0,35031
tempo_relativo:::1	-0,00081	0,00064	-1,25485	0,22755
tempo_relativo:::2	-0,00082	0,00062	-1,31724	0,20631
tempo_relativo:::3	-0,00062	0,00071	-0,87924	0,39229
tempo_relativo:::4	-0,00101	0,00047	-2,16159	0,04615 *
tempo_relativo:::5	0,00015	0,00088	0,16645	0,86989
tempo_relativo:::6	-0,00135	0,00069	-1,97166	0,06619 *
tempo_relativo:::7	0,00000	0,00083	-0,00052	0,99959
tempo_relativo:::8	-0,00041	0,00049	-0,83855	0,41407
tempo_relativo:::9	-0,00095	0,00055	-1,74337	0,10045
tempo_relativo:::10	-0,00089	0,00058	-1,52706	0,14627
tempo_relativo:::11	-0,00094	0,00060	-1,57326	0,13522
tempo_relativo:::12	-0,00124	0,00059	-2,09369	0,05257 *
tempo_relativo:::13	-0,00083	0,00044	-1,86537	0,08057 *
TomNeg	0,07789	0,02505	3,10933	0,00675 ***
Length	0,00048	0,00030	1,57218	0,13547
CommonWords	-0,01498	0,01715	-0,87340	0,39536
Legibilidade	0,00000	0,00003	-0,07297	0,94274
FinancialWords	-0,01044	0,02119	-0,49296	0,62874
TAM	0,00049	0,00048	1,01753	0,32404
TANGAT	0,00107	0,00209	0,51246	0,61533
VMP	-0,00002	0,00001	-1,30620	0,20995
EPU	0,03248	0,04149	0,78293	0,44510
R ²		0,47160		
R ² (efeito fixo)		0,18576		
Observações		215		

Legenda: *tempo_relativo* = indica a distância temporal em anos entre a observação e o momento da listagem via ADRs (t_0). Valores negativos representam períodos anteriores ao evento (*leads*), e valores positivos representam períodos posteriores (*lags*); TomNeg = proporção de palavras negativas no relatório da administração; Length = comprimento do relatório (log do número total de palavras); CommonWords = proporção de palavras comuns no texto; Legibilidade = grau de facilidade de leitura do texto; FinancialWords = proporção de termos financeiros no texto; TAM = tamanho da empresa (log do total de ativos); TANGAT = tangibilidade dos ativos; VMP = valor de mercado sobre o patrimônio líquido; EPU = índice de incerteza da política econômica.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

Assim, mesmo com a exclusão das coortes mais recentes (empresas listadas a partir de 2021), as conclusões principais do estudo permanecem inalteradas. Esse

resultado reforça a robustez da estratégia de identificação, evidenciando que o efeito estimado não é sensível à presença de observações muito próximas ao final da série temporal analisa

Adicionalmente, com o objetivo de verificar a sensibilidade das estimativas ao período de observação, foi reestimado o modelo *event study* considerando uma janela temporal mais restrita, limitada a cinco anos antes e cinco anos depois do evento (t_0). Essa abordagem permite avaliar se os efeitos estimados no modelo principal são robustos quando a análise se concentra em um horizonte mais próximo ao evento de listagem via ADRs, reduzindo a influência de observações muito distantes no tempo.

Os resultados (**Tabela 5**) mostram que, no período pré-evento, nenhum coeficiente apresentou significância estatística, reforçando a suposição de tendências paralelas. Após o evento (t_0), também não se observam efeitos significativos da listagem sobre a incerteza textual dentro dessa janela restrita, indicando estabilidade das estimativas mesmo em horizonte mais curto. Esses achados indicam que a redução da janela de análise não altera as conclusões centrais do modelo, o que fortalece a robustez da estratégia de identificação.

Tabela 5

Resultado da Regressão DiD – Teste Placebo – Janela alternativa para t_0 .

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
tempo_relativo:::-5	0,00061	0,00064	0,94456	0,35243
tempo_relativo:::-4	0,00017	0,00048	0,35933	0,72187
tempo_relativo:::-3	0,00035	0,00050	0,70403	0,48684
tempo_relativo:::-2	0,00004	0,00032	0,12604	0,90054
tempo_relativo:::0	-0,00001	0,00041	-0,01848	0,98538
tempo_relativo:::1	-0,00025	0,00060	-0,42005	0,67744
tempo_relativo:::2	-0,00038	0,00067	-0,56808	0,57421
tempo_relativo:::3	-0,00040	0,00068	-0,58504	0,56290
tempo_relativo:::4	-0,00099	0,00071	-1,39137	0,17435
TomNeg	0,13596	0,04686	2,90150	0,00689 ***
Length	0,00052	0,00020	2,56912	0,01541 **
CommonWords	-0,00460	0,01068	-0,43051	0,66990
Legibilidade	-0,00001	0,00002	-0,26682	0,79143
FinancialWords	0,03573	0,03246	1,10051	0,27986
TAM	-0,00042	0,00035	-1,18806	0,24413
TANGAT	0,00064	0,00201	0,31959	0,75150
VMP	-0,00002	0,00001	-1,54583	0,13263
EPU	-0,09157	0,17099	-0,53553	0,59623
R ²			0,58656	
R ² (efeito fixo)			0,26112	

Tabela 5*Resultado da Regressão DiD – Teste Placebo – Janela alternativa para t_0 .*

Observações	230
Legenda: <i>tempo_relativo</i> = indica a distância temporal em anos entre a observação e o momento da listagem via ADRs (t_0). Valores negativos representam períodos anteriores ao evento (<i>leads</i>), e valores positivos representam períodos posteriores (<i>lags</i>); TomNeg = proporção de palavras negativas no relatório da administração; Length = comprimento do relatório (log do número total de palavras); CommonWords = proporção de palavras comuns no texto; Legibilidade = grau de facilidade de leitura do texto; FinancialWords = proporção de termos financeiros no texto; TAM = tamanho da empresa (log do total de ativos); TANGAT = tangibilidade dos ativos; VMP = valor de mercado sobre o patrimônio líquido; EPU = índice de incerteza da política econômica.	

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

Com base nas etapas descritas, incluindo o delineamento dos grupos, a definição temporal, a modelagem econométrica, a verificação da suposição de tendências paralelas e os testes placebo, consolida-se a estratégia empírica empregada neste estudo. Os modelos foram especificados com base em fundamentação teórica e empírica, e medidas de robustez, como o pareamento por PSM e os testes placebo, foram implementadas para validar as inferências. A **Tabela 6**, a seguir, sintetiza a articulação entre os objetivos do estudo, as hipóteses formuladas e os modelos econôméticos aplicados em cada etapa da investigação.

Tabela 6*Relação entre os objetivos, hipóteses e modelos por função.*

Objetivos	Hipóteses	Modelos por função
Analizar a associação entre a legibilidade dos relatórios da administração e a incerteza textual.	H ₁ : Existe uma associação negativa entre a legibilidade dos relatórios da administração e a incerteza textual.	Incerteza Textual = $f(\text{Legibilidade}, \text{ADR}, \text{Variáveis de controle})$
Analizar a associação entre o tom dos relatórios da administração e a incerteza textual.	H ₂ : Existe uma associação negativa entre a interação de legibilidade com o tom negativo dos relatórios da administração e a incerteza textual.	Incerteza Textual = $f(\text{TomNeg}, \text{ADR}, \text{Variáveis de controle})$
Estimar o efeito da listagem de empresas brasileiras nos Estados Unidos via ADRs sobre a incerteza textual.	H ₃ : Empresas brasileiras listadas via ADRs, ao adotarem padrões mais rigorosos de governança corporativa, apresentam maior legibilidade e uso mais disciplinado do tom em seus relatórios da administração, o que reduz a incerteza textual.	$\text{Incerteza}_{it} = \alpha + \beta_1 \text{ADR}_i + \beta_2 \text{Post}_t + \beta_3 (\text{ADR}_i \times \text{Post}_t) + X'_{it}\gamma + \delta_s + \varepsilon_{it}$

Fonte: Elaboração própria, 2025.

A **Tabela 7** apresenta de forma sintética as variáveis de controle utilizadas no PSM e aquelas incluídas no modelo DiD. A seleção das variáveis em cada etapa considerou os objetivos distintos de balanceamento entre grupos no PSM e de controle por heterogeneidade observável no DiD, com base em critérios de disponibilidade, relevância empírica e ausência de multicolinearidade.

Tabela 7

Variáveis incluídas no PSM e no modelo DiD.

Etapa Metodológica	Variáveis Incluídas
PSM	TAM (Tamanho), Idade, TANGAT (Tangibilidade)
DiD	TAM (Tamanho), Idade, TANGAT (Tangibilidade), VMP (Valor de Mercado sobre Patrimônio), EPU (Incerteza da Política Econômica), ICC (Índice de Confiança do Consumidor) e Crise (<i>dummy</i>)

Fonte: Elaboração Própria (2025).

Esse delineamento assegura as condições necessárias para a identificação do efeito estimado da listagem via ADRs sobre a incerteza textual, permitindo comparações válidas entre os grupos tratados e de controle.

3.5 Variável dependente

A variável dependente deste estudo é a incerteza textual expressa nos relatórios da administração das empresas brasileiras não financeiras listadas na B3. Sua mensuração segue a metodologia de Baker et al. (2016), originalmente proposta para o índice de incerteza econômica (*Economic Policy Uncertainty* - EPU), adaptada ao contexto de relatórios corporativos.

Para isso, utilizou-se o dicionário de Loughran e McDonald (2011, 2014), específico ao vocabulário contábil e financeiro, traduzido e ajustado para o português brasileiro. O índice foi calculado pela razão entre o número de palavras classificadas como relacionadas à incerteza, como “anomalia”, “contingência”, “cuidado”, “flutuante”, “indeciso”, “incerto”, “risco” e “volatilidade”, e o total de palavras do respectivo relatório da administração, conforme a Equação 4 abaixo:

$$Incerteza_{it} = \frac{Nº\ de\ palavras\ de\ incerteza\ (L\&M)}{Total\ de\ palavras\ do\ relatório} \quad (4)$$

Destaca-se que o índice foi calculado individualmente para cada relatório da administração, por empresa e por ano, permitindo captar variações intertemporais e entre diferentes empresas na opacidade textual. A métrica resultante reflete o grau de ambiguidade percebido na comunicação corporativa, sendo utilizada como *proxy* da incerteza informacional.

O processamento dos textos e o cálculo da variável foram realizados de forma automatizada, por meio de *scripts* desenvolvidos no *software R*, garantindo a replicabilidade dos resultados. O índice final de incerteza textual constitui a principal variável de resultado (*outcome*) da pesquisa e é utilizada exclusivamente na etapa de estimação do efeito estimado, por meio do modelo de diferenças em diferenças (DiD).

O quadro 1 a seguir resume a definição da variável dependente, sua fórmula de cálculo, fonte de referência e justificativa metodológica.

Quadro 1

Variável Dependente da Pesquisa.

Variável	Fórmula	Fonte/Referência	Justificativa para Uso
Incerteza Textual	Número de palavras de incerteza / Total de palavras do relatório	Adaptado de Baker et al. (2016), com vocabulário de Loughran e McDonald (2011, 2014)	Mede o grau de ambiguidade e imprecisão percebido no discurso corporativo. O dicionário de L&M foi escolhido por sua especificidade à linguagem financeira e maior acurácia na análise de textos corporativos, em comparação ao vocabulário político do EPU tradicional.

Fonte: Elaboração própria, 2025.

3.6 Variável de tratamento

A variável de tratamento considerada neste estudo corresponde à emissão de programas de ADRs por empresas brasileiras de capital aberto. Trata-se de uma variável binária que assume valor igual a 1 para os anos a partir da primeira emissão de ADRs pela firma, e 0 nos demais casos. A identificação das empresas listadas foi realizada com base nas informações da base pública disponibilizada pela J.P. Morgan.

Assim, a variável de tratamento funciona como uma *proxy* institucional para o acesso a padrões mais exigentes de governança e regulação internacional, sendo utilizada em ambas as etapas metodológicas do estudo. No PSM, a variável define o grupo tratado (empresas com ADRs) e o grupo de controle (empresas sem ADRs). Já no modelo DiD, ela compõe a interação com o tempo (pós-tratamento), permitindo estimar o efeito estimado da listagem sobre a incerteza textual. A definição precisa

dessa variável é, portanto, fundamental para a identificação do tratamento e a validade das inferências obtidas.

3.7 Variáveis explicativas e de controle

As variáveis a seguir são utilizadas com dois propósitos distintos: i) no PSM, são empregadas como covariáveis para formação de grupos comparáveis com base em características observáveis, e ii) no modelo DiD, são incluídas como variáveis de controle, a fim de aumentar a robustez das estimativas e reduzir o viés atribuído a fatores não observáveis.

Destaca-se que, as variáveis independentes (explicativas) são as *proxies* de legibilidade e tom dos relatórios da administração, detalhadas a seguir.

3.7.1 Legibilidade

3.7.1.1 Índice flesch adaptado para o português.

A mensuração da legibilidade dos relatórios da administração foi realizada a partir das Demonstrações Financeiras Padronizadas (DFPs), convertidas de “.pdf” para “.txt” com auxílio de *script* no software R. Arquivos com falhas na extração textual, decorrentes de problemas de codificação ou operadores incompatíveis, foram tratados manualmente no Adobe Acrobat Pro. Após essa etapa, os textos foram processados por meio de procedimentos automatizados, garantindo replicabilidade e precisão na análise.

A métrica principal utilizada foi o Índice Flesch adaptado ao português, conforme proposto por Martins et al. (1996). O índice varia de 0 a 100, sendo que valores mais altos indicam maior facilidade de leitura. A Equação (5) utilizada para o cálculo é apresentada a seguir:

$$\text{Índice Flesch Adaptado} = 248,835 - 1,015 \left(\frac{\text{Total de Palavras}}{\text{Total de Frases}} \right) - 84,6 \left(\frac{\text{Total de Sílabas}}{\text{Total de Palavras}} \right) \quad (5)$$

Em que:

Índice Flesch Adaptado: é a dificuldade de dificuldade textual, em português;

Total de Palavras é o número total de palavras no texto;

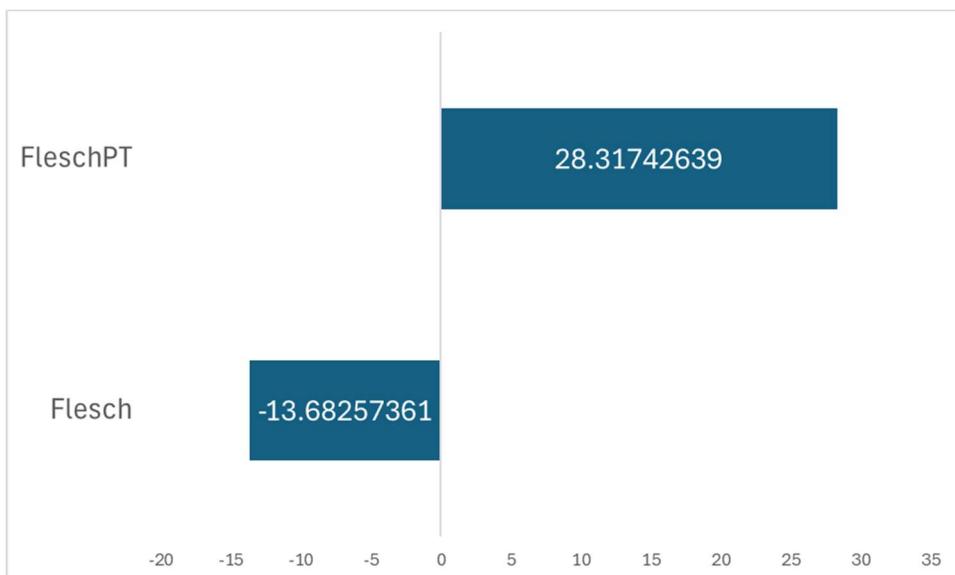
Total de Frases é o número total de frases no texto; e

Total de Sílabas: é o número total de sílabas no texto.

Além do Índice Flesch Adaptado, foram testados outros índices tradicionalmente aplicados à língua inglesa (*Flesch*, *Flesch-Kincaid*, *Gunning Fog*, *SMOG*, *ARI*, *Coleman-Liau* e *Average Sentence Length*). Esses índices classificaram os relatórios como mais difíceis quando aplicados a textos em português, reproduzindo o padrão identificado por Martins et al. (1996).

A Figura 7 compara diretamente o valor médio do índice Flesch original com o índice Flesch adaptado ao português. O Flesch original apresenta valor médio negativo (-13,7), classificando os relatórios como extremamente difíceis, enquanto o FleschPT (Legibilidade) apresenta valor médio de 28,3, enquadrando os mesmos textos na categoria de “texto difícil”.

Figura 7
Comparativo dos índices *Flesch Original* e *Adaptado*.



Fonte: Dados da Pesquisa (2025).

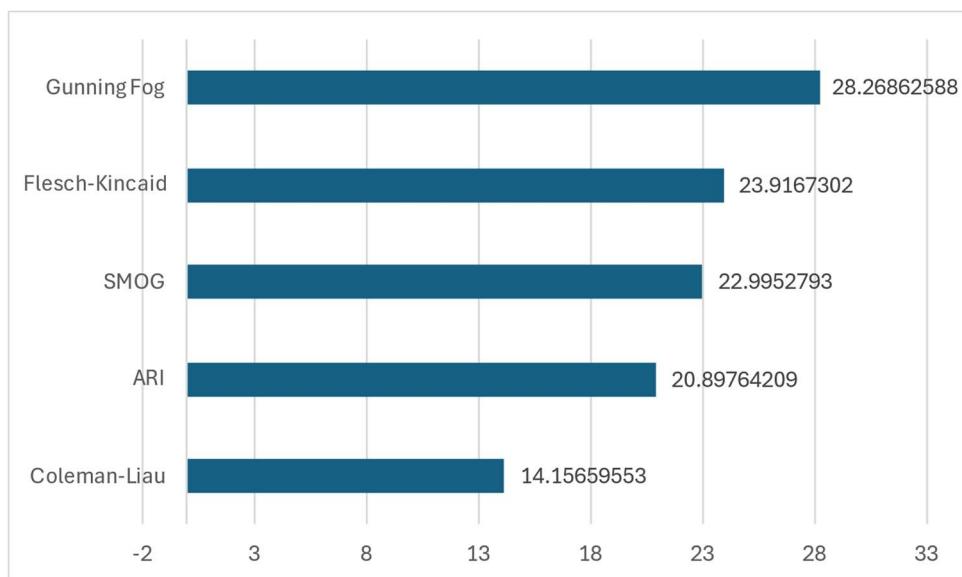
A Figura 8 apresenta os valores médios dos principais índices de legibilidade baseados em escolaridade, tradicionalmente aplicados à língua inglesa, calculados com base nos relatórios da administração entre 2010 e 2024. Os resultados apontam para médias elevadas: i) Gunning Fog, com média de 28,27, sugere que seria necessário nível de doutorado ou superior para compreensão plena; ii) Flesch-Kincaid Grade Level, com média de 23,92, aponta para mais de 23 anos de escolaridade formal, ultrapassando o nível de pós-graduação; iii) SMOG, com média de 22,99, também indica necessidade de formação universitária avançada; iv) ARI, com valor médio de 20,90, reforça a exigência de alto grau de escolaridade, compatível com

nível acadêmico superior; v) Coleman–Liau, com média de 14,16, sugere que a compreensão dos textos exigiria, 14 anos de escolaridade formal, o que equivale, aproximadamente, ao nível de ensino superior incompleto no Brasil.

Esses índices foram originalmente concebidos para estimar a quantidade de anos de escolaridade necessários à leitura fluente de textos em inglês. De acordo com essa lógica, valores entre 20 e 30 estariam associados à necessidade de formação universitária avançada. No entanto, tais interpretações devem ser consideradas com cautela no contexto brasileiro, especialmente quando se trata de textos institucionais formais, como relatórios da administração, redigidos em português.

Figura 8

Legibilidade média – escolaridade.

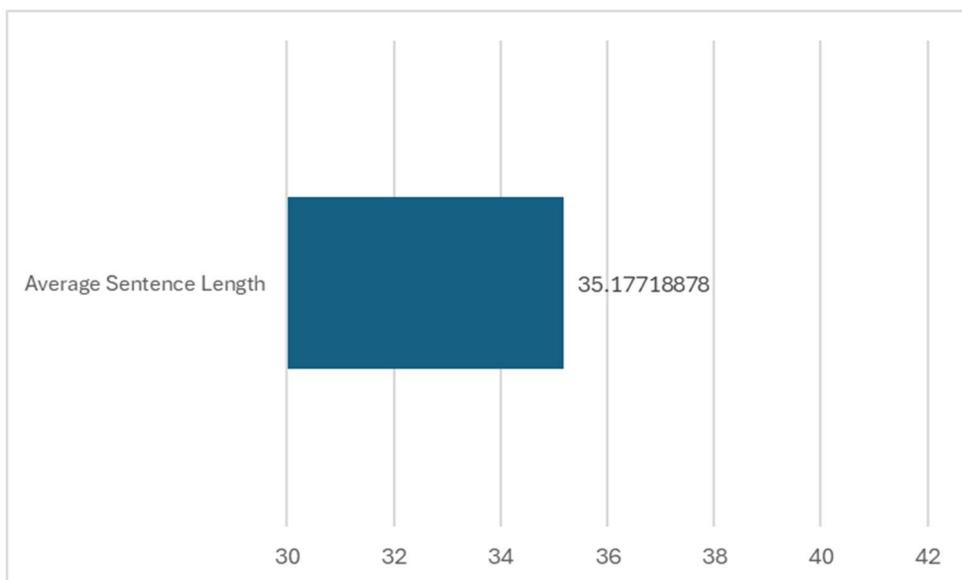


Fonte: Dados da Pesquisa (2025).

A **Figura 9** apresenta os índices de legibilidade baseados em estrutura textual, com destaque para o comprimento médio das sentenças (ASL – *Average Sentence Length*). O ASL atingiu uma média de 35,18 palavras por sentença, sugerindo uma estrutura sintática densa, típica de textos técnicos, com frases longas e subordinadas, o que dificulta a fluidez da leitura.

Figura 9

Legibilidade média – estrutura textual.



Fonte: Dados da Pesquisa (2025)

Apesar de ambos os índices terem limitações, o *Coleman-Liau* é possivelmente o mais neutro entre os índices originalmente desenvolvidos para a língua inglesa, por se basear em métricas objetivas como o número de letras por palavra e sentenças por parágrafo, e não depender da estrutura silábica ou do vocabulário da língua inglesa. Ainda assim, seu uso deve ser interpretado com cautela, já que os relatórios analisados são redigidos em português técnico-formal e direcionados a públicos especializados.

Esses resultados reforçam a necessidade de utilização de índices calibrados para o idioma e o contexto analisado, como o Flesch adaptado ao português. Para validação adicional, foi testado um relatório original em português, posteriormente traduzido para o inglês. Quando avaliado com os índices em inglês, o resultado da versão traduzida foi compatível com o índice Flesch adaptado, confirmando a adequação para o contexto da língua portuguesa.

Complementarmente, foram incorporadas ao modelo outras três métricas a análise da legibilidade, com base em estudos prévios da literatura contábil e financeira: (i) o comprimento do relatório (*Length*), (ii) a proporção de palavras comuns (*Common Words Index*) e (iii) a proporção de palavras financeiras (*Financial Words Index*).

3.7.1.2 Comprimento do documento (*length*).

A variável *Length* foi calculada como o logaritmo natural do número total de palavras do relatório. A métrica foi operacionalizada por meio de um *script* desenvolvido em R, que contabilizou o número total de palavras em cada relatório, a partir do texto processado. Após a remoção de cabeçalhos, rodapés, abreviações técnicas e elementos não textuais, cada documento foi *tokenizado* em palavras, e o total resultante foi submetido à função logarítmica natural (ln). O resultado, denominado *Length Index*, representa, portanto, a dimensão logarítmica do volume textual de cada relatório. A Equação 6, apresentada a seguir, foi utilizada para o cálculo:

$$\text{Length Index}_{it} = \ln(\text{Total de Palavras}_{it}) \quad (6)$$

Em que:

Length Index_{it} é o índice de comprimento do relatório da administração da empresa i no ano t ; e

$\text{Total de Palavras}_{it}$ é o número total de palavras no texto.

3.7.1.3 Proporção de palavras comuns (*common words index*).

O *Common Words Index* foi desenvolvido com base na frequência de termos recorrentes nos relatórios da administração. Para isso, foi construído um dicionário de palavras comuns a partir de um corpus textual contendo todos os documentos da amostra, processados no *software* R. Após limpeza linguística (minúsculas, remoção de pontuação, *stopwords* e aplicação de *stemming*), selecionaram-se os termos que apareciam em pelo menos 5% dos relatórios. A seguir, apresenta-se a Equação 7, utilizada no cálculo da variável:

$$\text{Common Words Index}_{it} = \frac{\text{PalavrasComuns}_{it}}{\text{TotalPalavras}_{it}} \quad (7)$$

Em que:

$\text{Common Words Index}_{it}$ representa a proporção de palavras comuns no relatório da empresa i no ano t ;

PalavrasComuns_{it} é o número de palavras encontradas no dicionário de termos recorrentes; e

Total de Palavras_{it} é o número total de palavras no relatório da administração.

3.7.1.4 Proporção de palavras financeiras (*financial words index*).

Para operacionalizar essa métrica, utilizou-se um dicionário de termos financeiros, inspirado no glossário elaborado por Campbell R. Harvey. O dicionário foi previamente traduzido, adaptado e submetido às mesmas etapas de pré-processamento linguístico aplicadas ao corpus principal: conversão para minúsculas, remoção de pontuação e *stopwords*, além da aplicação de *stemming*.

A identificação dos termos presentes no dicionário foi realizada por meio de um script automatizado em R, que analisou cada relatório e contabilizou as palavras que correspondiam aos radicais do vocabulário financeiro. A métrica foi então calculada como a razão entre o número de palavras financeiras e o número total de palavras do relatório. A Equação 8, utilizada nesta etapa, é expressa da seguinte forma:

$$Financial\ Words\ Index_{it} = \frac{Número\ de\ Palavras\ Financeiras_{it}}{Total\ de\ Palavras\ no\ Relatório\ _{it}} \quad (8)$$

Em que:

Financial Words Index_{it} representa a proporção de palavras financeiras no relatório da empresa *i* no ano *t*;

Número de Palavras Financeiras_{it} é o número de palavras do relatório identificadas como pertencentes ao vocabulário financeiro previamente definido; e

Total de Palavras no Relatório_{it} é o número total de palavras no relatório da administração.

3.7.2 Tom Negativo

Neste estudo, o tom textual é mensurado com base na predominância relativa de termos positivos e negativos nos relatórios da administração, conforme a metodologia proposta por Henry (2008). Utilizou-se o dicionário de tonalidade desenvolvido por Loughran e McDonald (2011), traduzido para o idioma português, o

qual contempla vocabulário específico do universo financeiro. A métrica tradicional de tom (Tom) é calculada segundo a seguinte equação:

$$TOM_{it} = \frac{Nº\ palavras\ positivas_{it} - Nº\ palavras\ negativas_{it}}{Nº\ total\ das\ palavras_{it}} \quad (9)$$

Adicionalmente, com o objetivo de verificar se o uso de linguagem negativa contribui para reduzir a incerteza textual, foi realizada uma adaptação metodológica para isolar o componente negativo da tonalidade. A métrica, denominada TomNeg, expressa a proporção de palavras negativas em relação ao total de palavras do relatório, conforme descrito abaixo:

$$TomNeg_{it} = \frac{Nº\ palavras\ negativas_{it}}{Nº\ total\ das\ palavras_{it}} \quad (10)$$

Essa adaptação visa capturar com maior precisão o grau de negatividade explícita nos textos, permitindo testar se a frequência de termos negativos contribui para mitigar a opacidade informacional.

A operacionalização do cálculo foi realizada de forma automatizada por meio de um *script* desenvolvido em R, que processa os relatórios da administração das empresas não financeiras listadas na B3 entre 2010 e 2024. O processamento textual inclui as seguintes etapas: i) remoção de *stopwords*; ii) aplicação de *stemming*; iii) e identificação de termos positivos e negativos com base nos dicionários de Loughran e McDonald (2011), traduzido para o idioma português.

Em seguida, calculou-se o total de palavras por relatório, bem como as ocorrências de termos classificados como positivos e negativos, conforme os radicais identificados nos dicionários. A métrica Tom Negativo é então derivada e armazenada para posterior análise estatística.

3.7.3 Variáveis de controle

Com o objetivo de isolar o efeito estimado da listagem de ADRs sobre a incerteza textual, este estudo incorpora um conjunto de variáveis de controle que refletem características estruturais das firmas, aspectos da qualidade da divulgação e condições macroeconômicas. A inclusão dessas variáveis visa mitigar possíveis

vieses relacionados à heterogeneidade observável e conferir maior robustez aos modelos estimados.

Baseado na literatura, foram selecionadas as seguintes variáveis de controle no nível da firma: o tamanho das empresas (TAM); idade da empresa (Idade); retorno sobre ativos (ROA); alavancagem financeira (ALV); Valor de Mercado sobre Patrimônio Líquido (VMP); tangibilidade dos ativos (TANGAT); Incerteza da Política Econômica (EPU); Índice de Confiança do Consumidor (ICC), e Crise. A seguir, são apresentadas suas justificativas para a inclusão do modelo de análise, bem como as relações esperadas e, ao final desta seção, os principais aspectos são consolidados na **Tabela 8**.

O tamanho da firma (**TAM**) foi mensurado pelo logaritmo natural do ativo total. Empresas maiores tendem a divulgar mais informações e atrair maior atenção do mercado, o que pode reduzir a incerteza textual. Por outro lado, a maior complexidade organizacional pode tornar os relatórios mais extensos e difíceis de interpretar, aumentando a ambiguidade.

A idade da empresa (**Idade**), medida em anos desde sua primeira aparição na B3. Empresas mais maduras tendem a apresentar práticas de divulgação mais consolidadas, podendo reduzir a ambiguidade textual.

O retorno sobre ativos (**ROA**), calculado como a razão entre lucro líquido e ativo total, mensura a eficiência da firma em utilizar seus recursos para gerar lucros. Empresas mais lucrativas tendem a divulgar informações de forma mais clara e acessível, dado que não enfrentam os mesmos incentivos para obscurecer resultados negativos. Como resultado, espera-se que maiores níveis de rentabilidade estejam associados a menor incerteza textual nos relatórios da administração.

A alavancagem financeira (**ALV**), operacionalizada como a razão entre dívida total e ativos totais, representa o grau de dependência da firma em relação a capital de terceiros. Elevados níveis de alavancagem implicam maior exposição ao risco financeiro e ampliam os conflitos de agência, especialmente entre acionistas e credores. Sendo assim, espera-se que maior alavancagem financeira esteja associada a maior incerteza textual nos relatórios da administração.

Valor de Mercado sobre Patrimônio Líquido (**VMP**) proxy de crescimento e precificação de mercado. Empresas com maior VMP podem estar sujeitas a expectativas elevadas dos investidores, recorrendo a estratégias de comunicação

mais técnicas, densas ou defensivas, o que se reflete em níveis mais elevados de incerteza textual nos relatórios da administração.

A tangibilidade dos ativos (**TANGAT**), calculada como a razão entre ativos imobilizados e ativos totais, reflete o grau de tangibilidade dos ativos operacionais da empresa. Espera-se que maior tangibilidade esteja associada a níveis mais baixos de incerteza textual nos relatórios da administração.

O modelo inclui três variáveis de controle no nível macroeconômico: o índice de Incerteza da Política Econômica (**EPU**), calculado com base na metodologia de Baker et al. (2016)Baker et al. (2016) individualmente para cada relatório da administração; o Índice de Confiança do Consumidor (**ICC**), divulgado pela Fundação Getúlio Vargas (FGV); e uma variável *dummy* de **Crise**, que assume o valor 1 para os anos marcados por choques sistêmicos relevantes e 0 caso contrário.

A seleção das variáveis de controle no modelo DiD segue lógica distinta daquela aplicada à estimação do escore de propensão (PSM). No PSM, foram incluídas apenas as variáveis associadas à decisão de listar ADRs, isto é, à propensão ao tratamento. Por essa razão, algumas variáveis que não apresentaram significância estatística ou correlação robusta com a decisão de listagem foram excluídas do pareamento.

No modelo DiD, por outro lado, foram incorporadas variáveis adicionais associadas ao desfecho de interesse (incerteza textual), com o objetivo de aumentar a robustez das estimativas. Essa prática metodológica é recomendada na literatura de avaliação de efeitos causais (Austin, 2011Austin, 2011; Rosenbaum & Rubin, 1983), que distingue as funções das variáveis no pareamento e na estimação do efeito estimado.

Todos os modelos incluem, adicionalmente, efeitos fixos por setor, para controlar heterogeneidades estruturais entre indústrias, e erros-padrão robustos agrupados por ano, conforme recomendação de Bertrand et al. (2004), assegurando validade na inferência estatística. A **Tabela 8** apresenta uma síntese das variáveis de controle.

Tabela 8*Variáveis de Controle.*

Variável	Descrição	Sinal Esperado	Fonte
TAM	Tamanho da empresa dado pelo logaritmo do ativo total.	+/-	(Healy & Palepu, 2001; Li, 2008)
Idade	Anos desde a primeira aparição na B3.	-	(Bar-Hod et al., 2021; Carlini et al., 2020; González et al., 2019, 2021)
ROA	Rentabilidade sobre o ativo total mensurada pela divisão entre lucro líquido e ativo total.	-	(Healy & Palepu, 2001)
ALV	Grau de dependência da firma em relação a capital de terceiros obtido pela divisão entre dívida total e ativo total.	+	(Courtis, 2004; Dalwai et al., 2021)
VMP	Crescimento e precificação de mercado obtido pela divisão entre valor de mercado e patrimônio líquido.	+	(Barth et al., 2017; Healy & Palepu, 2001)
TANGAT	Grau de tangibilidade dos ativos operacionais da empresa medido pela divisão entre imobilizado e ativo total.	-	(Hodgdon & Hughes, 2016)
EPU	Incerteza Política Econômica.	+	(Baker et al., 2016)
ICC	Índice de Confiança do Consumidor, usado como proxy de percepção econômica do consumidor	-	Contribuição inédita do estudo.
<i>Crise</i>	<i>Dummy para anos de choque sistêmico</i>	+	(González et al., 2019)

Fonte: Dados da Pesquisa (2025).

3.7.4 Variáveis Governança corporativa: considerações metodológicas e restrições empíricas

A governança corporativa constitui o eixo teórico central desta tese, operacionalizada pela listagem de ADRs. Essa variável funciona como um mecanismo externo de disciplinamento e transparência, com potencial para influenciar práticas de divulgação e reduzir a incerteza textual. Complementarmente, foram consideradas variáveis internas de governança comumente adotadas como controles adicionais na literatura, obtidas a partir da base Refinitiv Eikon. Entre elas, destacam-se:

- **BoardSize** número de membros do conselho de administração. Sua inclusão permite controlar variações na capacidade de supervisão e monitoramento da gestão (Ginesti et al., 2018; Hassanein & Albitar, 2024; Hussainey et al., 2022; Ivashkovskaya & Adamu, 2021).
- **FemaleBoard** proporção de mulheres no conselho. Componente da qualidade da governança, com impacto sobre a legibilidade e a transparência das informações divulgadas. A diversidade de gênero no board é associada a maior independência, menor viés gerencial e menor

incidência de práticas oportunistas (Ginesti et al., 2018; Ivashkovskaya & Adamu, 2021; Rajabalizadeh, 2023).

- **Governance Pillar Score** escore ESG referente ao pilar “G”, que abrange aspectos como independência do conselho, composição, comitês de auditoria e estrutura de controle. Mensura a qualidade da governança interna e sua influência sobre a divulgação financeira (Hussainey et al., 2022; Ishak et al., 2018).
- **ESG Score** indicador composto de práticas ambientais, sociais e de governança, utilizado como proxy de institucionalização de políticas de sustentabilidade e transparência (Barth et al., 2017; Sayari & Marcum, 2018).

Apesar da relevância empírica e do alinhamento com a temática deste estudo, a inclusão dessas variáveis foi inviabilizada pela elevada proporção de valores ausentes (*missing values*), o que comprometeria a consistência estatística dos modelos econometríticos. Conforme demonstrado na **Tabela 9**, enquanto variáveis centrais como TAM, ROA, Legibilidade, *Length* e Tom contam com 1.676 observações válidas, as variáveis de governança apresentam, em média, 890 observações, o que representa uma redução superior a 40% no tamanho da amostra.

Tabela 9
Número de Observações por Variável.

Variável	Obs
TAM (Tamanho)	1.676
ROA (Retorno s/ Ativo)	1.676
TANGAT (Tangibilidade dos Ativos)	1.676
FCO (Fluxo de Caixa Operacional)	1.676
Legibilidade	1.676
Length	1.676
Tom	1.676
EPU	1.676
Quantidade de Ações	1.675
ALV (Alavancagem)	1.674
LIQCX (Liquidez de Caixa)	1.673
POLDIV (Política de Dividendos)	1.649
CRESCREC (Crescimento da Receita)	1.649
VMP (Valor de Mercado / PL)	1.526
VOLNEG (Volume de Negociações)	1.503
RET (Retorno)	1.427
BidAsk	1.361
VOLPRC (Volatilidade do Preço)	1.300
BoardSize (Tamanho do Conselho)	891
FemaleBoard (% de Mulheres no Conselho)	891
<i>Governance Pillar Score (G)</i>	890
<i>ESGScore</i>	890
<i>ESG Controversies Score</i>	890
Quantidade de Funcionários	560

Fonte: Dados da Pesquisa (2025).

Adicionalmente, a inclusão dessas variáveis no modelo DiD resultou em uma redução significativa no número de observações válidas, comprometendo o poder estatístico e a estabilidade das estimativas. Em função dessa limitação amostral, observou-se instabilidade nos coeficientes principais, inclusive com alterações no sinal e na significância estatística da variável de tratamento (ADRs). Os resultados indicam que tais distorções decorrem da perda de casos disponíveis, e não de um efeito explicativo das variáveis adicionadas.

Diante desse cenário, adotou-se uma especificação mais parcimoniosa e robusta, priorizando variáveis amplamente disponíveis e estatisticamente relevantes, como TAM, TANGAT, VMP, além dos índices textuais (legibilidade, tom, *length*, *common words* e *financial words*). Essa decisão visa preservar a integridade do pareamento amostral realizado por meio do PSM e assegurar a robustez das inferências causais no modelo final.

Essa escolha metodológica segue recomendações consolidadas na literatura sobre modelagem em painéis desbalanceados, que enfatizam a importância de adotar especificações parcimoniosas para mitigar perdas amostrais e evitar vieses associados à redução do tamanho efetivo da amostra (Hsiao, 2003; Rubin, 1976; Wooldridge, 2010). A inclusão de variáveis adicionais, como as relacionadas à governança, implicaria uma redução superior a 40% da amostra total, o que comprometeria a precisão das estimativas e enfraqueceria a robustez dos testes de hipótese. A **Tabela 10** a seguir consolida os principais fundamentos metodológicos que justificam a adoção de especificações parcimoniosas em modelos com painéis desbalanceados.

Tabela 10

Fundamentação Metodológica para Especificações Parcimoniosas em Painéis Desbalanceados.

Autor	Conceito/Aplicação
Wooldridge (2010)	Especificações excessivamente complexas podem reduzir a eficiência, provocar ajuste excessivo aos dados (<u>sobreajuste</u>) e perdas amostrais. Recomenda-se parcimônia.
Hsiao (2003)	Painéis desbalanceados exigem cautela. Inclusão excessiva de variáveis pode reduzir o tamanho efetivo da amostra e comprometer a robustez.
Rubin (1976)	Recomenda-se modelagem que preserve o tamanho da amostra e evite viés associado a dados faltantes.
	O excesso de variáveis leva à perda de precisão, aumento do erro padrão e maior instabilidade dos modelos.
	A exclusão de observações com dados ausentes pode gerar viés, especialmente quando o mecanismo de ausência não é completamente aleatório.
	Modelos que preservam a simplicidade são aceitáveis quando o custo da complexidade implica perda amostral e instabilidade das estimativas.

Fonte: Elaborada a partir de Hsiao (2003), Rubin (1976) e Wooldridge (2010) (2025).

3.8 Limitações da pesquisa

Tal como ocorre em pesquisas aplicadas à análise textual e aos estudos de governança corporativa, este trabalho apresenta limitações que devem ser reconhecidas. A primeira delas diz respeito à própria natureza dos dados. A base construída para este estudo configura um painel desbalanceado, reflexo da dinâmica empresarial, na qual empresas ingressam ou se desligam da bolsa ao longo do tempo. Esse desbalanceamento é estrutural e não resulta de falhas no processo de coleta, mas de características temporais legítimas da amostra. Ainda assim, essa condição pode impactar a estabilidade das estimativas, especialmente em modelos econôméticos sensíveis à variação no tamanho da amostra (Hsiao, 2003; Wooldridge, 2010).

Reconhece-se que o número de empresas tratadas é relativamente reduzido ($n = 32$), sendo que mais da metade dessas emissões de ADRs concentrou-se nos anos mais recentes (2021 a 2023). Essa assimetria temporal impõe restrições ao poder estatístico da análise e limita a capacidade de explorar tendências de longo prazo nos efeitos do tratamento. Como teste complementar, foram retiradas da amostra as 12 empresas que emitiram ADRs em 2021 (ano com maior concentração de listagens) e os resultados se mantiveram consistentes, com sinais esperados e significância estatística preservada, o que reforça a robustez empírica do modelo.

Outro aspecto relevante refere-se às limitações específicas derivadas da indisponibilidade de algumas variáveis de governança corporativa, como BoardSize, FemaleBoard, Governance Pillar Score e ESG Score, cuja alta proporção de dados ausentes inviabilizou sua inclusão nos modelos finais. Embora tais variáveis sejam reconhecidamente relevantes na literatura de governança e divulgação de informações, sua exclusão foi necessária para preservar a robustez estatística e evitar perdas amostrais superiores a 40%, conforme discutido na seção metodológica. Além disso, reconhece-se que o uso da listagem via ADRs como *proxy* de governança, embora amplamente adotado, não contempla dimensões internas da estrutura de governança, como composição do conselho ou presença de comitês.

Também se reconhecem limitações inerentes à própria análise textual automatizada. Apesar de os índices de legibilidade, tom e incerteza textual serem amplamente utilizados na literatura (Li, 2008; Loughran & McDonald, 2014; Saravanan et al., 2024), essas métricas não capturam dimensões subjetivas da interpretação dos

textos pelos leitores humanos. Elementos como ironia, contexto discursivo, ambiguidades semânticas ou estruturas narrativas complexas podem não ser completamente detectados pelos algoritmos, representando uma restrição típica dos estudos de processamento de linguagem natural (PLN). Adicionalmente, a ausência de validação perceptual junto a leitores qualificados, usuários reais da informação contábil, como investidores ou analistas, representa um limite importante à interpretação dos efeitos comunicacionais identificados.

No tocante à estratégia de identificação do efeito estimado, embora a combinação de PSM e DiD seja amplamente reconhecida na literatura por sua capacidade de mitigar viés de seleção observável e controlar heterogeneidade não observável invariável no tempo (Abadie, 2005; Heckman et al., 1997), tais métodos não são capazes de eliminar completamente vieses decorrentes de fatores não observáveis que variam ao longo do tempo e que simultaneamente afetam a decisão de emitir ADRs e a incerteza textual.

Do ponto de vista da validade externa, os resultados devem ser interpretados com cautela em relação à sua generalização para outros países. O foco no mercado brasileiro, caracterizado por menor *enforcement* regulatório e estruturas institucionais distintas, limita a extração direta dos achados para economias desenvolvidas. No entanto, os resultados podem ser informativos para mercados emergentes com características semelhantes.

Por fim, mudanças no ambiente macroeconômico, alterações regulatórias, eventos externos ou choques específicos ocorridos durante o período de 2010 a 2024 podem ter impactado tanto o comportamento das firmas quanto os padrões de divulgação textual, o que representa uma limitação inerente a qualquer estudo empírico de natureza longitudinal.

Apesar dessas limitações, o estudo foi conduzido com procedimentos metodológicos rigorosos, que incluem coleta sistemática, pareamento por PSM, controles econômicos robustos e testes placebo. A transparência na exposição das escolhas metodológicas reforça a confiabilidade das evidências empíricas obtidas.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

4.1 Descrição da amostra e aplicação do PSM

A base de dados empregada neste estudo está estruturada em painel desbalanceado, contemplando informações anuais de empresas não financeiras listadas na B3, no período de 2010 a 2024. Foram excluídas da amostra as empresas cuja listagem via ADRs ocorreu antes de 2011.

Após esse procedimento, a amostra é composta por 32 empresas tratadas (com ADRs) e 280 empresas controle (sem ADRs), totalizando 312 empresas. Em termos de observações no painel (empresa-ano), isso corresponde a 480 registros no grupo tratado e 4.209 no grupo controle. A **Tabela 11** apresenta a distribuição das empresas tratadas segundo o ano de adesão aos ADRs, considerando apenas aquelas cuja listagem ocorreu a partir de 2010, conforme os critérios adotados para composição da amostra:

Tabela 11

Distribuição das Empresas Tratadas por Ano de Anúncio e Adoção dos ADRs.

Ano de Anúncio (anopre)	Número de Empresas	Ano de Emissão (AnoDR)	Número de Empresas
2010	4	2010	4
2011	4	2011	3
2012	2	2012	1
2013	-	2013	2
2016	1	2016	1
2017	1	2017	1
2018	2	2018	-
2019	3	2019	1
2020	11	2020	4
2021	2	2021	12
2022	2	2022	-
2023	-	2023	2
2024	-	2024	1

Nota: Nos casos sem anúncio ou comunicação ao mercado, definiu o tempo zero pela data de início do DR, conforme dados públicos na J.P. Morgan e adotou-se uma definição prévia que antecipa esse marco em 12 meses, refletindo o ciclo anual de preparação e divulgação financeira.

Fonte: Dados da Pesquisa (2025).

Essa distribuição reforça a necessidade do uso do PSM, dado o alto grau de dispersão temporal entre os anos de listagem das empresas tratadas. Como a base está organizada em estrutura de painel empresa-ano, o número total de observações reflete a presença repetida das mesmas empresas ao longo do tempo. A **Tabela 12**, a seguir, apresenta esse detalhamento.

Tabela 12

Amostra: Empresas e Observações no Painel (2010–2024).

Situação	Quantidade de Empresas	Quantidade de Observações (empresa-ano)
Tratadas	32	480
Não Tratadas	280	4.209
Total	312	4.689

Fonte: Dados da Pesquisa (2025).

Diante das diferenças estruturais entre os grupos, aplicou-se o método de PSM. A escolha das covariáveis para estimação do escore de propensão seguiu critérios operacionais baseados na relevância empírica e ausência de colinearidade.

Os testes de diferença de médias (*t-test*) realizados para o ano base de 2010 indicaram desequilíbrios significativos entre os grupos nas variáveis *Legibilidade* ($p = 0,0423$), *Length* ($p = 0,0040$) e *TAM* ($p < 0,001$), evidenciando um viés de seleção inicial. Por outro lado, as variáveis *TomNeg*, *CommonWords*, *FinancialWords*, *Idade* e *TANGAT* não apresentaram diferenças estatísticas relevantes, mas foram mantidas no modelo de escore por serem consideradas relevantes teoricamente e na literatura.

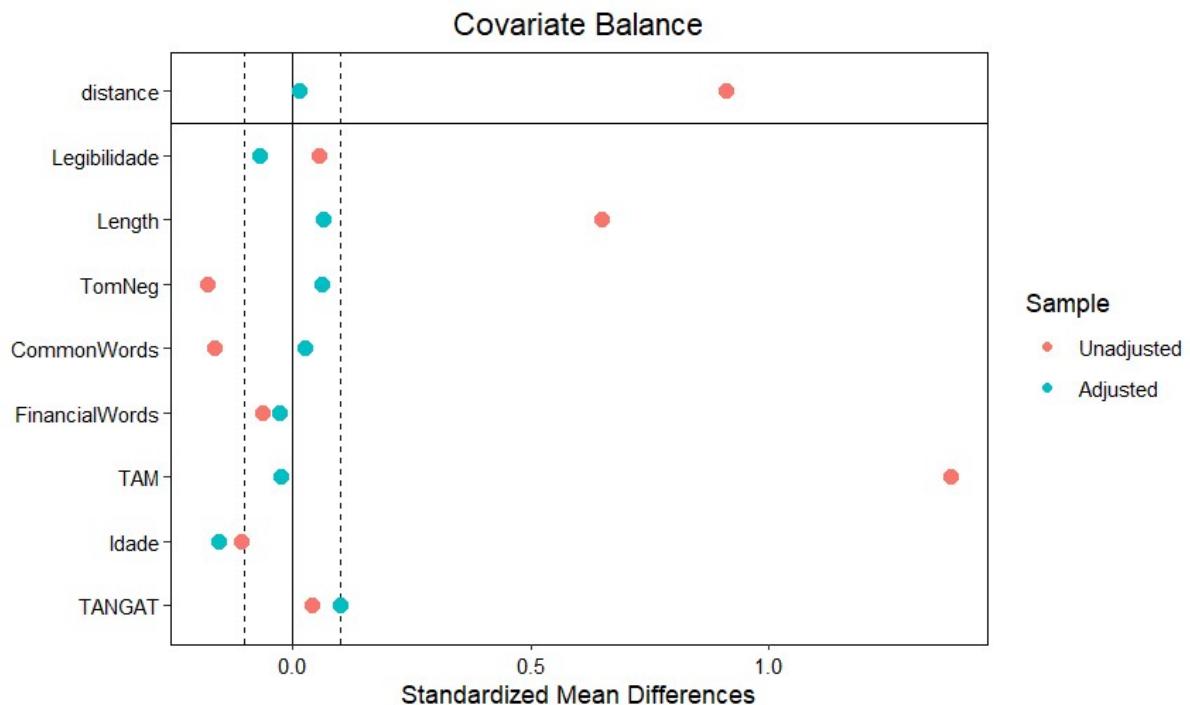
A matriz de correlação indicou uma associação forte entre *Legibilidade* e *CommonWords* (-0,78) e moderada entre *Length* e *TAM* (0,67), sem configurar colinearidade severa. Os valores do VIF confirmaram esse diagnóstico, com todos os indicadores abaixo do limiar crítico de 5, sendo o maior valor observado igual a 4,23 (*CommonWords*).

A avaliação da qualidade do pareamento baseou-se na Diferença Padronizada de Médias (*Standardized Mean Difference* - SMD), considerando valores inferiores a ($\pm 0,1$) como indicativos de bom balanceamento (Austin, 2011), e na Razão de Variância, idealmente situada entre 0,5 e 2,0. Os gráficos complementares confirmaram a existência de suporte comum, sem necessidade de exclusão de observações.

A Figura 10 apresenta o gráfico de diferenças padronizadas das médias (*Love Plot*), evidenciando que, após o pareamento, todas as covariáveis passaram a se situar dentro dos limites de balanceamento recomendados.

Figura 10

Balanceamento das covariáveis antes e depois do pareamento.

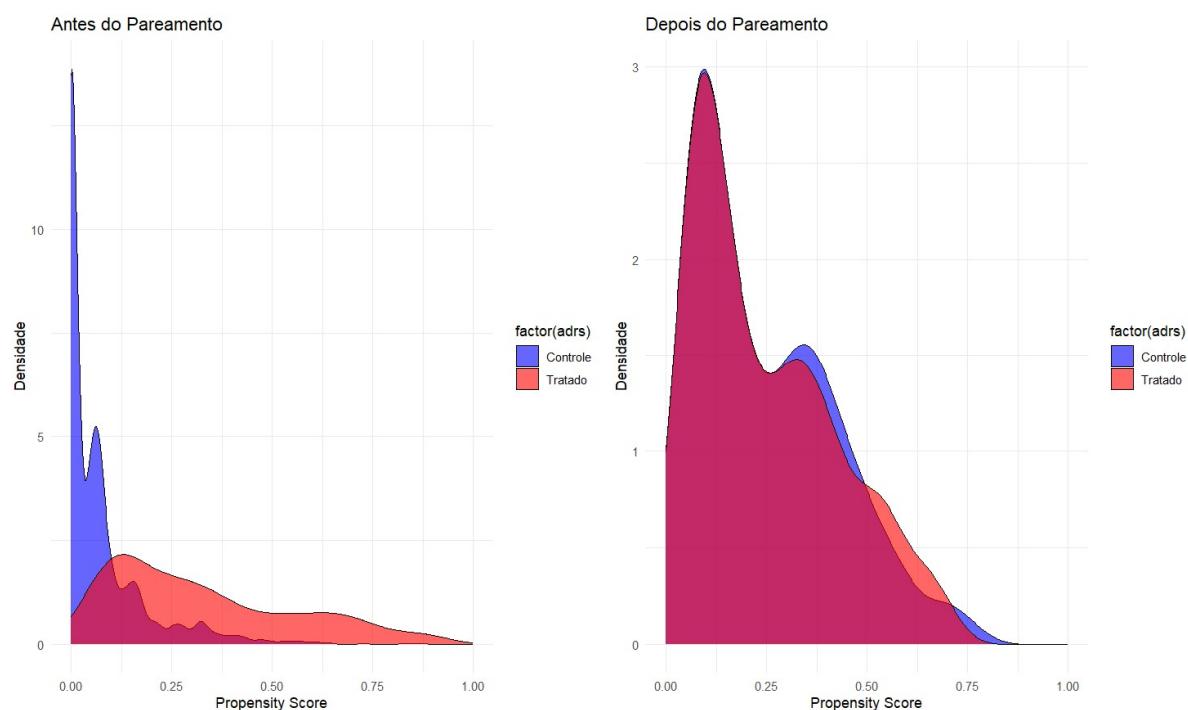


Nota: As diferenças padronizadas de médias foram calculadas para cada covariável do modelo de escore. Os pontos vermelhos indicam os valores antes do pareamento (*unadjusted*), e os azuis, após o pareamento (*adjusted*). As linhas tracejadas delimitam o limiar de balanceamento aceitável ($|0,1|$).
Fonte: Dados da Pesquisa (2025).

A Figura 11 mostra a distribuição dos escores de propensão nos grupos tratado e controle antes e depois do pareamento. Observa-se que, inicialmente, havia separação considerável entre os grupos, especialmente nas faixas superiores do escore. Após o pareamento, as curvas se tornam substancialmente sobrepostas, indicando maior comparabilidade entre as unidades analisadas.

Figura 11

Distribuição dos escores de propensão antes e depois do pareamento.



Nota: As curvas representam a concentração dos escores de propensão em cada grupo. Antes do pareamento (gráfico da esquerda), há clara separação entre os grupos. Após o pareamento (gráfico da direita), observa-se sobreposição substancial, indicando que o método conseguiu equilibrar as características observáveis entre empresas com e sem ADRs.

Fonte: Dados da Pesquisa (2025).

Após a aplicação do método de *optimal matching*, não foi necessário excluir observações por falta de sobreposição entre os grupos. Todas as empresas tratadas apresentaram escores compatíveis com empresas controle, o que reforça a comparabilidade no espaço das covariáveis. A **Tabela 13** apresenta os resultados do balanceamento antes do pareamento. Variáveis como *TAM*, *Length* e *Tom* apresentavam diferenças padronizadas superiores a 0,3, sendo o *TAM* o caso mais crítico (1,3801).

Tabela 13

Estatísticas de Balanceamento Antes do Pareamento (PSM).

Variável	Média Tratado	Média Controle	SMD	Razão de Variância
<i>Distance</i>	0,2478	0,0858	0,9093	3,955
Legibilidade	31,0038	30,3224	0,0565	1,306
<i>Length</i>	8,5660	7,9664	0,6508	0,8266
<i>TomNeg</i>	0,0174	0,0183	-0,1770	0,6815
<i>CommonWords</i>	0,3478	0,3537	-0,1612	1,2479
<i>FinancialWords</i>	0,0380	0,0384	-0,0613	0,7384

Tabela 13*Estatísticas de Balanceamento Antes do Pareamento (PSM).*

Variável	Média Tratado	Média Controle	SMD	Razão de Variância
TAM	16,0482	14,0605	1,3801	0,4788
Idade	18,6875	19,5647	-0,1051	0,5288
TANGAT	0,2190	0,2123	0,0406	0,6474

Nota:

- SMD: Diferença Padronizada de Médias. Valores abaixo de |0,10| indicam ótimo balanceamento.
- Razão de Variância ideal: entre 0,5 e 2,0.

Fonte: Dados da Pesquisa (2025).

A **Tabela 14** mostra os valores após o pareamento. As diferenças padronizadas foram reduzidas a níveis aceitáveis, com todas as covariáveis abaixo de 0,1, o que confirma o bom desempenho do PSM.

Tabela 14*Estatísticas de Balanceamento Após do Pareamento (PSM).*

Variável	Média Tratado	Média Controle	SMD	Razão de Variância
Distance	0,2478	0,2449	0,0162	1,0358
Legibilidade	31,0038	31,8285	-0,0684	0,9636
Length	8,5660	8,5066	0,0645	1,0462
TomNeg	0,0174	0,0171	0,0616	1,1561
CommonWords	0,3478	0,3469	0,0264	1,2747
FinancialWords	0,0380	0,0382	-0,0252	0,7537
TAM	16,0482	16,0798	-0,0219	1,0302
Idade	18,6875	19,9625	-0,1528	0,6954
TANGAT	0,2190	0,2022	0,1011	0,6291

Nota:

- SMD: Diferença Padronizada de Médias. Valores abaixo de |0,10| indicam ótimo balanceamento.
- Razão de Variância ideal: entre 0,5 e 2,0.

Fonte: Dados da Pesquisa (2025).

Por fim, observa-se que as empresas tratadas, em média, apresentam relatórios mais longos (*Length*), ligeiramente mais legíveis (*Legibilidade*) e com *Tom* marginalmente mais positivo. Por sua vez, a variável *Incerteza* apresentou média inferior no grupo tratado já no estágio descritivo, sugerindo menor ambiguidade nos relatos das empresas com ADRs. Esse resultado preliminar é compatível com a expectativa de que a listagem via ADRs, associada à adoção de padrões mais elevados de governança, se relacione com maior qualidade informacional nos relatórios.

4.2 Análise das hipóteses

4.2.1 Relação entre legibilidade e incerteza (H1)

Antes da análise econométrica, realizou-se uma comparação descritiva entre os grupos de empresas listadas via ADRs (grupo tratado) e aquelas restritas ao mercado doméstico (grupo controle). Os resultados indicam níveis significativamente superior de legibilidade nos relatórios das empresas listadas via ADRs (35,15 vs. 28,26; $p = 0,042$), o que é coerente com a hipótese de que a clareza textual contribui para reduzir a incerteza informacional percebida.

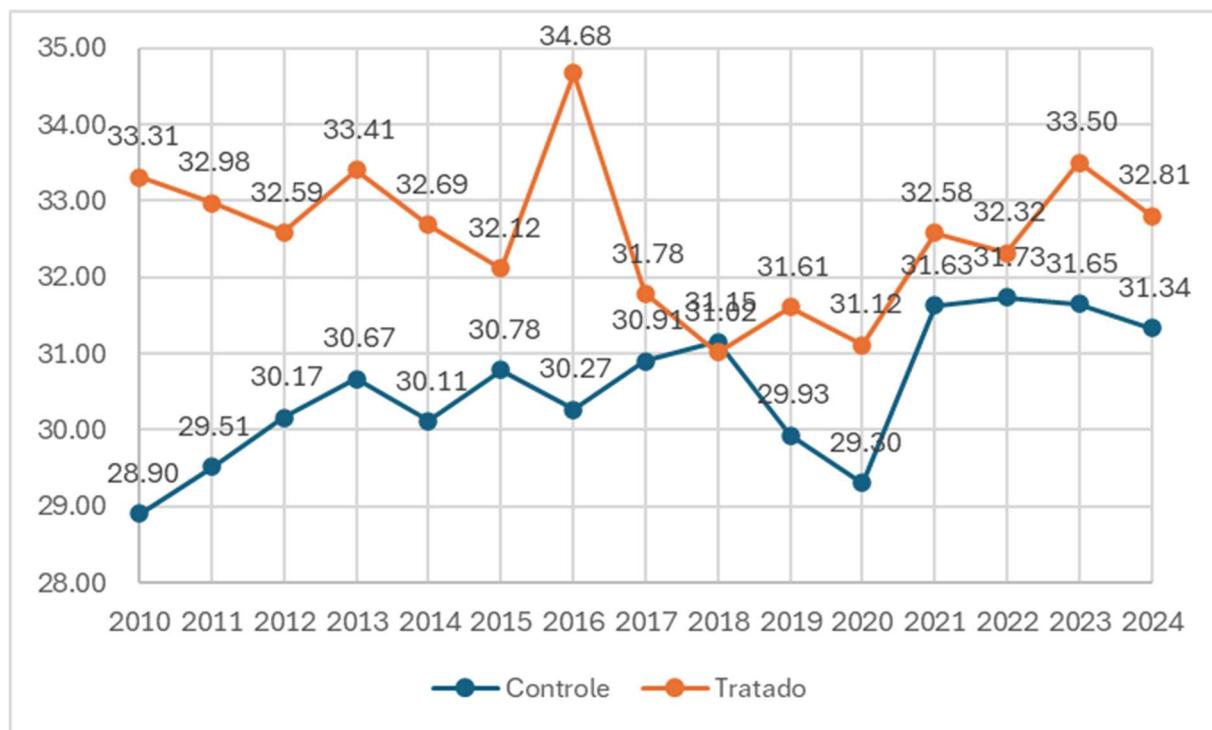
A maior extensão média dos relatórios no grupo tratado (8,57 vs. 7,76; $p = 0,004$) também reforça essa leitura, indicando maior detalhamento informativo. A simultaneidade entre relatórios mais longos e mais legíveis indica um esforço deliberado por parte dessas empresas em fornecer comunicações mais densas, mas também mais acessíveis, coerente com pressões institucionais do mercado norte-americano e de analistas globais.

Embora variáveis como proporção de palavras comuns (0,347 vs. 0,360; $p = 0,135$) e termos financeiros (0,0386 vs. 0,0374; $p = 0,590$) não tenham apresentado diferenças estatisticamente significativas, os resultados apontam um padrão de comunicação textual mais estruturado e tecnicamente orientado entre o grupo tratado.

A Figura 12 ilustra a evolução anual da legibilidade média dos relatórios da administração entre 2010 e 2024, confirmando a persistência da diferença entre os grupos. Em 2010, a diferença era de 4,4 pontos na Legibilidade (33,31 vs. 28,90), e se manteve favorável ao grupo tratado nos anos seguintes. Observa-se uma elevação acentuada da legibilidade em 2016 (34,68), seguida por uma convergência parcial até 2020 e nova divergência a partir de 2021, com o grupo tratado atingindo 33,50 em 2023, frente a 31,65 no controle. Essa tendência reforça a hipótese de que a adesão a padrões internacionais estimula maior clareza textual.

Figura 12

Evolução da legibilidade média dos RAs por grupo de empresas.



Nota: Elaboração própria, 2025.

Para testar a hipótese **H1**, foi estimado um modelo de regressão com dados em painel desbalanceado, com efeitos fixos por setor e controles por ano. O objetivo foi avaliar a associação entre o índice de legibilidade dos relatórios e o grau de incerteza textual, medido pela frequência relativa de palavras ambíguas. O modelo incorpora variáveis linguísticas (comprimento, termos comuns e financeiros), características da empresa e controles macroeconômicos. Os erros padrão foram ajustados por clusterização temporal, conforme prática recomendada para painéis com estrutura desbalanceada.

A partir dos resultados apresentados na **Tabela 15**, verifica-se uma relação negativa, entre a legibilidade e a incerteza. Contudo, a ausência de significância estatística impede confirmar, com base nos dados, que a clareza textual exerça efeito direto sobre a incerteza. Ainda assim, o resultado mantém aderência à literatura que associa a clareza informacional à mitigação da assimetria (Li, 2008), mas sugere que, no contexto amostral analisado, esse efeito pode ser diluído por outras dimensões textuais e estruturais, sendo assim, não se confirma H1.

Tabela 15*Resultado da Regressão (H1).*

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
Legibilidade	-0,00001	0,00001	-1,31206	0,21061 **
Adrs	-0,00021	0,00014	-1,47377	0,16267
Idade	-0,00001	0,00001	-1,38380	0,18809
Length	0,00066	0,00011	5,89612	0,00004 ***
CommonWords	0,00946	0,00356	2,65585	0,01881 **
FinancialWords	-0,01730	0,00895	-1,93248	0,07379 *
TAM	-0,00012	0,00005	-2,32785	0,03543 **
TANGAT	0,00015	0,00024	0,62585	0,54148
VMP	0,00000	0,00001	-0,04674	0,96338
EPU	0,08871	0,04983	1,78022	0,09674 *
ICC	-0,00001	0,00000	-1,62681	0,12607
Crise	0,00016	0,00014	1,18125	0,25718
R ²			0,230364	
R ² (efeito fixo)			0,190303	
Observações			732	

Legenda: Legibilidade = grau de facilidade de leitura do texto; adrs = *dummy* para empresas listadas em ADRs; Idade = idade da empresa (em anos); Length = comprimento do relatório (log do número total de palavras); CommonWords = proporção de palavras comuns no texto; Legibilidade = índice Flesch adaptado ao português; FinancialWords = proporção de termos financeiros no texto; TAM = tamanho da empresa (log do total de ativos); TANGAT = tangibilidade dos ativos; VMP = valor de mercado sobre o patrimônio líquido; EPU = índice de incerteza da política econômica; ICC = índice de confiança do consumidor; Crise = *dummy* para anos com choques sistêmicos.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

Além da variável principal, os resultados evidenciam que o comprimento do relatório (*Length*) exerce um impacto positivo e significativo ao nível de 1% sobre a incerteza textual. Ou seja, relatórios mais longos estão associados a maior opacidade informacional. Esse achado sugere que o excesso de conteúdo pode dificultar a extração das mensagens principais ou gerar dispersão cognitiva, ampliando a incerteza percebida. Esse resultado reforça o argumento de que volume informacional não se traduz necessariamente em transparência, e que complexidade textual pode atuar como barreira à compreensão (Dalwai et al., 2021; Ginesti et al., 2017).

Em contrapartida, o resultado para o tamanho da empresa (*TAM*) revelou um coeficiente negativo e estatisticamente significativo ao nível de 10%, indicando que empresas de maior porte tendem a apresentar menor incerteza textual. Esse resultado é compatível com a existência de estruturas de governança mais consolidadas, maior capacidade de investimento em comunicação corporativa e um compromisso institucional mais robusto com a transparência (Hassan et al., 2022; Li, 2008; Loughran & McDonald, 2016).

A variável *CommonWords* apresentou coeficiente positivo e significativo em nível de 5%, indicando que o uso intensivo de palavras genéricas ou triviais não contribui para reduzir a incerteza textual. Em certos casos, pode até ampliá-la, ao empobrecer a precisão técnica e o valor informativo dos relatórios. Esse resultado está alinhado à literatura que associa a linguagem excessivamente simplificada, vaga ou redundante a maiores níveis de opacidade informacional e dificuldade de interpretação por parte dos *stakeholders* (Abernathy et al., 2019; Courtis, 2004; Dalwai et al., 2021; Li, 2008; Luo et al., 2018).

A tangibilidade dos Ativos (*TANGAT*) apresentou coeficiente positivo, porém sem significância estatística, o que não permite afirmar a existência de uma relação robusta entre maior proporção de ativos tangíveis e níveis de incerteza textual. Ainda assim, o sinal do coeficiente sugere que empresas com estruturas mais intensivas em ativos fixos poderiam apresentar padrões de divulgação menos detalhados no conteúdo textual, possivelmente pela ênfase em aspectos técnicos ou operacionais em detrimento da comunicação estratégica. Essa interpretação, contudo, deve ser vista com cautela, já que a ausência de significância impede inferências conclusivas. Estudos anteriores indicam que organizações com modelos de negócio mais padronizados e previsíveis enfrentam menor pressão do mercado por explicações complexas ou projeções estratégicas, o que pode reduzir o incentivo à elaboração de relatórios com maior clareza e profundidade qualitativa (Hodgdon & Hughes, 2016).

A variável *Crise* a presentou coeficiente positivo, porém sem significância estatística, o que não permite estabelecer uma relação estatisticamente robusta entre períodos de choques sistêmicos, como recessões econômicas, crises políticas ou instabilidades institucionais e a incerteza textual nos relatórios da administração. Apesar disso, o sinal do coeficiente sugere que contextos macroeconômicos adversos poderiam estar associados a padrões de comunicação mais cautelosos ou menos assertivos, refletindo maior dificuldade dos gestores em antecipar cenários ou fornecer

informações precisas. Tal interpretação deve ser considerada com cautela, uma vez que a ausência de significância impede conclusões definitivas. Pesquisas anteriores documentam que condições macroeconômicas desfavoráveis podem induzir ajustes na narrativa corporativa, tornando-a mais prudente ou defensiva (Ertugrul et al., 2017; González et al., 2019; Hassan et al., 2022; Pouryousof et al., 2022).

A variável *adrs*, não apresentou significância estatística neste modelo, o que indica que não se identificou associação direta entre o *status* de listagem via ADRs e a incerteza textual. Ressalta-se que neste modelo, a variável *adrs* atua como variável de controle estrutural, permitindo distinguir empresas listadas internacionalmente daquelas restritas ao mercado doméstico.

No caso do Valor de Mercado sobre o Patrimônio Líquido (*VMP*), o coeficiente estimado foi positivo, mas não apresentou significância estatística, o que não permite confirmar uma associação robusta entre maior valorização relativa e a incerteza textual dos relatórios. Ainda assim, o sinal positivo sugere que empresas mais valorizadas em termos relativos podem não necessariamente apresentar relatórios mais claros, possivelmente pela presença de ativos intangíveis relevantes, como reputação, capital intelectual ou expectativas de crescimento, que não são integralmente refletidos nas demonstrações contábeis e que podem ampliar a assimetria informacional. Essa interpretação, no entanto, deve ser feita com cautela, dado que a ausência de significância estatística impede conclusões definitivas. Evidências anteriores indicam que o valor de mercado elevado nem sempre se traduz em maior previsibilidade ou transparência na comunicação corporativa (Barth et al., 2017; Healy & Palepu, 2001).

A incerteza política (*EPU*), apresentou coeficiente positivo e estatisticamente significativo, indicando que níveis mais elevados de incerteza associada à política econômica tendem a aumentar a incerteza textual nos relatórios da administração. Esse resultado está alinhado à literatura que trata o *EPU* como uma proxy de risco sistêmico capaz de afetar decisões corporativas e padrões de divulgação (Baker et al., 2016). Em cenários de maior instabilidade política ou regulatória, as empresas podem adotar narrativas mais cautelosas ou ambíguas, refletindo a dificuldade de antecipar impactos sobre operações, investimentos e estratégias de longo prazo. Essa resposta comunicacional sugere que o ambiente político-econômico exerce influência direta na qualidade informacional, reforçando a relevância de incorporar variáveis macroeconômicas na análise da incerteza textual.

Por fim, o Índice de Confiança do Consumidor (*ICC*) apresentou coeficiente negativo, embora não tenha alcançado significância estatística ao nível tradicional. Apesar disso, o sinal estimado permanece compatível com a hipótese de que, em contextos macroeconômicos mais favoráveis, os gestores tendem a adotar um discurso mais direto, reduzindo a ambiguidade textual. Esse resultado indica que, em contextos macroeconômicos mais estáveis, os gestores tendem a adotar discursos mais diretos e objetivos, reduzindo a ambiguidade percebida nas mensagens corporativas (Courtis, 2004; Hassan et al., 2022).

De forma geral, os resultados indicam que a incerteza textual não é determinada apenas por aspectos linguísticos, mas por um conjunto de fatores estruturais e contextuais. Embora a legibilidade apresente relação negativa com a incerteza textual, a ausência de significância estatística impede confirmar a hipótese de efeito direto, sugerindo que seu impacto pode ser condicionado por outras dimensões textuais e contextuais.

Essas evidências empíricas possuem implicações práticas relevantes, sobretudo em ambientes como o brasileiro, marcados por elevada assimetria informacional e fragilidade institucional. Promover a clareza textual não apenas reduz custos de agência, como também fortalece a confiança dos investidores e a eficiência dos mercados de capitais (Cho et al., 2024).

4.2.2 Relação entre Tom Negativo e Incerteza (H2)

A hipótese H2 propôs que o tom negativo dos relatórios da administração estaria associado à redução da incerteza textual, desde que comunicado com clareza. Para testar essa hipótese, foi estimado um modelo com interação entre a proporção de palavras negativas no relatório da administração (*TomNeg*) e o grau de legibilidade dos relatórios (*Legibilidade*), considerando variáveis linguísticas, estruturais e macroeconômicas, com efeitos fixos por setor e erros padrão ajustados por ano.

Os resultados apresentados na **Tabela 16** indicam que a variável *TomNeg* apresentou coeficiente positivo e significativo ao nível de 1%, sugerindo que, de forma isolada, o uso de termos com carga negativa está associado a maior incerteza textual. No entanto, a interação entre *TomNeg* e *Legibilidade* exibiu coeficiente negativo e significativo ao nível de 1%, indicando que o efeito do tom negativo sobre a incerteza textual depende do nível de legibilidade do relatório. Em relatórios mais claros, a

negatividade tende a sinalizar transparência e reduzir a ambiguidade percebida; já em textos menos legíveis, o mesmo tom contribui para elevar a opacidade da comunicação.

Tabela 16

Resultado da Regressão (H2).

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor	
TomNeg	0,13619	0,02812	4.84262	0,00026	***
Legibilidade	0,00002	0,00001	2.21093	0,04419	*
adrs	-0,00025	0,00014	-1.79197	0,09477	*
Idade	-0,00002	0,00001	-2.42169	0,02961	**
Length	0,00070	0,00009	7.44610	0,00000	***
CommonWords	0,00711	0,00300	2.36654	0,03291	*
FinancialWords	-0,00645	0,00663	-0.97190	0,34759	
TAM	-0,00009	0,00005	-2.05657	0,05886	*
TANGAT	0,00006	0,00023	0.27066	0,79060	
VMP	0,00000	0,00001	-0.39436	0,69926	
EPU	0,08099	0,04580	1.76842	0,09876	*
Crise	0,00025	0,00008	2.98413	0,00986	**
TomNeg x Legibilidade	-0,00176	0,00080	-2.19011	0,04594	**
R ²		0,264115			
R ² (efeito fixo)		0,226914			
Observações		732			

Legenda: TomNeg = proporção de palavras negativas no relatório da administração; Legibilidade = grau de facilidade de leitura do texto, no modelo com interação, seu efeito isolado no modelo reflete apenas situações em que o tom negativo está ausente; adrs = *dummy* para empresas listadas em ADRs; Idade = idade da empresa (em anos); Length = comprimento do relatório (log do número total de palavras); CommonWords = proporção de palavras comuns no texto; FinancialWords = proporção de termos financeiros no texto; TAM = tamanho da empresa (log do total de ativos); TANGAT = tangibilidade dos ativos; VMP = valor de mercado sobre o patrimônio líquido; EPU = índice de incerteza da política econômica; ICC = índice de confiança do consumidor; Crise = *dummy* para anos com choques sistêmicos; TomNeg x Legibilidade = termo de interação entre o tom negativo do relatório e sua legibilidade; representa o efeito moderador da clareza textual sobre a relação entre o tom negativo e a incerteza.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

A variável Legibilidade, por sua vez, apresentou coeficiente positivo e significativo ao nível de 5%. Embora isso indique um aumento da incerteza quando o

tom negativo é nulo, condição limite no modelo, seu papel mais relevante reside no efeito moderador evidenciado pela interação com *TomNeg*. Na qual a clareza textual potencializa os efeitos do tom negativo, funcionando como um canal para sua interpretação como sinal informacional.

Sendo assim, o resultado está alinhado à Teoria da Sinalização, uma vez que confirma que a negatividade textual pode atuar como sinal de transparência e prudência, desde que seja comunicada de forma clara e estruturada. Essa interpretação encontra suporte em estudos anteriores, Elliott et al. (2024), Hassan et al. (2022) e Reichmann (2023) que argumenta que, na ausência de clareza, o tom negativo pode ser mal interpretado, gerando confusão ou exacerbando percepções de risco.

Outros achados complementam essa evidência central. O comprimento dos relatórios (*Length*) apresentou coeficiente positivo e significativo ao nível de 1%, confirmando que relatórios mais longos tendem a ampliar a incerteza textual. Esse resultado sugere que volume informacional não se traduz, necessariamente, em clareza. Contudo, a extensão excessiva pode dificultar a extração dos conteúdos essenciais e aumentar a ambiguidade da comunicação (Dalwai et al., 2021; Ginesti et al., 2017).

Entre os controles, o tamanho da empresa (*TAM*) apresentou coeficiente negativo e estatisticamente significativo ao nível de 10%, sugerindo que empresas maiores tendem a produzir relatórios mais claros, possivelmente em razão de estruturas mais robustas de governança, padronização da comunicação e maior capacidade de gestão da informação (Hassan et al., 2022; Loughran & McDonald, 2016).

De forma semelhante, a variável *Crise* apresentou coeficiente positivo e estatisticamente significativo ao nível de 1%. Esse resultado sugere que, em contextos de crise, como recessões econômicas, instabilidade política ou choques institucionais, os relatórios da administração tendem a expressar maior incerteza textual. Essa elevação pode refletir a dificuldade dos gestores em oferecer informações claras diante de cenários voláteis, ou ainda uma estratégia de comunicação mais vaga e previdosa, motivada pelo risco de julgamentos equivocados por parte do mercado. Esse achado é compatível com estudos que apontam que, em momentos de adversidade, o conteúdo dos relatórios tende a ser mais ambíguo, refletindo a

insegurança institucional e a cautela na exposição pública de estimativas e projeções (González et al., 2019; Pouryousof et al., 2022).

A variável *adrs*, apresentou significância estatística, mesmo sendo utilizada no modelo apenas como controle estrutural. Esse resultado sugere que o status de listagem internacional, isoladamente, exerce influência sobre a incerteza textual, ainda que não se considere sua interação temporal. A variável *Idade*, indica que empresas mais maduras tendem a apresentar menor incerteza textual, possivelmente em função de maior experiência na elaboração e divulgação de relatórios, bem como da consolidação de práticas de governança e comunicação ao longo do tempo.

A variável *CommonWords*, apresentou coeficiente positivo e significante, indicando que o uso mais frequente de palavras comuns, de vocabulário genérico ou pouco técnico, está associado ao aumento da incerteza textual percebida. Esse achado reforça a evidência de que a predominância de termos pouco específicos pode empobrecer a precisão da linguagem e comprometer a efetividade da comunicação, dificultando a interpretação por parte dos leitores (Abernathy et al., 2019; Courtis, 2004; Li, 2008; Luo et al., 2018).

A tangibilidade dos ativos (*TANGAT*) apresentou coeficiente positivo e não significativo. Embora sem significância estatística, o sinal positivo é compatível com a hipótese de que empresas com maior proporção de ativos concretos tendem a apresentar estruturas operacionais mais simples ou previsíveis, o que reduz a demanda por justificativas complexas, mas não necessariamente se traduz em clareza textual (Hodgdon & Hughes, 2016).

A variável *VMP*, apresentou coeficiente positivo e não significativo. Embora sem evidência estatística robusta, o sinal sugere que empresas mais valorizadas podem exibir níveis mais elevados de clareza comunicacional, possivelmente em razão de maior visibilidade no mercado e pressão por transparência. No entanto, a ausência de significância estatística impede afirmar que essa relação seja consistente, sendo plausível que fatores como a presença de ativos intangíveis ainda contribuam para a manutenção de assimetrias informacionais (Barth et al., 2017; Healy & Palepu, 2001).

No plano macroeconômico, a variável *EPU*, apresentou coeficiente positivo e estatisticamente significativo, indicando que níveis mais elevados de incerteza de política econômica estão associados ao aumento da incerteza textual nos relatórios da administração. Esse resultado sugere que, em cenários de instabilidade política e econômica, os gestores tendem a adotar comunicações mais cautelosas e,

possivelmente, menos assertivas, o que pode ampliar a percepção de opacidade informacional (Dalwai et al., 2021).

De modo geral, os resultados do modelo H2 reforçam a relevância do tom negativo como elemento estratégico na comunicação financeira, desde que inserido em um discurso claro e inteligível. A significância da interação entre *TomNeg* e *Legibilidade* confirma que a clareza textual não apenas reduz a opacidade informacional por si só, mas também modera os efeitos emocionais da linguagem, ampliando sua função sinalizadora. Essa evidência fortalece a ideia de que o valor informativo do tom depende de sua forma de transmissão, sendo a legibilidade um canal fundamental para a credibilidade discursiva. Ao mesmo tempo, os resultados reiteram que a incerteza textual é moldada por um conjunto amplo de fatores linguísticos, estruturais e contextuais, não sendo atribuída exclusivamente à escolha lexical ou ao tom predominante.

4.2.3 Efeito da Governança via ADRs sobre a Incerteza (H3)

A hipótese H3 postula que a adoção de padrões mais rigorosos de governança corporativa, por meio da listagem de ADRs no mercado norte-americano, contribui para a redução da incerteza textual nos relatórios da administração. Para testar essa hipótese, foi estimado um modelo econômético com estrutura DiD, aplicado após o PSM. Foram conduzidas duas especificações: uma considerando como marco temporal o ano do anúncio (anopre) e outra tomando como referência o ano efetivo da listagem (AnoDR). O modelo inclui, características textuais (*Legibilidade*, *Length*, *Commonwords*, *Financialwords* e *TomNeg*, características estruturais das empresas (tamanho, idade, tangibilidade e valor de mercado) e variáveis macroeconômicas (EPU, ICC e Crise). Os erros padrão foram ajustados por clusterização temporal.

Os resultados apresentados na **Tabela 17** revelam que os impactos da listagem via ADRs sobre a incerteza textual se concentram em janelas temporais específicas após o tratamento. No modelo baseado no anúncio (anopre), identifica-se reduções estatisticamente significativas da incerteza textual no quarto ano, no sexto ano e no nono ano após o anúncio. No modelo baseado na listagem oficial (AnoDR), os efeitos significativos surgem no quarto e no nono ano após a listagem, ambos com coeficientes negativos, reforçando a consistência do padrão identificado no modelo baseado no anúncio.

Tabela 17*Resultado da Regressão DiD (H3).*

Variável	Anúncio (anopre)	Sig	Listagem (AnoDR)	Sig
tempo_relativo = -12	-0,0006 (0,0007)		0,0017 (0,0013)	
tempo_relativo = -11	0,0002 (0,0008)		0,0003 (0,0008)	
tempo_relativo = -10	-0,0002 (0,0008)		0,0006 (0,0009)	
tempo_relativo = -9	0,0006 (0,0009)		0,0007 (0,0007)	
tempo_relativo = -8	0,0004 (0,0007)		0,0008 (0,0007)	
tempo_relativo = -7	0,0004 (0,0007)		0,0001 (0,0007)	
tempo_relativo = -6	0,0003 (0,0007)		-0,0003 (0,0006)	
tempo_relativo = -5	0,00002 (0,0006)		-0,00005 (0,0006)	
tempo_relativo = -4	-0,0005 (0,0006)		-0,0003 (0,0005)	
tempo_relativo = -3	-0,00005 (0,0006)		0,0002 (0,0005)	
tempo_relativo = -2	0,0001 (0,0005)		-0,0001 (0,0003)	
tempo_relativo = 0	-0,0004 (0,0003)		-0,000037 (0,0004)	
tempo_relativo = 1	-0,0002 (0,0004)		-0,0002 (0,0004)	
tempo_relativo = 2	-0,0004 (0,0004)		-0,0003 (0,0005)	
tempo_relativo = 3	-0,0005 (0,0004)		-0,0004 (0,0004)	
tempo_relativo = 4	-0,0007 (0,0004)		-0,0009 (0,0004)	*
tempo_relativo = 5	-0,00006 (0,0006)		0,0004 (0,0009)	
tempo_relativo = 6	-0,0009 (0,0004)	*	-0,0009 (0,0004)	*
tempo_relativo = 7	0,0003 (0,0008)		0,0001 (0,0008)	
tempo_relativo = 8	-0,0003 (0,0004)		-0,0003 (0,0004)	
tempo_relativo = 9	-0,0011 (0,0004)	**	-0,0009 (0,0004)	*
tempo_relativo = 10	-0,0008 (0,0006)		-0,0006 (0,0004)	
tempo_relativo = 11	-0,0006 (0,0007)		-0,0007 (0,0006)	
tempo_relativo = 12	-0,0009 (0,0004)		-0,0009 (0,0004)	*
tempo_relativo = 13	-0,0006 (0,0003)		-0,0008 (0,0004)	*
TomNeg	0,0861 (0,0253)	**	0,0860 (0,0269)	**
Length	0,0005 (0,0001)	***	0,0005 (0,0001)	***

Tabela 17

Resultado da Regressão DiD (H3).

Variável	Anúncio (anopre)	Sig	Listagem (AnoDR)	Sig
CommonWords	0,0017 (0,0063)		0,0006 (0,0066)	
Legibilidade	-0,0000 (0,0000)		-0,0000 (0,0000)	
FinancialWords	0,0182 (0,0150)		0,0212 (0,0156)	
TAM	-0,00003 (0,0003)		-0,0000 (0,0003)	
TANGAT	0,0020 (0,0010)	*	0,0019 (0,0010)	
VMP	-0,00001 (0,0000)		-0,0000 (0,0000)	
EPU	-0,0534 (0,0592)		-0,0303 (0,0565)	
R ²	0,50500		0,51133	
R ² (efeito fixo)	0,21758		0,22759	
Observações	382		382	

Legenda: *tempo_relativo* = indica a distância temporal em anos entre a observação e o momento da listagem via ADRs (t_0). Valores negativos representam períodos anteriores ao evento (*leads*), e valores positivos representam períodos posteriores (*lags*); Idade = idade da empresa (em anos); TomNeg = proporção de palavras negativas no relatório da administração; representa o tom textual com carga emocional negativa quando a Legibilidade é zero; Length = comprimento do relatório (log do número total de palavras); CommonWords = proporção de palavras comuns no texto; Legibilidade = grau de facilidade de leitura do texto, no modelo com interação, seu efeito isolado no modelo reflete apenas situações em que o tom negativo está ausente; FinancialWords = proporção de termos financeiros no texto; TAM = tamanho da empresa (log do total de ativos); TANGAT = tangibilidade dos ativos; VMP = valor de mercado sobre o patrimônio líquido; EPU = índice de incerteza da política econômica; ICC = Índice de confiança do consumidor; Crise = *dummy* para anos com choques sistêmicos; TomNeg x Legibilidade = termo de interação entre o tom negativo do relatório e sua legibilidade; representa o efeito moderador da clareza textual sobre a relação entre o tom negativo e a incerteza.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

Sendo assim, apesar dos resultados indicarem que, em determinados momentos, a listagem esteve associada a menor incerteza textual em relação ao grupo de controle, o padrão não se mostrou persistente ao longo de todo o horizonte analisado. Dessa forma, os achados não oferecem evidência empírica suficiente para confirmar H3, sugerindo que o efeito da governança aprimorada sobre a opacidade informacional pode ser contingente a condições específicas do período ou do contexto da divulgação.

O modelo DiD também permite observar os efeitos marginais das variáveis textuais incluídas como controles, especialmente *Legibilidade* e *TomNeg*. A variável *Legibilidade* apresentou coeficiente negativo, porém, não significativo, sugerindo que relatórios com maior clareza textual estão associados a menores níveis de incerteza.

Esse padrão, ainda que não robusto, é coerente com a premissa de que práticas de governança mais rigorosas podem estimular melhorias qualitativas na linguagem dos reportes corporativos.

Em consonância com a Teoria da Agência, a maior legibilidade pode ser interpretada como uma tentativa de reduzir os custos de monitoramento e os riscos de assimetria informacional (Healy & Palepu, 2001; Li, 2008). Além disso, sob a lente da Teoria da Sinalização, a clareza textual atua como um indicativo de transparência e alinhamento com as expectativas do mercado (Dalwai et al., 2021; Ginesti et al., 2017). No entanto, a falta de significância sugere que esse efeito não é sistemático na amostra analisada, podendo depender de fatores contextuais ou de outras dimensões qualitativas da divulgação.

Por outro lado, a variável *TomNeg* apresentou coeficiente positivo e estatisticamente significativo ao nível de 5%, indicando que o uso mais intenso de termos negativos nos relatórios está associado ao aumento da incerteza textual. Esse resultado pode ser compreendido à luz de estudos que apontam o uso estratégico da negatividade como mecanismo de evasão informacional. Em contextos de baixa legibilidade ou ausência de suporte contextual, o tom negativo tende a gerar ambiguidade interpretativa e a amplificar a percepção de risco, especialmente entre investidores externos menos familiarizados com o ambiente institucional brasileiro (Rajabalizadeh, 2023; Reichmann, 2023).

Para as demais características textuais, o comprimento dos relatórios (*Length*) mostrou coeficiente positivo e significativo ao nível de 1%, indicando que relatórios mais longos tendem a aumentar a incerteza, possivelmente devido à sobrecarga informacional e à dificuldade de síntese (Bar-Hod et al., 2021; Ertugrul et al., 2017; Li, 2008; Saravanan et al., 2024).

Contudo, a variável *Commonwords* também apresentou coeficiente positivo, porém não significativo, indicando que, neste modelo, não se encontrou evidência estatística robusta de que o uso de vocabulário genérico esteja diretamente associado à elevação da incerteza textual. Contudo, o sinal positivo é compatível com a literatura, que sugere que o uso excessivo de vocabulário genérico pode comprometer a precisão técnica do texto, elevando a percepção de ambiguidade (Abernathy et al., 2019; Dalwai et al., 2021; Li, 2008).

No mesmo sentido, a variável *Financialwords* apresentou coeficiente positivo e não significativo, sugerindo que o uso mais frequente de termos financeiros está

associado, ainda que de forma estatisticamente não robusta, a níveis mais elevados de incerteza textual. Esse resultado indica que a simples incorporação de jargões técnicos não garante maior precisão comunicacional, podendo inclusive gerar ambiguidade se não acompanhada de explicações claras e contextualizadas (Loughran & McDonald, 2014; Luo et al., 2018).

Entre os controles estruturais, o tamanho da empresa (*TAM*) apresentou coeficiente positivo e não significativo, o que impede afirmar com evidência estatística robusta que empresas maiores divulguem relatórios com menor incerteza textual. Ainda assim, o sinal do coeficiente está alinhado à literatura que associa empresas de maior porte a estruturas de governança mais consolidadas e maior capacidade de processamento e divulgação de informações, fatores que podem favorecer uma comunicação mais clara e menos ambígua (Hassan et al., 2022; Healy & Palepu, 2001; Li, 2008; Loughran & McDonald, 2016).

Por outro lado, a tangibilidade dos ativos (*TANGAT*), apresentou coeficiente positivo e estatisticamente significativo, apenas para o modelo baseado no anúncio, sugerindo que empresas com maior proporção de ativos tangíveis e, portanto, com operações mais previsíveis e facilmente mensuráveis, tendem a apresentar níveis mais elevados de incerteza textual. Esse resultado indica que, mesmo com uma base operacional mais concreta, tais empresas podem não priorizar a clareza comunicacional em seus relatórios, possivelmente por confiarem que a transparência operacional decorrente de seus ativos físicos reduza a necessidade de detalhamento textual. Esse achado está em consonância com discussões como as de Hodgdon e Hughes (2016), que apontam que a natureza dos ativos influencia, mas não garante, padrões uniformes de *disclosure*.

A variável *VMP* apresentou coeficiente negativo e não significativo, sugerindo que empresas mais valorizadas em termos relativos não se beneficiam automaticamente de maior clareza comunicacional. Isso reforça a ideia de que o valor de mercado, ao refletir ativos intangíveis e expectativas futuras, não se traduz diretamente em qualidade textual (Barth et al., 2017; Healy & Palepu, 2001).

Em relação ao índice EPU, o coeficiente foi negativo e não significativo, o que não confirma a expectativa de que ambientes econômicos instáveis aumentem a opacidade textual nos relatórios da administração. Esse resultado sugere que, em contextos de maior incerteza política e econômica, os gestores podem adotar um discurso mais claro e direto, possivelmente como forma de mitigar interpretações

adversas do mercado (Dalwai et al., 2021), embora tal associação não tenha se confirmado estatisticamente neste modelo.

Os resultados não oferecem suporte estatisticamente robusto à hipótese H3, embora indiquem sinais compatíveis com a expectativa teórica de que a governança corporativa aprimorada, viabilizada pela listagem via ADRs, possa contribuir para a redução da incerteza textual nos relatórios da administração. A concentração dos efeitos em poucos períodos e o número reduzido de empresas tratadas, aliados à dispersão temporal entre o anúncio e a efetiva listagem, podem ter limitado o poder estatístico do modelo e influenciado a detecção de efeitos consistentes.

Ainda assim, esses achados preservam relevância prática e teórica ao apontar que avanços em governança e padrões mais rigorosos de divulgação narrativa permanecem como vetores potenciais de mitigação da opacidade informacional e fortalecimento da eficiência dos mercados.

4.3 Considerações Finais da Análise das Hipóteses

Os resultados apresentados neste capítulo ajudam a entender melhor o que torna um relatório da administração mais ou menos incerto para quem o lê. A análise mostrou que fatores como a clareza do texto, o tom da mensagem e o ambiente institucional em que a empresa está inserida influenciam diretamente a forma como a informação é percebida pelos usuários.

A primeira hipótese testada (H1) previa que relatórios mais fáceis de ler tenderiam a gerar menos incerteza. Embora o coeficiente da variável legibilidade tenha apresentado o sinal esperado (negativo), o efeito não foi estatisticamente significativo, o que impede confirmar a hipótese. Ainda assim, o resultado é coerente com a literatura, que associa maior clareza textual à redução da opacidade informacional, reforçando a importância de uma comunicação objetiva, especialmente em ambientes com maior assimetria de informação, como o mercado brasileiro.

Já no teste da segunda hipótese (H2), que investigou o papel do tom negativo, ou seja, a presença de palavras que expressam cautela, alerta ou preocupação, os resultados foram mais complexos. Quando analisado isoladamente, o uso de termos negativos está associado a maior incerteza. No entanto, quando combinado a textos bem escritos e organizados, o tom negativo tem o efeito contrário: ajuda a reduzir a opacidade da informação. Isso significa que não basta adotar uma postura prudente,

é preciso comunicá-la de forma clara para que ela gere confiança. Relatórios com linguagem negativa e pouco claros tendem a confundir mais do que informar.

A terceira e principal hipótese da pesquisa (H3) avaliou se a listagem de empresas brasileiras no mercado norte-americano por meio de ADRs ajuda a melhorar a qualidade da comunicação nos relatórios. Os resultados não confirmaram essa expectativa de forma consistente ao longo do período analisado. Apesar de alguns indícios positivos pontuais, o conjunto de resultados sugere que o efeito potencial da governança corporativa aprimorada, viabilizada pelas exigências da SEC e pelos padrões de *disclosure* norte-americanos, pode ser limitado ou sensível a condições específicas.

Em conjunto, esses achados mostram que a incerteza textual não depende apenas do conteúdo, mas também da forma e do contexto em que a mensagem é construída. A clareza do texto, o tom adotado e os compromissos com boas práticas de governança se combinam para gerar comunicações mais confiáveis, transparentes e úteis para o mercado. Isso tem implicações importantes para reguladores, investidores e profissionais de relações com investidores, pois reforça o valor estratégico da linguagem nas decisões corporativas.

4.4 Teste de Robustez

Com o objetivo de verificar a robustez dos resultados do modelo principal em relação à definição da variável dependente, foi estimado um modelo DiD escalonado substituindo a medida de incerteza textual (Incerteza) pelo índice EPU (*Economic Policy Uncertainty*). Para evitar sobreposição conceitual, o EPU foi retirado do conjunto de variáveis de controle, mantendo-se as demais covariáveis originais e aplicando efeitos fixos por empresa e erros-padrão agrupados por ID.

A **Tabela 18** apresenta os coeficientes estimados para os períodos relativos ao evento (listagem via ADRs), abrangendo todos os anos anteriores e posteriores ao ano base ($t = 0$). Os resultados indicam que não foram identificados efeitos estatisticamente significativos no período pré-tratamento, o que reforça a plausibilidade da suposição de tendências paralelas. No período pós-tratamento, apenas $t = 5$ apresentou significância estatística ao nível de 5% (coeficiente = -0,00105) e $t = 9$ apresentou significância estatística ao nível de 10% (coeficiente =

-0,00081), mas com magnitude muito reduzida e sem configuração de padrão sistemático de impacto ao longo dos anos posteriores.

Esse comportamento sugere que a relação identificada nos modelos principais não se deve a flutuações conjunturais capturadas pelo EPU, aumentando a confiança na robustez dos resultados. Além disso, as variáveis de controle mantiveram comportamento consistente com a literatura: *Length* (comprimento do relatório) apresentou coeficiente negativo e significativo a 1%, enquanto *CommonWords* manteve coeficiente positivo e significativo a 1%, indicando que textos mais longos e mais densos em vocabulário comum estão associados a níveis diferentes do índice de incerteza macroeconômica.

Tabela 18

Resultado da Regressão DiD – Teste de Robustez – Incerteza Alternativa (EPU).

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
tempo_relativo:::-14	0.00023	0.00085	0.27420	0.78575
tempo_relativo:::-13	-0.00004	0.00067	-0.06325	0.94997
tempo_relativo:::-12	-0.00061	0.00056	-1.09768	0.28080
tempo_relativo:::-11	-0.00016	0.00050	-0.32241	0.74931
tempo_relativo:::-10	0.00003	0.00045	0.06921	0.94527
tempo_relativo:::-9	0.00082	0.00066	1.24025	0.22419
tempo_relativo:::-8	0.00050	0.00062	0.81447	0.42159
tempo_relativo:::-7	0.00018	0.00040	0.45132	0.65490
tempo_relativo:::-6	-0.00030	0.00045	-0.65973	0.51430
tempo_relativo:::-5	-0.00036	0.00032	-1.13791	0.26388
tempo_relativo:::-4	-0.00022	0.00027	-0.81523	0.42116
tempo_relativo:::-3	0.00002	0.00020	0.11736	0.90733
tempo_relativo:::-2	0.00038	0.00030	1.26456	0.21545
tempo_relativo:::-0	0.00009	0.00025	0.35509	0.72493
tempo_relativo:::1	-0.00009	0.00025	-0.36234	0.71956
tempo_relativo:::2	0.00010	0.00030	0.31758	0.75293
tempo_relativo:::3	-0.00009	0.00027	-0.34495	0.73246
tempo_relativo:::4	-0.00043	0.00028	-1.52023	0.13859
tempo_relativo:::5	-0.00105	0.00049	-2.12972	0.04124 **
tempo_relativo:::6	-0.00017	0.00040	-0.42573	0.67324
tempo_relativo:::7	-0.00011	0.00042	-0.25784	0.79823
tempo_relativo:::8	-0.00041	0.00041	-0.99799	0.32601
tempo_relativo:::9	-0.00081	0.00041	-1.98563	0.05598 *
tempo_relativo:::10	-0.00040	0.00082	-0.48240	0.63291
tempo_relativo:::11	0.00036	0.00104	0.34688	0.73103
tempo_relativo:::12	-0.00072	0.00060	-1.20735	0.23643
tempo_relativo:::13	-0.00076	0.00057	-1.34975	0.18687
tempo_relativo:::14	-0.00100	0.00061	-1.62693	0.11388
TomNeg	0.01114	0.01979	0.56261	0.57775
Length	-0.00045	0.00012	-3.81708	0.00061 ***
CommonWords	0.01782	0.00454	3.92165	0.00045 ***
Legibilidade	0.00000	0.00002	-0.03996	0.96838
FinancialWords	-0.02175	0.02045	-1.06349	0.29578
TAM	0.00039	0.00020	1.97527	0.05720 *
TANGAT	-0.00046	0.00098	-0.47379	0.63897

Tabela 18*Resultado da Regressão DiD – Teste de Robustez – Incerteza Alternativa (EPU).*

Variável	Coeficiente	Erro Padrão	Estatística t	p-valor
VMP	0.00001	0.00001	0.79877	0.43050
Crise	0.00011	0.00015	0.75010	0.45885
ICC	0.00001	0.00001	1.23917	0.22458
R ²		0,57183		
R ² (efeito fixo)		0,32499		
Observações		382		

Legenda: *tempo_relativo* = indica a distância temporal em anos entre a observação e o momento da listagem via ADRs (t_0). Valores negativos representam períodos anteriores ao evento (*leads*), e valores positivos representam períodos posteriores (*lags*); *TomNeg* = proporção de palavras negativas no relatório da administração; *Length* = comprimento do relatório (log do número total de palavras); *CommonWords* = proporção de palavras comuns no texto; *Legibilidade* = grau de facilidade de leitura do texto; *FinancialWords* = proporção de termos financeiros no texto; *TAM* = tamanho da empresa (log do total de ativos); *TANGAT* = tangibilidade dos ativos (ativo imobilizado / total de ativos); *VMP* = razão entre valor de mercado e patrimônio líquido; *EPU* = índice de incerteza da política econômica; *ICC* = Índice de confiança do consumidor; *Crise* = *dummy* para anos com choques sistêmicos.

Nota: *, ** e *** representa significância estatística de 10%, 5% e 1% respectivamente.

Fonte: Dados da Pesquisa (2025).

5. CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo avaliar o impacto da listagem de empresas brasileiras no mercado norte-americano, por meio de ADRs, sobre a legibilidade, o tom e a incerteza textual dos relatórios da administração. A pesquisa foi motivada pela premissa de que a governança corporativa exerce um papel fundamental na qualidade da divulgação financeira, sobretudo no contexto de assimetria informacional característico dos mercados emergentes.

Os resultados não confirmam que a listagem via ADRs implica a adoção de padrões de governança mais rigorosos, os quais repercutem positivamente na qualidade da comunicação textual, segundo a literatura. A hipótese **H1** não foi confirmada, uma vez que, apesar de o coeficiente da legibilidade ter apresentado o sinal esperado (negativo), o efeito não foi estatisticamente significativo. Ainda assim, o resultado é coerente com a literatura, reforçando o entendimento de que a clareza na comunicação contribui para mitigar a assimetria informacional, ao facilitar a interpretação das mensagens pelos *stakeholders*.

A hipótese **H2** se confirma, ou seja, quando os relatórios apresentam maior clareza, o uso do tom negativo contribui para reduzir a opacidade informacional. Em outras palavras, a legibilidade atua como moderadora, ou seja, a clareza dos relatórios transforma o impacto do tom negativo, permitindo que mensagens prudentes e realistas sinalizem transparência, ao invés de confusão.

A hipótese **H3** não foi confirmada, representando o principal achado deste estudo. A listagem no mercado norte-americano, via ADRs, está significativamente associada à redução da incerteza textual, indicando que empresas sujeitas a regulação mais rigorosa aprimoram substancialmente a qualidade da comunicação informacional. Esse efeito evidencia o papel disciplinador de ambientes regulatórios mais exigentes sobre a divulgação financeira.

O modelo revelou ainda achados complementares. O comprimento dos relatórios (*Length*) apresentou associação positiva com a incerteza, indicando que documentos excessivamente longos tendem a aumentar a complexidade e dificultar a compreensão. A variável *CommonWords* apresentou coeficiente positivo e significativo, sugerindo que vocabulário genérico excessivo comprometer a precisão técnica da comunicação.

Esses resultados indicam que a incerteza textual é influenciada não apenas por atributos formais da linguagem, como legibilidade e tom, mas também pelo contexto estrutural e macroeconômico em que os relatórios são produzidos. Os achados dialogam com os fundamentos da Teoria da Agência e da Teoria da Sinalização, ainda que com evidências mistas, sugerindo que práticas de governança mais rigorosas e estratégias comunicacionais prudentes funcionam como instrumentos de redução da assimetria.

Assim, os resultados reforçam os argumentos centrais da Teoria da Sinalização, especialmente no que se refere ao uso do tom negativo aliado à clareza textual (H2), pois, operam como sinais críveis de transparência, qualidade da governança e alinhamento com os interesses dos investidores.

Sob a ótica da Teoria da Agência, os achados não permitem afirmar que a listagem via ADRs funcione consistentemente como um mecanismo eficaz de mitigação dos problemas de agência, embora o contexto regulatório mais exigentes possa oferecer incentivos para o aprimoramento da qualidade informacional.

De forma integrada, os resultados desta pesquisa sugerem que fatores como o uso adequado do tom negativo em relatórios claros pode, contribuir para reduzir a incerteza textual, fortalecendo a qualidade da comunicação financeira e contribuindo para mercados mais transparentes e eficientes, ainda que os efeitos não sejam uniformes em todos os contextos analisados.

Do ponto de vista acadêmico, a principal contribuição deste estudo reside na análise integrada dos efeitos da governança corporativa, da legibilidade textual e do tom sobre a incerteza informacional. Essa abordagem preenche uma lacuna importante na literatura, que historicamente tem tratado essas dimensões de forma isolada. Ao reunir esses construtos em um mesmo arcabouço empírico, o estudo permite compreender de maneira mais abrangente como mecanismos formais de governança e atributos qualitativos da comunicação podem interagir na mitigação da assimetria informacional, especialmente em mercados emergentes como o brasileiro.

Do ponto de vista metodológico, a pesquisa inova ao aplicar técnicas avançadas de *Machine Learning* à análise textual de relatórios corporativos, uma abordagem ainda pouco explorada na literatura nacional. A utilização de algoritmos para mensuração de legibilidade, classificação de tom e quantificação de incerteza, combinada a métodos econôméticos robustos, como o PSM e o DiD, eleva o rigor

analítico e oferece uma contribuição replicável para futuros estudos em contabilidade, finanças e governança corporativa.

Do ponto de vista prático, os resultados oferecem subsídios relevantes para o ambiente regulatório e para formuladores de políticas públicas no Brasil. Embora a relação esperada entre a legibilidade e a redução da incerteza não tenha se confirmado de forma estatisticamente significativa, observou-se que o uso do tom negativo, quando combinado com maior clareza textual, pode contribuir para reduzir a opacidade informacional. Dessa forma, os resultados fornecem evidências que podem auxiliar no desenvolvimento de diretrizes que promovam uma comunicação financeira mais clara, objetiva e transparente.

Para as empresas, os achados sugerem que a adoção de boas práticas de governança, incluindo, em alguns contextos, a listagem de ADRs, pode representar um diferencial estratégico, embora os efeitos positivos sobre a incerteza textual não tenham se manifestado de forma consistente em todo o período analisado. A clareza textual e o uso adequado do tom, sobretudo em contextos de risco ou adversidade, podem contribuir para a redução da percepção de risco pelos investidores, a melhora da reputação corporativa e o aumento da atratividade perante o mercado de capitais. Isso reforça que investir na qualidade da comunicação financeira é uma decisão alinhada aos interesses de longo prazo da organização.

Para investidores e analistas, os resultados oferecem indícios que podem ser explorados no aprimoramento dos modelos de avaliação de risco informacional. A pesquisa evidencia que a qualidade textual dos relatórios, medida pela legibilidade, pelo tom e pelo grau de incerteza linguística, pode refletir as práticas de governança adotadas pelas empresas. Dessa forma, a análise qualitativa dos relatórios pode ser incorporada como uma variável adicional na avaliação de risco, na precificação de ativos e na tomada de decisão de investimentos, com o suporte de técnicas de processamento de linguagem natural.

Apesar dos avanços teóricos, metodológicos e empíricos oferecidos, esta pesquisa apresenta algumas limitações que devem ser reconhecidas. Uma primeira limitação reside na natureza do painel desbalanceado, característico de estudos que utilizam dados de empresas listadas em diferentes períodos, o que pode gerar pequenas perdas de precisão em determinadas estimativas, bem como, limitar a generalização dos resultados, especialmente quando o número de empresas tratadas é reduzido e a dispersão temporal dos eventos são maiores.

Além disso, embora os algoritmos de *Machine Learning* ofereçam elevada capacidade de mensuração objetiva, eles não capturam nuances subjetivas, como ironia, ambiguidade, contexto cultural ou intencionalidade discursiva, o que pode gerar simplificações na interpretação do conteúdo. Recomenda-se, portanto, a incorporação de modelos baseados em inteligência artificial generativa, capazes de captar aspectos discursivos mais subjetivos e contextuais.

REFERÊNCIAS

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1), 1–19. <https://doi.org/10.1111/0034-6527.00321>
- Abernathy, J. L., Guo, F., Kubick, T. R., & Masli, A. (2019). Financial statement footnote readability and corporate audit outcomes. *Auditing*, 38(2), 1–26. <https://doi.org/10.2308/ajpt-52243>
- Aerts, W. (2005). Picking up the pieces: Impression management in the retrospective attributional framing of accounting outcomes. *Accounting, Organizations and Society*, 30(6), 493–517. <https://doi.org/10.1016/j.aos.2004.07.001>
- Aldahray, A. (2024). Notes Readability and Discretionary Accruals | Legibilidad de las notas y devengo discrecional. *Revista de Contabilidad-Spanish Accounting Review*, 27(2), 229–238. <https://doi.org/10.6018/rccsr.459281>
- Ammer, M. A., & Salim, A. S. (2024). INVESTMENT AND FINANCING DECISIONS IMPACT ON FINANCIAL SUSTAINABILITY WITH MODERATING EFFECT OF CORPORATE GOVERNANCE INDEX: A DYNAMIC PANEL DATA APPROACH. *International Journal of Economics and Finance Studies*, 16(1), 62–82. <https://doi.org/10.34109/ijefs.202416103>
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Ed.).
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Ayuningtyas, E. S., & Harymawan, I. (2021). Negative Tone and Readability in Management Discussion and Analysis Reports: Impact on The Cost of Debt. *Jurnal Manajemen Teori Dan Terapan| Journal of Theory and Applied Management*, 14(2), 129. <https://doi.org/10.20473/jmtt.v14i2.26681>
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/https://doi.org/10.1093/qje/qjw024>
- Bar-Hod, A., Chen, E., & Gavious, I. (2021). The economic consequences of fair value disclosures: a manifestation of the buried facts doctrine. *Accounting and Finance*, 61(S1), 1363–1413. <https://doi.org/10.1111/acfi.12630>
- Barth, M. E., Cahan, S. F., Chen, L., & Venter, E. R. (2017). The economic consequences associated with integrated report quality: Capital market and real effects. *Accounting, Organizations and Society*, 62, 43–64. <https://doi.org/10.1016/j.aos.2017.08.005>
- Ben-Amar, W., & Belgacem, I. (2018). Do socially responsible firms provide more readable disclosures in annual reports? *Corporate Social Responsibility and Environmental Management*, 25(5), 1009–1018. <https://doi.org/10.1002/csr.1517>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Biehl, H., Bleibtreu, C., & Stefani, U. (2024). The real effects of financial reporting: Evidence and suggestions for future research. *Journal of International Accounting, Auditing and Taxation*, 54, 100594. <https://doi.org/10.1016/j.intaccaudtax.2023.100594>
- Brasil. Congresso Nacional. (2023). *Projeto de Lei n. 2925*.
- Brunelli, S., Carlino, C., Castellano, R., & Giosi, A. (2021). Going concern modifications and related disclosures in the Italian stock market: do regulatory

- improvements help investors in capturing financial distress? *Journal of Management and Governance*, 25(2), 433–473. <https://doi.org/10.1007/s10997-020-09537-7>
- Carlini, F., Cucinelli, D., Previtali, D., & Soana, M. G. (2020). Don't talk too bad! stock market reactions to bank corporate governance news. *Journal of Banking & Finance*, 121, 105962. <https://doi.org/10.1016/j.jbankfin.2020.105962>
- Cho, S. J., Choi, C., & Chung, C. Y. (2024). Firm information and risk: Evidence from the role of 10-K report readability. *Bulletin of Economic Research*, 76(2), 488–507. <https://doi.org/10.1111/boer.12435>
- Coffee, J. C. (2002). Racing towards the Top?: The Impact of Cross-Listings and Stock Market Competition on International Corporate Governance. *Columbia Law Review*, 102(7), 1757. <https://doi.org/10.2307/1123661>
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67. <https://doi.org/10.1177/0149206310388419>
- Courtis, J. K. (2004). Corporate report obfuscation: Artefact or phenomenon? *British Accounting Review*, 36(3), 291–312. <https://doi.org/10.1016/j.bar.2004.03.005>
- Dalwai, T., Chinnasamy, G., & Mohammadi, S. S. (2021). Annual report readability, agency costs, firm performance: an investigation of Oman's financial sector. *Journal of Accounting in Emerging Economies*, 11(2), 247–277. <https://doi.org/10.1108/JAEE-06-2020-0142>
- Dau, N. H., Nguyen, D. Van, & Diem, H. T. T. (2024). Annual report readability and firms' investment decisions. *Cogent Economics & Finance*, 12(1). <https://doi.org/10.1080/23322039.2023.2296230>
- Dodge, C., Karolyi, G. A., & Stulz, R. M. (2004). Why are foreign firms listed in the U.S. worth more? *Journal of Financial Economics*, 71(2), 205–238. [https://doi.org/10.1016/S0304-405X\(03\)00183-1](https://doi.org/10.1016/S0304-405X(03)00183-1)
- Dye, R. A. (1985). Disclosure of Nonproprietary Information. In *Journal of Accounting Research* (Vol. 23, Issue 1).
- Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, 14(1), 57. <https://doi.org/10.2307/258191>
- Elliott, W. B., Loftus, S., & Winn, A. (2024). To read or to listen? Does disclosure delivery mode impact investors' reactions to managers' tone language? *Contemporary Accounting Research*, 41(1), 7–38. <https://doi.org/10.1111/1911-3846.12898>
- Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. *Journal of Financial and Quantitative Analysis*, 52(2), 811–836. <https://doi.org/10.1017/S0022109017000187>
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383. <https://doi.org/10.2307/2325486>
- Ferreira, T. S. V., & Martins, O. S. (2024). Relato integrado e criação de valor para os acionistas: evidências internacionais. *Revista Contabilidade & Finanças*, 35(95). <https://doi.org/10.1590/1808-057x20241896.pt>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Ginesti, G., Drago, C., Macchioni, R., & Sannino, G. (2018). Female board participation and annual report readability in firms with boardroom connections. *Gender in Management: An International Journal*, 33(4), 296–314. <https://doi.org/10.1108/GM-07-2017-0079>

- Ginesti, G., Sannino, G., & Drago, C. (2017). Board connections and management commentary readability: the role of information sharing in Italy. *Corporate Governance (Bingley)*, 17(1), 30–47. <https://doi.org/10.1108/CG-01-2016-0015>
- Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate Governance and Equity Prices. *The Quarterly Journal of Economics*, 118(1), 107–156. <https://doi.org/10.1162/00335530360535162>
- González, M., Guzmán, A., Téllez, D. F., & Trujillo, M. A. (2021). What you say and how you say it: Information disclosure in Latin American firms. *Journal of Business Research*, 127, 427–443. <https://doi.org/10.1016/j.jbusres.2019.05.014>
- González, M., Guzmán, A., Tellez-Falla, D. F., & Trujillo, M. A. (2019). Governance, sentiment analysis, and initial public offering underpricing. *Corporate Governance: An International Review*, 27(3), 226–244. <https://doi.org/10.1111/corg.12272>
- Grove, H., Patelli, L., Victoravich, L. M., & Xu, P. T. (2011). Corporate Governance and Performance in the Wake of the Financial Crisis: Evidence from US Commercial Banks. *Corporate Governance: An International Review*, 19(5), 418–436. <https://doi.org/10.1111/j.1467-8683.2011.00882.x>
- Hassan, M. K., Abu-Abbas, B., & Kamel, H. (2022). Tone, readability and financial risk: the case of GCC banks. *Journal of Accounting in Emerging Economies*, 12(4), 716–740. <https://doi.org/10.1108/JAEE-06-2021-0192>
- Hassanein, A., & Albitar, K. (2024). An inverted U-shaped relationship between reporting risk information and corporate value: evidence from the UK. *Review of Managerial Science*. <https://doi.org/10.1007/s11846-024-00832-3>
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1–3), 405–440. [https://doi.org/10.1016/S0165-4101\(01\)00018-0](https://doi.org/10.1016/S0165-4101(01)00018-0)
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4), 605–654. <https://doi.org/10.2307/2971733>
- Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication*, 45(4), 363–407. <https://doi.org/10.1177/0021943608319388>
- Hodgdon, C., & Hughes, S. B. (2016). The effect of corporate governance, auditor choice and global activities on EU company disclosures of estimates and judgments. *Journal of International Accounting, Auditing and Taxation*, 26, 28–46. <https://doi.org/10.1016/j.intaccaudtax.2016.02.002>
- Hrasky, S., Mason, C., & Wills, D. (2009). The textual complexity of annual report narratives: A comparison of high- and lowperformance companies. *New Zealand Journal of Applied Business Research*, 7(2), 31–45.
- Hsiao, C. (2003). *Analysis of Panel Data, Third Edition* (Cambridge University Press, Ed.; 2. ed.).
- Huang, X., Krishnan, S., & Lin, P. (2016). Readability of compensation discussion and analysis: The impact of the Plain English initiative. *International Journal of Services and Standards*, 11(1), 20–42. <https://doi.org/10.1504/IJSS.2016.076969>
- Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone Management. *The Accounting Review*, 89(3), 1083–1113. <https://doi.org/10.2308/accr-50684>
- Hussainey, K., Albitar, K., & Alkaraan, F. (2022). Corporate narrative reporting on Industry 4.0 technologies: does governance matter? *International Journal of*

- Accounting and Information Management*, 30(4), 457–476.
<https://doi.org/10.1108/IJAIM-02-2022-0024>
- Ishak, R., Amran, N. A., & Manaf, K. B. A. (2018). Firm characteristics and financial reporting quality: The moderating role of Malaysian corporate governance index. *Journal of Social Sciences Research, Special Iss*, 924–932.
<https://doi.org/10.32861/jssr.spi6.924.932>
- Ivashkovskaya, I., & Adamu, M. U. (2021). Corporate Governance and Risk Disclosure in Emerging Countries. *Journal of Corporate Finance Research / Корпоративные Финансы | ISSN: 2073-0438*, 15(4), 5–17.
<https://doi.org/10.17323/j.jcfr.2073-0438.15.4.2021.5-17>
- Jayasree, M., & Shette, R. (2021). Readability of Annual Reports and Operating Performance of Indian Banking Companies. *IIM Kozhikode Society and Management Review*, 10(1), 20–30. <https://doi.org/10.1177/2277975220941946>
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure. *Journal of Financial Economics*, 3, 305–360.
- Kang, T., Park, D.-H., & Han, I. (2018). Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics. *Telematics and Informatics*, 35(2), 370–381. <https://doi.org/10.1016/j.tele.2017.12.014>
- Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55(12), 2049–2075. <https://doi.org/10.1016/j.infsof.2013.07.010>
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, 10(1), 62–102.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6), 1113–1155.
<https://doi.org/10.1086/250042>
- Lajili, K. (2015). Embedding human capital into governance design: a conceptual framework. *Journal of Management & Governance*, 19(4), 741–762.
<https://doi.org/10.1007/s10997-014-9295-8>
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2–3), 221–247.
<https://doi.org/10.1016/j.jacceco.2008.02.003>
- Li, M., Wu, H., Xiao, M., & You, J. (2019). Beyond cheap talk: management's informative tone in corporate disclosures. *Accounting and Finance*, 59(5), 2905–2959. <https://doi.org/10.1111/acfi.12554>
- Liu, J. (2021). Does negative information in MD&A can reduce stock crash risk? *Nankai Business Review International*, 12(4), 537–552.
<https://doi.org/10.1108/NBRI-04-2021-0027>
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, LXVI(1).
- Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, 69(4), 1643–1671. <https://doi.org/10.1111/jofi.12162>
- Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4), 1187–1230.
<https://doi.org/10.1111/1475-679X.12123>
- Luo, J., Li, X., & Chen, H. (2018). Annual report readability and corporate agency costs. *China Journal of Accounting Research*, 11(3), 187–212.
<https://doi.org/10.1016/j.cjar.2018.04.001>

- Martins, M. E. F., Ghiraldello, C. M., Nunes, M. das G. V., & Oliveira Junior, O. N. de. (1996). Readability formulas applied to textbooks in Brazilian Portuguese. In *Instituto de Ciências Matemáticas e de Computação*.
- Pajuste, A., Poriente, E., & Novickis, R. (2021). Management reporting complexity and earnings management: evidence from the Baltic markets. *Baltic Journal of Management*, 16(1), 47–69. <https://doi.org/10.1108/BJM-01-2020-0019>
- Pashalian, S., & Crissy, W. J. E. (1950). How readable are corporate annual reports? *Journal of Applied Psychology*, 34(4), 244–248. <https://doi.org/10.1037/h0061918>
- Pouryousof, A., Nassirzadeh, F., & Askarany, D. (2023). Inconsistency in Managers' Disclosure Tone: The Signalling Perspective. *Risks*, 11(12), 205. <https://doi.org/10.3390/risks11120205>
- Pouryousof, A., Nassirzadeh, F., Hesarzadeh, R., & Askarany, D. (2022). The Relationship between Managers' Disclosure Tone and the Trading Volume of Investors. *Journal of Risk and Financial Management*, 15(12), 618. <https://doi.org/10.3390/jrfm15120618>
- Rajabalizadeh, J. (2023). CEO overconfidence and financial reporting complexity: evidence from textual analysis. *Management Decision*, 61(13), 356–385. <https://doi.org/10.1108/MD-06-2023-1033>
- Reichmann, D. (2023). Tone management and stock price crash risk. *Journal of Accounting and Public Policy*, 42(6). <https://doi.org/10.1016/j.jaccpubpol.2023.107155>
- Resolução CVM Nº 59, de 22 de Dezembro de 2021 (2021). <https://conteudo.cvm.gov.br/legislacao/resolucoes/resol059.html>
- Resolução CVM Nº 87, de 31 de Março de 2022 (2022). <https://conteudo.cvm.gov.br/legislacao/resolucoes/resol087.html>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (1976). Biometrika Trust Inference and Missing Data. *Biometrika*, 63(3), 581–592. <https://about.jstor.org/terms>
- Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: gap-spotting or problematization? *Organization*, 18(1), 23–44. <https://doi.org/10.1177/1350508410372151>
- Saravanan, R., Mohammad, F., & Kumar, P. (2024). Does IFRS convergence affect the readability of annual reports by Indian listed companies? *Journal of Applied Accounting Research*, 25(3), 547–569. <https://doi.org/10.1108/JAAR-10-2022-0284>
- Sayari, N., & Marcum, B. (2018). Reducing risk in the emerging markets: Does enhancing corporate governance work? *BRQ Business Research Quarterly*, 21(2), 124–139. <https://doi.org/10.1016/j.brq.2018.01.002>
- Shleifer, A., & Vishny, R. W. (1997). A Survey of Corporate Governance. *The Journal of Finance*, LII(2).
- Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355. <https://doi.org/10.2307/1882010>
- Stulz, R. M. (2022). Globalization, Corporate Finance, and the Cost of Capital. *Journal of Applied Corporate Finance*, 34(1), 8–23. <https://doi.org/10.1111/jacf.12484>

- Tahat, E. A., Al-Mawali, H. H., & Tahat, Y. A. (2021). The value relevance of the financial statements' information: The moderating role of the board of directors. *Jordan Journal of Business Administration*, 17(2), 177–193.
- Tan, H., Wang, E. Y., & Zhou, B. (2014). When the Use of Positive Language Backfires: The Joint Effect of Tone, Readability, and Investor Sophistication on Earnings Judgments. *Journal of Accounting Research*, 52(1), 273–302.
<https://doi.org/10.1111/1475-679X.12039>
- Toerien, F. E., & du Toit, E. (2024). Fighting through the Flesch and Fog: the readability of risk disclosures. *Accounting Research Journal*, 37(1), 39–56.
<https://doi.org/10.1108/ARJ-03-2023-0094>
- Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32, 97–180.
- Waheed, A., & Malik, Q. A. (2019). Board characteristics, ownership concentration and firms' performance. *South Asian Journal of Business Studies*, 8(2), 146–165. <https://doi.org/10.1108/SAJBS-03-2018-0031>
- Wang, C., Xin, J., Sun, F., Shi, Y., & Du, Y. (2024). The effects of manager sentiment in financial disclosure: Perspectives of operational efficiency and market reaction. *Finance Research Letters*, 64, 105425.
<https://doi.org/10.1016/j.frl.2024.105425>
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Ed.; 2. ed.).
- Yusuf, J.-E. W., & Jordan, M. M. (2017). Accessibility of the Management's Discussion and Analysis to Citizen Users of Government Financial Reports. *Public Budgeting and Finance*, 37(4), 74–91. <https://doi.org/10.1111/pbaf.12170>
- Zeng, Y., Chan, K. C., Chen, Y., & Li, L. (2023). Do analysts pay attention to managerial sentiment? Evidence from analysts' following decisions and earnings forecasts. *Borsa Istanbul Review*, 23(2), 269–284.
<https://doi.org/10.1016/j.bir.2022.10.008>

Apêndice A – Protocolo de Revisão Sistemática da Literatura (RSL)

Este protocolo foi elaborado com base nas diretrizes de Kitchenham e Brereton (2013) Kitchenham e Brereton (2013), com o objetivo de assegurar rigor metodológico, reproduzibilidade e transparência na condução da revisão sistemática da literatura. A revisão teve como foco a identificação de estudos que investigassem a relação entre governança corporativa, legibilidade, tom e incerteza textual nos relatórios da administração de empresas brasileiras, especialmente no contexto da listagem de *American Depository Receipts* (ADRs). A revisão seguiu cinco estágios principais: (i) definição do tema de pesquisa; (ii) identificação dos constructos e palavras-chave; (iii) busca e seleção manual; (iv) processo de triagem; e (v) extração, análise e síntese dos dados.

A busca foi realizada no Portal de Periódicos da CAPES, com acesso às bases Scopus e Web of Science (WoS). Ressalta-se que alguns artigos só são acessíveis por meio da rede institucional da Universidade Federal da Paraíba (UFPB). Sendo assim, utilizou-se o navegador Firefox UFPB, que permite autenticação remota e garante acesso completo a determinados periódicos disponíveis. Pesquisadores vinculados a outras instituições podem precisar de soluções equivalentes para acessar os mesmos materiais.

No segundo estágio, definiu-se os constructos da pesquisa e os respectivos termos de busca. O Quadro XXXX apresenta os constructos e respectivas as palavras-chave para busca bibliográfica.

Quadro 2

Constructos e palavras-chave para busca bibliográfica.

Constructo	Palavras-chave
Governança Corporativa	"corporate governance" OR "board independence" OR "audit committee" OR "ownership structure" OR "governance mechanisms" OR "investor protection" OR "american depositary receipts" OR ADR OR "cross-listing" OR "foreign listing" OR "listing abroad"
Características Textuais dos Relatórios	"readability" OR "text complexity" OR "length" OR "common words" OR "financial words" OR "linguistic complexity" OR "textual analysis" OR "textual obfuscation" OR "flesch" OR "flesch-kincaid" OR "plain english" OR "tone" OR "managerial tone" OR "sentiment analysis" OR "disclosure sentiment" OR "positive tone" OR "negative tone" OR "annual report" OR "management report" OR "MD&A" OR "management's discussion and analysis" OR "narrative disclosure" OR "financial disclosure"
Incerteza	"uncertainty" OR "textual uncertainty" OR "textual ambiguity" OR "disclosure uncertainty" OR "financial uncertainty" OR "policy uncertainty" OR EPU

Fonte: Elaboração Própria (2025).

As combinações de palavras-chave foram aplicadas nas bases selecionadas com a utilização de operadores booleanos (AND, OR), adaptadas à sintaxe de cada plataforma. A formulação das *strings* buscou refletir os constructos diretamente relacionados às hipóteses da pesquisa, assegurando um equilíbrio entre abrangência e especificidade. O Quadro XXX apresenta as *strings* utilizadas nas bases Scopus e Web of Science, bem como a quantidade de artigos recuperados em cada uma delas.

Quadro 3

Strings de busca por base prioritária e relação de busca.

Base	Relação de busca	Quantidade
Scopus	<p>TITLE-ABS-KEY(</p> <p>("corporate governance" OR "board independence" OR "audit committee" OR "ownership structure" OR "governance mechanisms" OR "investor protection" OR "american depositary receipts" OR ADR OR "cross-listing" OR "foreign listing" OR "listing abroad")</p> <p>AND</p> <p>("readability" OR "text complexity" OR "length" OR "common words" OR "financial words" OR "linguistic complexity" OR "textual analysis" OR "textual obfuscation" OR "flesch" OR "flesch-kincaid" OR "plain english" OR "tone" OR "managerial tone" OR "sentiment analysis" OR "disclosure sentiment" OR "positive tone" OR "negative tone" OR "annual report" OR "management report" OR "MD&A" OR "management's discussion and analysis" OR "narrative disclosure" OR "financial disclosure")</p> <p>AND</p> <p>("uncertainty" OR "textual uncertainty" OR "textual ambiguity" OR "disclosure uncertainty" OR "financial uncertainty" OR "policy uncertainty" OR EPU)</p> <p>)</p>	34
Wos	<p>TS=(</p> <p>("corporate governance" OR "board independence" OR "audit committee" OR "ownership structure" OR "governance mechanisms" OR "investor protection" OR "american depositary receipts" OR "ADR" OR "cross-listing" OR "foreign listing" OR "listing abroad")</p> <p>AND</p> <p>("readability" OR "text complexity" OR "length" OR "common words" OR "financial words" OR "linguistic complexity" OR "textual analysis"</p>	50

Quadro 3

Strings de busca por base prioritária e relação de busca.

Base	Relação de busca	Quantidade
	OR "textual obfuscation" OR "flesch" OR "flesch-kincaid" OR "plain english" OR "tone" OR "managerial tone" OR "sentiment analysis" OR "disclosure sentiment" OR "positive tone" OR "negative tone" OR "annual report" OR "management report" OR "MD&A" OR "management's discussion and analysis" OR "narrative disclosure" OR "financial disclosure") AND ("uncertainty" OR "textual uncertainty" OR "textual ambiguity" OR "disclosure uncertainty" OR "financial uncertainty" OR "policy uncertainty" OR "EPU"))	

Fonte: Elaboração Própria (2025).

No quarto estágio, realizou-se o processo de triagem. Os títulos e resumos dos estudos identificados foram analisados com base nos seguintes critérios de inclusão: (a) disponibilidade integral nas bases selecionadas; (b) idioma em português, inglês e espanhol; (c) abordagem teórica ou empírica de pelo menos um dos seguintes temas centrais, legibilidade, tom, incerteza textual ou governança corporativa; e (d) análise de documentos corporativos, como relatório da administração, relatório anual ou seção MD&A. Foram excluídos artigos indisponíveis na íntegra, redigidos em outros idiomas, sem relação com os eixos temáticos definidos ou identificados como duplicatas.

Na sequência, os estudos válidos foram organizados em planilha, contendo dados bibliográficos como autores, título, periódico, ano, volume, número, páginas, DOI, *link* de acesso, idioma, tipo de documento, base de origem e classificação no *Scimago*. Procedeu-se, então, à leitura integral dos artigos e ao fichamento estruturado, que reuniu informações sobre o tema principal, objetivos, metodologia, amostra, resultados e conclusões gerais. Essa sistematização buscou garantir padronização, rastreabilidade e integridade no processo de extração.

No estágio final, os estudos incluídos foram analisados de forma integrada, e os principais achados serviram de base para a construção do referencial teórico da tese. O Quadro 4, apresenta o número de artigos identificados em cada base, os

selecionados após a leitura dos resumos, os analisados integralmente, os duplicados e os efetivamente incluídos.

Quadro 4

Resumo da RSL.

Base de Dados	Total de artigos	Duplicados	Selecionados após leitura do resumo	Artigos selecionados
Scopus	34	-	16	10
WoS	50	17	19	8
Total	84	17	35	18

Fonte: Elaboração Própria (2025).

Apêndice B – Script R Extração do RA's e Conversão “.pdf” para “.txt”.

O *script* abaixo foi desenvolvido para automatizar a extração do texto referente ao "Relatório da Administração" de arquivos PDF e salvá-lo em arquivos .txt separados.

```
# Instalar o pacote necessário (execute apenas uma vez)
install.packages("pdftools")

# Carregar o pacote necessário
library(pdftools)

# Remoção de Títulos e Rodapés
remove_headers_and_footers <- function(text) {
  header_footer_pattern <- "DFP - Demonstrações Financeiras
Padronizadas|Versão : \d+|PÁGINA: \d+ de \d+"
  cleaned_text <- gsub(header_footer_pattern, "", text)
  cleaned_text <- gsub("\s+", " ", cleaned_text) # Remove múltiplos
espaços extras
  return(cleaned_text)
}

# Extrair Texto do "Relatório da Administração"
extract_text_from_pdf <- function(pdf_path) {
  all_text <- pdftools::pdf_text(pdf_path)

  # Verificar se as primeiras duas páginas são índices
  index_page_count <- ifelse(grepl("PÁGINA: \d+ de \d+", all_text[1]) ||
grepl("PÁGINA: \d+ de \d+", all_text[2]), 1, 2)

  # Padrões para encontrar as páginas no índice
  pattern_start <- "Relatório da Administração/Comentário do
Desempenho|Relatório da Administração|Comentário do Desempenho"
  pattern_end <- "Notas Explicativas"

  # Encontrar a página inicial e final
  index_text <- paste(all_text[1:index_page_count], collapse = "\n")
  start_match <- regexpr(pattern_start, index_text)
  if (start_match == -1) {
    stop("Não foi possível encontrar a seção desejada no índice do arquivo:
", basename(pdf_path))
  }
  start_page_number <- as.numeric(sub(".?(\d+).?", "\1",
substring(index_text, start_match))) + index_page_count

  end_match <- regexpr(pattern_end, index_text)
  if (end_match == -1) {
    end_page_number <- length(all_text) # Se não encontrar o padrão, usar
o final do documento
  } else {
    end_page_number <- as.numeric(sub(".?(\d+).?", "\1",
substring(index_text, end_match))) + index_page_count - 1
  }

  # Verificação se os números de página são válidos
  if (start_page_number > end_page_number || start_page_number >
length(all_text)) {
```

```

    stop("Erro na identificação das páginas. Verifique o índice do arquivo:
", basename(pdf_path))
}

# Extrair o texto das páginas relevantes
extracted_text <- paste(all_text[start_page_number:end_page_number],
collapse = "\n")

# Dividir por linhas e remover cabeçalhos e rodapés
extracted_text_lines <- unlist(strsplit(extracted_text, "\n"))
cleaned_text <- remove_headers_and_footers(extracted_text_lines)

if (nchar(paste(cleaned_text, collapse = "\n")) == 0) {
  stop("Não foi possível extrair texto limpo do arquivo: ",
  basename(pdf_path))
}

return(paste(cleaned_text, collapse = "\n"))
}

# Salvar Texto em Arquivo TXT
save_text_to_file <- function(text, file_path) {
  output_file <- sub("\\.pdf$", ".txt", file_path)
  writeLines(text, output_file)
  return(output_file)
}

# Definir diretórios (Substitua conforme necessário)
pdf_directory <- "Caminho/DO/DIRETÓRIO/COM/PDFS/"
output_directory <- "Caminho/DO/DIRETÓRIO/DE/SAÍDA/"

# Listar arquivos PDF
pdf_files <- list.files(pdf_directory, pattern = "\\*.pdf$", full.names =
TRUE)

# Loop através de cada arquivo PDF
for (pdf_file in pdf_files) {
  tryCatch({
    # Extrair o texto com base nas páginas identificadas
    extracted_text <- suppressWarnings(extract_text_from_pdf(pdf_file))

    if (nchar(extracted_text) == 0) {
      next
    }

    # Salvar o texto limpo no diretório especificado
    txt_file <- save_text_to_file(extracted_text, pdf_file, save_directory)

    }, error = function(e) {
      next
    })
}
}

```

Apêndice C – Script R para calcular características textuais – Linguagem R

O *script* abaixo foi desenvolvido para automatizar o cálculo das características textuais dos "Relatório da Administração".

```
# URLs dos arquivos online
url_relatorio <- "https://raw.githubusercontent.com/castroufpb/Relat-rio-Modelo/main/Relat%C3%B3rio%20TESTE.txt"
url_positivas <-
"https://raw.githubusercontent.com/castroufpb/positivas/main/positivas.csv"
url_negativas <-
"https://raw.githubusercontent.com/castroufpb/negativas/main/Negativas.csv"
url_incerteza <-
"https://raw.githubusercontent.com/castroufpb/incerteza/main/incerteza_PT.csv"

# Instalar pacotes necessários
if (!require(stringr)) install.packages("stringr", dependencies = TRUE)
if (!require(tokenizers)) install.packages("tokenizers", dependencies = TRUE)
if (!require(openxlsx)) install.packages("openxlsx", dependencies = TRUE)
if (!require(SnowballC)) install.packages("SnowballC", dependencies = TRUE)
if (!require(tm)) install.packages("tm", dependencies = TRUE)
if (!require(dplyr)) install.packages("dplyr", dependencies = TRUE)
if (!require(httr)) install.packages("httr", dependencies = TRUE)

# Carregar pacotes necessários
library(stringr)
library(tokenizers)
library(openxlsx)
library(SnowballC)
library(tm)
library(dplyr)
library(httr)

# Função para ler CSV online
ler_csv_online <- function(url) {
  temp <- tempfile()
  download.file(url, temp, quiet = TRUE)
  dados <- read.csv(temp, header = FALSE, stringsAsFactors = FALSE, sep =
"\n", fileEncoding = "UTF-8")
  unlink(temp)
  return(dados)
}

# Carregar dicionários online
dicionarios <- list(
  positivas = ler_csv_online(url_positivas),
  negativas = ler_csv_online(url_negativas),
  lm = ler_csv_online(url_incerteza)
)

# Função para processar dicionários com stemming e sem duplicatas
processar_dicionario <- function(dados) {
  palavras <- dados %>% pull(1) %>% str_trim() %>% tolower()
  palavras_stem <- wordStem(palavras, language = "portuguese")
  data.frame(Original = palavras, Radical = palavras_stem, stringsAsFactors
= FALSE) %>%
  distinct(Radical, .keep_all = TRUE)
```

```

}

# Processar todos os dicionários
dicionarios <- lapply(dicionarios, processar_dicionario)

# Lista de abreviações
abreviacoes <- c(
  "a\\.a\\.", "n°", "CVM", "ibovespa", "BOVESPA", "Selic", "SELIC",
  "TI", "TV", "PIB", "IPCA", "IGPM", "EBITDA", "BM&FBOVESPA",
  "PROSOFT", "IDC", "ERP", "ERM", "SPED", "NF-e", "Co\\.", "S\\.A\\.",
  "S/A", "vs\\.", "V\\.sas\\.", "CLIA", "CPC", "CPCs", "IFRIC", "ABECS"
)

# Função para limpar cabeçalhos, rodapés e linhas vazias
limpar_cabecalho_rodape <- function(texto) {
  # Remove linhas com padrões de cabeçalhos e rodapés
  texto <- texto[!grepl("^- \\d{2}/\\d{2}/\\d{4} - ", texto)] # Cabeçalho com data
  texto <- texto[!grepl("^DFP", texto)] # Cabeçalho DFP
  texto <- texto[!grepl("^PÁGINA:\\s*\\d+\\s*(de)?\\s*\\d*", texto,
  ignore.case = TRUE)] # Rodapé de página
  texto <- texto[!grepl("PÁGINA\\s*:\\s*\\d+", texto, ignore.case = TRUE)]
  # Rodapé com "PÁGINA"
  texto <- texto[!grepl("Relatório da Administração/Comentário do Desempenho", texto, ignore.case = TRUE)] # Seções específicas

  # Remove linhas vazias ou espaços extras
  texto <- texto[texto != ""]
  texto <- trimws(texto)
  return(texto)
}

# Função para substituir abreviações
tratar_abreviacoes <- function(texto) {
  for (abv in abreviacoes) texto <- gsub(abv, "[ABV]", texto)
  texto
}

# Função para gerar n-gramas (bigrams e trigrams)
gerar_ngramas <- function(tokens, n) {
  lapply(seq_along(tokens), function(i) {
    if (i + n - 1 <= length(tokens)) {
      paste(tokens[i:(i + n - 1)], collapse = " ")
    } else {
      NA
    }
  }) %>% unlist() %>% na.omit()
}

# Função para buscar palavras-chave com contexto
encontrar_palavras_com_contexto <- function(frases, dicionario, categoria) {
  resultados <- lapply(frases, function(frase) {
    tokens <- unlist(str_split(frase, "\\s+"))
    tokens <- tokens[!tokens %in% stopwords("pt")]
    tokens_stem <- wordStem(tokens, language = "portuguese")

    unigrams <- tokens_stem
    bigrams <- gerar_ngramas(tokens_stem, 2)
    trigrams <- gerar_ngramas(tokens_stem, 3)
    todos_ngramas <- c(unigrams, bigrams, trigrams)
  })
}

# Função para gerar n-gramas (bigrams e trigrams)
gerar_ngramas <- function(tokens, n) {
  lapply(seq_along(tokens), function(i) {
    if (i + n - 1 <= length(tokens)) {
      paste(tokens[i:(i + n - 1)], collapse = " ")
    } else {
      NA
    }
  }) %>% unlist() %>% na.omit()
}

```

```

    palavras_encontradas <- todos_ngramas[todos_ngramas %in%
dicionario$Radical]

    if (length(palavras_encontradas) > 0) {
      data.frame(
        Frase = frase,
        Palavra = palavras_encontradas,
        Contexto = paste(tokens, collapse = " "),
        stringsAsFactors = FALSE
      )
    } else {
      NULL
    }
  }

  do.call(rbind, resultados)
}

# Função corrigida para processar relatórios online
processar_relatorio_online <- function(url) {
  tryCatch({
    # Baixar e ler o relatório com tratamento de encoding
    conteudo <- readLines(url, encoding = "UTF-8", warn = FALSE)

    # Verificar se o conteúdo foi lido corretamente
    if(length(conteudo) == 0) stop("Arquivo vazio ou não pôde ser lido")

    # Limpar conteúdo
    conteudo <- limpar_cabecalho_rodape(conteudo)
    texto_unificado <- paste(conteudo, collapse = " ")

    # Verificar se há conteúdo válido após limpeza
    if(nchar(texto_unificado) == 0) stop("Nenhum conteúdo válido após
limpeza")

    # Contar palavras (método mais robusto)
    palavras <- unlist(strsplit(texto_unificado, "\\s+"))
    palavras <- palavras[palavras != ""]
    total_palavras <- length(palavras)

    # Verificar se há palavras para análise
    if(total_palavras == 0) stop("Nenhuma palavra encontrada para análise")

    # Tokenização e stemming
    tokens_limpos <- palavras[!palavras %in% stopwords("pt")]
    tokens_stem <- wordStem(tokens_limpos, language = "portuguese")

    # Contagem de palavras nos dicionários (com verificação)
    palavras_encontradas <- sapply(dicionarios, function(dic) {
      if(nrow(dic) == 0) return(0) # Proteção contra dicionários vazios
      sum(tokens_stem %in% dic$Radical)
    })

    # Criar resultados com verificação de divisão por zero
    resultados <- data.frame(
      Arquivo = basename(url),
      TotalPalavras = total_palavras,
      Positivas = palavras_encontradas["positivas"],
      Negativas = palavras_encontradas["negativas"],
      LM = palavras_encontradas["lm"],

```

```

    Tom = ifelse(total_palavras > 0,
                 round((palavras_encontradas["positivas"] -
palavras_encontradas["negativas"]) / total_palavras, 4),
                 0),
    EPULM = ifelse(total_palavras > 0, round(palavras_encontradas["lm"] /
total_palavras, 4), 0),
    stringsAsFactors = FALSE
  )

  return(resultados)
}, error = function(e) {
  warning("Erro ao processar ", url, ":", e$message)
  return(data.frame(
    Arquivo = basename(url),
    TotalPalavras = NA,
    Positivas = NA,
    Negativas = NA,
    Baker = NA,
    Litigio = NA,
    LM = NA,
    Financas = NA,
    CommonWords = NA,
    Tom = NA,
    EPUBk = NA,
    EPULit = NA,
    EPULM = NA,
    FinancialIndex = NA,
    CommonIndex = NA,
    stringsAsFactors = FALSE
  ))
})
}
}

# Exemplo de uso com o relatório teste
resultado <- processar_relatorio_online(url_relatorio)
print(resultado)

# Salvar resultados em CSV (opcional)
write.csv(resultado, "resultado_analise_sentimento.csv", row.names = FALSE)

```

Apêndice D – Script em R para PSM e DiD

O *script* abaixo foi desenvolvido para aplicar as etapas principais da estratégia empírica da tese, combinando PSM e DiD. Inicialmente, o código realiza o pré-processamento da base de dados, define os grupos de tratamento e controle (empresas com e sem ADRs a partir de 2011) e aplica o PSM com pareamento ótimo para controlar o viés de seleção.

Em seguida, verifica o balanceamento entre os grupos, estima os escores de propensão e gera gráficos comparativos antes e depois do pareamento. Também são incluídos testes de robustez. Por fim, o *script* estima os modelos econôméticos principais (H1, H2 e H3), incluindo o modelo DiD.

```
# =====
# PSM + Event Study - Tese
# =====

rm(list = ls()); gc()

if(!require(pacman)) install.packages('pacman')
pacman::p_load(dplyr, tidyr, readxl, ggplot2, fixest, MatchIt, cobalt, broom, patchwork)

# -----
# 1) Importação e pré-processamento
# -----
setwd("informar o caminho do diretório")
dados <- read_excel("informar o nome do arquivo") %>%
  mutate(across(c(
    ID, Idade, AnoTrat, AnoDR, anopre, Ano, TAM, ALV, ROA, RET, VOLNEG, VOLPRC,
    Sharpe, PU, LnQtdAções, LnVM, VMP, LIQCX, TANGAT, POLDIV, FCO, CRESCREC,
    GovPillarScore, ESGScore, ESGControversies, BidAsk, BoardSize, FemaleBoard,
    FleschPT, Length, Tom, TomNeg, EPU, Incerteza2, Incerteza, FinancialWords,
    CommonWords, Crise, CRESCPIB, IIEBr, MídiaIIEBr, EPUBkBr, ICC, Legibilidade
  ), suppressWarnings(as.numeric))),
  adrs = ifelse(ADRs == "Sim", 1L, 0L),
  AnoDR = tidyr::replace_na(AnoDR, 0),
  anopre = tidyr::replace_na(anopre, 0),
  Setor = as.factor(SetorEconomatica)
)

# Filtra empresas tratadas após 2010
ids_remove <- dados %>% filter(AnoDR > 0 & AnoDR < 2010) %>% pull(ID)
dados <- dados %>% filter(!ID %in% ids_remove)

# -----
# 2) Contagem de empresas tratadas e não tratadas
# -----
contagem_tratadas <- dados %>%
  filter(AnoDR > 0) %>%
  distinct(ID, .keep_all = TRUE) %>%
  group_by(AnoDR) %>%
  summarise(Numero_de_Empresas = n()) %>%
  arrange(AnoDR)

print(contagem_tratadas)

contagem_nao_tratadas <- dados %>%
  filter(AnoDR == 0) %>%
  summarise(Total_Nao_Tratadas = n_distinct(ID))

print(contagem_nao_tratadas)
```

```

# Contagem de empresas e observações por grupo
contagem_atualizada <- dados %>%
  mutate(grupo = ifelse(AnoDR > 0, "Tratadas", "Não Tratadas")) %>%
  group_by(grupo) %>%
  summarise(
    Quantidade_Empresas = n_distinct(ID),
    Quantidade_Observacoes = n(),
    .groups = "drop"
  )

# Adicionar o total
contagem_total <- contagem_atualizada %>%
  summarise(
    grupo = "Total",
    Quantidade_Empresas = sum(Quantidade_Empresas),
    Quantidade_Observacoes = sum(Quantidade_Observacoes)
  )

# -----
# 3) PSM
# -----
psm_vars <- c("Legibilidade", "Length", "TomNeg", "CommonWords",
             "FinancialWords", "TAM", "Idade", "TANGAT")

dados <- dados %>%
  mutate(across(all_of(psm_vars), ~ifelse(is.na(.), median(., na.rm = TRUE), .)))

matching_model <- matchit(
  adrs ~ Legibilidade + Length + TomNeg + CommonWords + FinancialWords +
    TAM + Idade + TANGAT,
  data = dados,
  method = "optimal",
  distance = "logit"
)

# Base pareada
matched_data <- match.data(matching_model)

# Gráfico de balanceamento
love.plot(matching_model, binary = "std", threshold = 0.10)

# Propensity scores antes
propensity_model_before <- glm(
  adrs ~ Legibilidade + Length + TomNeg + CommonWords + FinancialWords +
    TAM + Idade + TANGAT + factor(SetorEconomatica),
  data = dados,
  family = binomial(link = "logit")
)
dados$prop_score <- predict(propensity_model_before, type = "response")

plot_before <- ggplot(dados, aes(x = prop_score, fill = factor(adrs))) +
  geom_density(alpha = 0.6) +
  xlim(0, 1) +
  labs(title = "Antes do Pareamento", x = "Propensity Score", y = "Densidade") +
  scale_fill_manual(values = c("0" = "blue", "1" = "red"),
                    labels = c("0" = "Controle", "1" = "Tratado")) +
  theme_minimal()

# Propensity scores depois
plot_after <- ggplot(matched_data, aes(x = distance, fill = factor(adrs))) +
  geom_density(alpha = 0.6) +
  xlim(0, 1) +
  labs(title = "Depois do Pareamento", x = "Propensity Score", y = "Densidade") +
  scale_fill_manual(values = c("0" = "blue", "1" = "red"),
                    labels = c("0" = "Controle", "1" = "Tratado")) +
  theme_minimal()

plot_before + plot_after

# Verificar pareamento
summary(matching_model)

# -----
# 4) MODELOS H1 e H2
# -----
# Modelo H1 - Legibilidade e Incerteza
Model_H1 <- feols(Incerteza ~ Legibilidade + adrs + Idade + Length + CommonWords +

```

```

        FinancialWords + TAM + TANGAT + VMP + EPU + ICC + Crise | Setor,
        data = matched_data, cluster = ~Ano)

summary(Model_H1)

# Extrair coeficientes com IC e p-valores
h1_results <- tidy(Model_H1, conf.int = TRUE)

# Salvar no Excel
write.xlsx(h1_results,
           file = "resultado_H1.xlsx",
           sheetName = "Modelo_H1",
           rowNames = FALSE,
           overwrite = TRUE)

# Modelo H2 - Tom Negativo (via variável TOM) e Incerteza
Model_H2 <- feols(Incerteza ~ TomNeg + Legibilidade + Legibilidade*TomNeg + adrs + Idade +
Length + CommonWords +
        FinancialWords + TAM + TANGAT + VMP + EPU + Crise | Setor,
        data = matched_data, cluster = ~Ano)
summary(Model_H2)

# Extrair coeficientes com IC e p-valores
h2_results <- tidy(Model_H2, conf.int = TRUE)

# Salvar no Excel
write.xlsx(h2_results,
           file = "resultado_H2.xlsx",
           sheetName = "Modelo_H2",
           rowNames = FALSE,
           overwrite = TRUE)

# -----
# 5) Função para Event Study
# -----
run_event_study <- function(data, gvar, yvar, covs, leads = NULL, lags = NULL) {

  # Determinar leads e lags automaticamente, se não fornecidos
  if (is.null(leads) | is.null(lags)) {
    # Calcular o tempo relativo para todos os tratados
    df_temp <- data %>% filter (!!sym(gvar) > 0) %>%
      mutate(tempo_relativo = Ano - !!sym(gvar))

    # Lead = valor absoluto mínimo (anos antes)
    min_rel <- min(df_temp$tempo_relativo, na.rm = TRUE)
    leads <- ifelse(is.null(leads), abs(min_rel), leads)

    # Lag = valor máximo (anos depois)
    max_rel <- max(df_temp$tempo_relativo, na.rm = TRUE)
    lags <- ifelse(is.null(lags), max_rel, lags)
  }

  df <- data %>%
    filter (!!sym(gvar) > 0) %>%
    mutate(tempo_relativo = Ano - !!sym(gvar)) %>%
    filter(tempo_relativo >= -leads & tempo_relativo <= lags)

  fmla <- as.formula(
    paste0(yvar, " ~ i(tempo_relativo, ref = -1) + ", paste(covs, collapse = " + ")))
  )

  est <- feols(fmla, data = df, fixef = c("ID", "Ano"), cluster = "ID")

  plot_df <- broom::tidy(est, conf.int = TRUE) %>%
    filter(grepl("tempo_relativo:", term)) %>%
    mutate(
      leadlag = as.numeric(gsub("tempo_relativo:", "", term)),
      periodo = ifelse(leadlag < 0, as.character(leadlag), paste0("+", leadlag))
    )

  list(model = est, plot_data = plot_df, leads = leads, lags = lags)
}

# -----
# 6) Rodar Event Study
# -----

```

```

covs_did <- c("Idade", "TomNeg", "Length", "CommonWords", "Legibilidade",
            "FinancialWords", "TAM", "TANGAT", "VMP", "EPU", "Crise", "ICC")

res_anopre <- run_event_study(matched_data, "anopre", "Incerteza", covs_did, leads = 14, lags
= 14)
res_AnoDR <- run_event_study(matched_data, "AnoDR", "Incerteza", covs_did, leads = 14, lags
= 14)

etable(res_anopre$model, res_AnoDR$model, se.below = TRUE)

# Extrair resultados detalhados com p-valor
anopre_results <- tidy(res_anopre$model, conf.int = TRUE)
AnoDR_results <- tidy(res_AnoDR$model, conf.int = TRUE)

# Criar workbook e exportar
wb <- createWorkbook()
addWorksheet(wb, "EventStudy_anopre")
writeData(wb, "EventStudy_anopre", anopre_results)

addWorksheet(wb, "EventStudy_AnoDR")
writeData(wb, "EventStudy_AnoDR", AnoDR_results)

# Salvar arquivo
saveWorkbook(wb, file = "resultados_event_study.xlsx", overwrite = TRUE)

# Visualizar rapidamente no console
print(anopre_results %>% select(term, estimate, std.error, statistic, p.value))
print(AnoDR_results %>% select(term, estimate, std.error, statistic, p.value))

# -----
# 6) Gráficos Event Study
# -----
p1 <- ggplot(res_anopre$plot_data, aes(x = leadlag, y = estimate)) +
  geom_point() +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
  geom_hline(yintercept = 0, linetype = "dotted") +
  labs(title = "Event Study - Ano do Anúncio (anopre)",
       x = "Período Relativo", y = "Efeito sobre Incerteza") +
  theme_minimal()

p2 <- ggplot(res_AnoDR$plot_data, aes(x = leadlag, y = estimate)) +
  geom_point() +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
  geom_hline(yintercept = 0, linetype = "dotted") +
  labs(title = "Event Study - Ano da Listagem (AnoDR)",
       x = "Período Relativo", y = "Efeito sobre Incerteza") +
  theme_minimal()

p1 + p2

# -----
# 7) Verificação visual da suposição de tendências paralelas
# -----
grafico_trends <- matched_data %>%
  group_by(Ano, adrs) %>%
  summarise(media_incerteza = mean(Incerteza, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = Ano, y = media_incerteza, color = as.factor(adrs), group = adrs)) +
  geom_line(size = 1.1) +
  geom_point(size = 2) +
  geom_vline(xintercept = min(matched_data$Ano[matched_data$Ano > 0]),
             linetype = "dashed", color = "black") +
  scale_x_continuous(breaks = seq(min(matched_data$Ano), max(matched_data$Ano), 1)) +
  labs(x = "Ano", y = "Incerteza média",
       color = "Grupo",
       title = "",
       subtitle = "") +
  scale_color_manual(values = c("0" = "blue", "1" = "red"),
                     labels = c("0" = "Controle", "1" = "Tratado")) +
  theme_minimal(base_size = 12) +
  theme(panel.grid.minor = element_blank(),
        plot.title = element_text(face = "bold"),
        legend.position = "bottom")

print(grafico_trends)

```

```

# -----
# 8) Testes de Placebo e de Estresse
# -----
# ---- 8.1) Placebo: datas falsas de adoção (2 anos antes)
set.seed(123)
placebo_data <- matched_data %>%
  mutate(AnoDR_placebo = ifelse(adrs == 1, AnoDR - 2, 0)) %>%
  filter(AnoDR_placebo > 0)

res_placebo <- run_event_study(placebo_data, "AnoDR_placebo", "Incerteza", covs_did, leads =
14, lags = 14)
etable(res_placebo$model, se.below = TRUE)

# ---- 8.2) Estresse: excluir coortes sensíveis (empresas com ADRs após 2021)
stress_data <- matched_data %>%
  filter(!AnoDR >= 2021 & AnoDR > 0))

res_stress <- run_event_study(stress_data, "AnoDR", "Incerteza", covs_did, leads = 14, lags =
14)
etable(res_stress$model, se.below = TRUE)

# ---- 8.3) Janela alternativa para t0 (-5 a +5 anos)
res_janela5 <- run_event_study(matched_data, "AnoDR", "Incerteza", covs_did, leads = 5, lags =
5)
etable(res_janela5$model, se.below = TRUE)

# ---- 8.4) Definições alternativas de variáveis
## Incerteza alternativa: Incerteza2
# 1) Tirar EPU das covariáveis
covs_no_epu <- setdiff(covs_did, "EPU")

# 2) Versão da função sem FE de Ano para este caso específico
run_event_study_no_year_fe <- function(data, gvar, yvar, covs, leads = 14, lags = 14) {
  df <- data %>%
    filter(!!(sym(gvar)) > 0) %>%
    mutate(tempo_relativo = Ano - !!(sym(gvar))) %>%
    filter(tempo_relativo >= -leads & tempo_relativo <= lags)

  fmla <- as.formula(
    paste0(yvar, " ~ i(tempo_relativo, ref = -1) + ", paste(covs, collapse = " + )))
  )

  est <- feols(fmla, data = df, fixef = c("ID"), cluster = "ID") # <-- sem "Ano"

  plot_df <- broom::tidy(est, conf.int = TRUE) %>%
    dplyr::filter(grepl("tempo_relativo:::", term)) %>%
    mutate(
      leadlag = as.numeric(gsub("tempo_relativo:::", "", term)),
      periodo = ifelse(leadlag < 0, as.character(leadlag), paste0("+", leadlag))
    )
  list(model = est, plot_data = plot_df)
}

# 3) Rodar com EPU macro como Y
res_epu_macro <- run_event_study_no_year_fe(
  matched_data,
  gvar = "AnoDR",
  yvar = "EPU",           # índice macro anual
  covs = covs_no_epu,
  leads = 14, lags = 14
)

# 4) Ver no console
etable(res_epu_macro$model, se.below = TRUE)

```

Apêndice E – Lista das Palavras Positivas

abundância; abundante; aclamado; acolhedor; adequadamente; agradável; agradavelmente; alcança; alcançado; alcançando; alcançar; alcançou; aliança; alianças; amigável; aperfeiçoa; aperfeiçoado; apesar de; aprecia; apreciação; apreciando; aprimorado; aproveitar; ascensão; ascensões; assegurando; assegurar; atingir; atraente; atratividade; aumento; avançando; avanço; avanços; bem-sucedido; beneficamente; beneficiado; beneficiando; beneficiar; benéfico; bom; brilhante; capaz; caritativo; colabora; colaboração; colaborações; colaborador; colaboradores; colaborando; colaborar; colaborativo; colaborou; com sucesso; conclusivamente; conclusivo; condutivo; confiabilidade; confiante; confiável; conquista; conquistas; consegue; conseguindo; construtivamente; construtivo; cortês; cortesia; crescendo; crescimento; criativamente; criatividade; criativo; curtiu; de valor; delícias; delicioso; desejado; desejável; destinado; diligente; diligentemente; distinção; distinções; distintamente; distintividade; distintivo; eficaz; eficiência; eficiências; eficiente; eficientemente; elogiado; elogiando; elogio; empodera; empoderado; empoderando; empoderar; empolgante; encantado; encantador; encantadoramente; encoraja; encorajado; encorajador; encorajamento; engenhosidade; entusiasmado; entusiasmo; entusiasticamente; entusiástico; espetacular; espetacularmente; estabilidade; estabiliza; estabilização; estabilizações; estabilizado; estabilizando; estabilizar; estável; estimulante; estreia; excede; excelência; excelente; excepcional; excepcionalmente; excitação; excitante; exclusivamente; exclusividade; exclusivo; exclusivos; exemplar; expansão; fácil; facilmente; fantástico; favorável; favoravelmente; favorecendo; favorito; favoritos; felicidade; feliz; felizmente; força; fortalece; fortalecer; fortalecido; fortalecimento; forte; ganhador; ganhando; ganhar; ganho; ganhos; ganhou; garante; grandeza; habilitado; habilitando; habilitar; honorável; honra; honrado; honrar; honras; honroso; ideal; impressiona; impressionado; impressionante; impressionantemente; impressionar; impulsionado; impulsionar; incomparável; incrível; incrivelmente; influente; informativo; inigualável; inova; inovação; inovações; inovado; inovador; inovadores; inovando; inovar; inspiração; inspirador; insuperável; integridade; invenção; invenções; inventado; inventando; inventar; inventividade; inventivo; inventor; inventores; leal; líder; liderança; lindamente; lindo; lucrativo; maior; mais alto; mais fácil; mais feliz; mais forte; melhor; melhora; melhorando; melhorar melhoria; melhorias; melhorou; meritório; motivado; muito; o melhor; oportunidade; oportunidades; otimista; ótimo; perfeitamente; perfeito; perfeitos; permite; perspicaz; pontos fortes; popular; popularidade; positivamente; positivo; prazer; prazeroso; preeminência; preeminente; prestígio; prestigioso; proativamente; proativo; proficiência; proficiente; proficientemente; progredindo; progrediu; progresso; progressos; progride; propício; próspera; prosperar; prosperidade; próspero; prosperou; realiza; realização; realizações; realizado; realizando; realizar; receptivo; recompensa; recompensado; recompensador; recompensas; recuperação; recuperado; recuperando; recuperando-se; recuperar; reforçado; reforço; reforços; rentabilidade; rentável; rentávelmente; resolve; resolvendo; resolver; revolucionaria; revolucionado; revolucionando;

revolucionar; satisfação; satisfatoriamente; satisfatório; satisfaz; satisfazer; satisfeito; sonhar; suave; suavemente; suaviza; suavização; sucesso; sucessos; supera; superado; superando; superando o desempenho; superar; superar o desempenho; superior; superou o desempenho; ter sucesso; transparência; tremendamente; tremendo; valioso; vantagem; vantagens; vantajosamente; vantajoso; vencedores; vencendo; versátil; versatilidade; vibração; vibrante.

Apêndice F – Listas das Palavras Negativas

abaixo do padrão; abandona; abandonado; abandonando; abandonar; abandono; abandonos; abdica; abdicação; abdicções; abdicado; abdicando; aberração; aberracional; aberrações; aberrante; abole; abolição; abolido; abolir; aborrecimento; aborrecimentos; abruptamente; abrupto; absentismo; absolve; absolvendo; absolver; absolvção; absolvões; absolvido; abusado; abusando; abusivamente; abusividade; abusivo; abuso; abusos; ação penal; accidental; accidentalmente; acidente; acidentes; acusa; acusação; acusações; acusado; acusando; acusar; acusável; adiamento; adiamentos; adiando; adiar; adulteração; adulterações; adulterado; adulterando; adulterar; adversamente; adversarial; adversário; adversários; adversidade; adversidades; adverso; advertido; advertindo; afeta; afirma; afligido; agitação; agrava; agravações; agravado; agravamento; agravante; gravar; agredido; agredindo; agressão; ajuste de relatórios financeiros; alega; alegação; alegações; alegadamente; alegado; alegando; alegar; alerta; alertando; aliena; alienação; alienações; alienado; alienando; alienar; ameaça; ameaçado; ameaçador; ameaçar; ameaças; amortecer; amortecido; anomalamente; anomalia; anomalias; anômalo; anormal; anormalidade; anormalidades; anormalmente; anticompetitivo; antieconômico; antiético; antiético; antitruste; anula; anulação; anulação de julgamento anulações; anulações de julgamento; anulado; anulando; anular; apela; apelo; aperto; aplicação errada; aplicações erradas; apreende; apreendendo; apreender; apreendido; apreensões; apresentando; apropriação indevida; apropriações indébitas; apropriado indevidamente; argumentativo; argumento; argumentos; arrastar; arriscado; artificialmente; aspereza; assaltado; assaltando; assaltos; assediado; assediando; assédio; asserções; assombroso; ataque cibernético; ataques cibernéticos; atenuação; atenuante; atrasado; atrasando; atraso; atrasos; atrito; ausência; ausências; autor; autores; avarias; avisado; avisar; aviso; aviso legal; avisos; baixas; baixo desempenho; barrado; barreira; barreiras; bloqueio; bloqueios; boicotado; boicotando; boicote; boicotes; caducado; caducando; calamidade; calamidades; calamitoso; calcula mal; calculado mal; calculando errado; calcular errado; cálculos errados; calúnia; caluniado; calúnias; cancela; cancelado; cancelamento; cancelamentos; cancelando; cancelar; caro; catástrofe; catástrofes catastroficamente; catastrófico; cautelar; censura; censurado; censurando; censuras; cepas; cessando; cessar; chateado; chato; chocado; cibercriminoso; cibercriminosos; citações; citada; classificação errada; classificações erradas; classificado mal; classificar erradamente; coação; coage; coagido; coagindo; coagir; coercão; coercivo; colapsando; colapso; colapsos; colisão; colisões; com falha; com mal funcionamento; com problemas; cometendo erro; competindo; complica; complicação; complicações; complicado; complicando; complicar; compreender errado; compulsão; concede; concedendo; conceder; concedido; conciliação; conciliações; conciliando; condena; condenação; condenações; condenado; condenando; condenar; confessa; confessado; confessando; confessar; confinado; confinamento; confinamentos; confinando; confiscado; confiscados; confiscando; confiscar; confisco; confissos; confissão; conflitante; conflito; conflitos; conformando; conformar; conformou; confronta; confrontacional; confrontações; confrontado; confrontando; confronto; confunde; confundindo; confundir; confusão; confuso; conluiado; conluio; conscientemente; consequências; conspira; conspiração; conspirações; conspirado; conspirador; conspiradores; conspirando; conspirar; conspiratório; constrange; stranger; constrangimentos; construção excessiva; construções excessivas; construído em excesso; contaminação; contaminado; contenção; contenciosamente; contencioso; contenções; contend; contender; contestadas; contestando; contestou; contornado; contornando; contornar; contorno; contornos; contra; contração; contrações; contradição; contradições; contradisse; contradito; contraditório; contradiz; contradizendo; contradizer; contramedida; contramedidas; contrariamente; contrário; contratempos; contravenção; contravenções; controvérsia; controvérsias; controverso; convicção; convicções; convocação; convocações; convocado; correção; correções; corrigir; corrigindo; corrói; corrompendo; corromper; corrompido; corrupção; corrupções; corruptamente; corrupto; cortado; cortar; corte; corte inferior; crime; crime cibernético; crimes; crimes cibernéticos; crimes graves; criminal; criminalmente; criminosos; crise; crises; critica; crítica; criticado; criticamente; criticando; criticar; críticas; crítico; crucial; crucialmente; cuidado; cuidados; culpabilidade; culpadamente; culpado; culpável; cyberbullying; danificação; danificado; danificados; dano; danos;

danoso; de forma confusa; de forma desmedida; de forma imprudente; de forma inaceitável; decepção; declaração incorreta; declarações erradas; declarações incorretas; declínio; declínios; defeito; defeitos; defeituoso; defende; defendendo; defender; defendido; defensiva; deficiência; deficiências; deficiente; déficit; déficits; defraudado; defraudando; defraudar; defraudes; defunto; degrada; degradação; degradações; degradado; degradando; degradar; deletério; deliberadamente; deliberado; deliberar; delinquência; delinquências; delinquente; delinquentemente; delinquentes; delitos; demasiadamente; demissão; demissões; demite; demitido; demitindo; demitir; demole; demolição; demolições; demolido; demolindo; demolir; denigre; denigrindo; denegrir; denegriu; denigração; depleções; depreciação; depreciações; deprime; deprimente; deprimido; deprimir; derrogatório; derrota; derrotado; derrotando; derrotas; derrubado; derrubando; desacelerações; desacelerando; desacelerar; desaconselhável; desacordo; desacordos; desacredita; desacreditado; desacreditando; desacreditar; desafiado; desafiante; desafio; desafios; desagradável; desajuste; desajustes; desalinhamento; desalinhamentos; desaparece; desaparecendo; desaparecer; desaparecido; desaparecimento; desaparecimentos; desaponta; desapontado; desapontadoramente; desapontamento; desapontamentos; desapontante; desapontar; desapropria; desapropriação; desapropriações; desapropriado; desapropriando; desapropriar; desaprova; desaprovação; desaprovações; desaprovando; desaprovar; desassocia; desassociando; desastre; desastres; desastrosamente; desastroso; desatenção; desativa; desativado; desativando; desatualizado; descaracterização; descargas; descarte; descoberto; descobre; descobrindo; descobrir; desconhecendo; desconhecido; desconsiderações; desconsiderado; desconsiderar; descontinuação; descontinuações; descontinuado; descontinuando; descontinuar; desconto; descontos; descontroladamente; descontrolado; descrédito; descuidado; descuidadosamente; desculpar; desempenho inferior; desempregado; desemprego; desencoraja; desencorajado; desencorajador; desencorajar; desequilíbrio; desequilíbrios; desestabilização; desestabilizando; desestabilizar; desfavor; desfavorável; desfavorecendo ; desfavoravelmente; desfavorecido; desfavores; desfiguração; desfigurado; desfigurar; desfiliação; desgostado; desgostando; desgostar; desgraça; desigualdade; desigualdades; desigualmente; desincentivos; desinforma; desinformação; desinformado; desinformando; desinformar; desinteressadamente; desinteressado; desinteresse; desinveste; desinvestido; desinvestimento; desinvestimentos; desinvestindo; desinvestir; desistindo; desistir; desleal; deslealdade; deslealmente; desligamentos; desligar; deslistado; deslistar; deslizamento; deslizes; deslocados; deslocamento; deslocamentos; deslocando; deslocar; desmoronado; desmoronando; desnecessariamente; desnecessário; desocupado; desonestidade; desonesto; desonra; desonrado; desonrando; desonras; desonrosamente; desonroso; desordenado; despedido; despeja; despejado; despejando; despejar; despejo; despejos; desperdiçado; desperdício; desprevenido; desprezado; desprezando; desprezível; desprezo; desprezos; desproporção; desproporcional; desproporcionalmente; desqualifica; desqualificação; desqualificações; desqualificado; desqualificando; desqualificar; desestabilizado; destrói; destruição; destruído; destruindo; destruir; destrutivo; desvaloriza; desvalorizado; desvalorizando; desvalorizar; desvantagem; desvantagens; desvantajoso; desvia; desviado; desviando; desviando fundos; desviar; desvio; desvio de fundos; desvios; desvios de fundos; detenção; detenções; deter; deteriora; deterioração; deteriorações; deteriorado; deteriorando; deteriorar; detido; detrair; detratões; detratado; detratando; detrimento; detrimetos; deturpa; deturpando; deturpar; devagar; devastação; devastado; devastador; devistar; devolução; devoluções; devolve; devolver; devolvido; difama; difamação; difamações; difamado; difamando; difamar; difamatório; difamou; difícil; dificuldade; dificuldades; dificulta; dificultado; dificultando; dificultar; diminui; diminuição; diminuído; diminuindo; diminuir; direcionado incorretamente; disciplinar; discorda; discordando; discordo; discordou; discrepancia; discrepancias; discutido; discutindo; discutir; disfunção; disfuncional; disfunções; disparidade; disparidades; disparo; dispensa; disputa; disputado; disputando; disputas; disruptivo; dissensos; dissidente; dissidentes; dissidência; dissolução; dissoluções; dissuado; dissuasão; dissuassões; dissusores; distorção; distorce; distorcendo; distorcer; distorcido; distorções; distração; distrações; distrai; distraído; distraindo; distrair; distúrbios; dividido; divorciado; divórcio; divulgação; divulgado;

divulgando; divulgar; doente; dominado; drasticamente; drástico; dúvida; duvidoso; duvidou; egrégios; eliminar; em declínio; em perigo; embaraço; embaraços; embaraçoso; embargado; embargo; embargos; encarcerado; encarcerados; encarceramento; encarceramentos; encarcerando; encarcerar; encerramento; encerramentos; encolhimento; encolhimentos; enfraquece; enfraquecer; enfraquecido; enfraquecimento; enfrentar; engana; enganado; enganador; enganando; enganar; engano; enganos; enganosamente; enganoso; envergonhado; erodido; erodindo; erodir; erosão; erra no julgamento; errado; errando; erraticamente; errático; erro; erro de cálculo; erro de julgamento; erroneamente; errôneo; erros; erros de julgamento; escala; escalado; escalando; escalar; escandalosamente; escândalos; escandaloso; escassez; escrutinando; escrutinar; escrutínio; esforço; esgota; esgotado; esgotamento; esgotando; esgotar; esquecido; estagna; estagnação; estagnado; estagnar; estressado; estressante; estresse; evade; evadindo; evadir; evasão; evasivo; evasões; evitada; evitar; exacerbação; exacerbações; exacerbado; exacerbando; exacerbar; exagera; exagerado; exagerando; exagerar; exagero; exageros; examinado; excessivamente; excessivo; excesso; excesso de capacidade; excesso de oferta; excluído; exclusão; exclusão da lista; exclusões; exculpa; exculpação; exculpações; exculpado; exculpando; excuplatório; execução hipotecária; execuções hipotecárias; executado; executando hipotecária; exonera; exoneração; exonerações; exonerado; exonerando; exonerar; explicitamente; exploração; explorações; explorado; explorando; explorar; explorativo; expõe; expondo; expor; exposto; expropria; expropriação; expropriado; expropriando; expropriar; expulsão; expulsões; extenuante; extorsão; extorsionista; extorsivo; extraviado; facilitando; falecendo; falecido; falênci;a falências; falha; falhar; falhas; falido; falido; falindo; falsamente; falsidade; falsifica; falsificação; falsificações; falsificado; falsificador; falsificadores; falsificando; falsificar; falso; falta; falta de comunicação; falta de desempenho; falta de pagamento; falta de vontade; faltas; faltas de execução; faltava; fardo; fardos; fatalidade; fatalidades; fatalmente; fecha; fechado; fechando; fechar fere ; ferida; ferido; ferindo; ferir; fiança; fictício; fim; foge; fora de moda; forçado; forçando; fracamente; fracassado; fraco; fraqueza; fraquezas; fraude; fraudes; fraudulentemente; fraudulento; frívolo; frustra; frustração; frustrações; frustrado; frustrante; frustrantemente; frustrar, fugitivo; fugitivos; gargalo; gargalos; gratuitamente; gratuito; gravidade; gravidades; grosseiramente; hostil; hostilidade; hostilmente; identificado incorretamente; ignora; ignorado; ignorando; ignorar; ilegal; ilegalidade; ilegalidades; ilegalmente; ilegível; ilicitamente; ilícito; iliquidex; ilíquido; imaturo; Incompatibilidade; Incompatibilidades; iminente; imoral; impasse; impasses; impede; impedem; impedido; impedimento; impedimentos; impedindo; impedir; imperativo; imperfeição; imperfeições; implica; implicação; implicado; implorou; impopular; impossibilidade; impossível; impraticabilidade; impraticabilidades; impraticável; imprecisamente; imprecisão; impreciso; imprecisões; imprevisibilidade; imprevisível; imprevisivelmente; imprevisto; improdutivo; imprópria; impropriedade; impropriedades; impróprio; imprudência; imprudente; imprudentemente; inação; inaceitável; inaceitavelmente; inacessível; inações; inadequação; inadequações; inadequadamente; inadequado; inadimplente; inadmissível; inadvertidamente; inadvertido; inalienável; inativa; inativação; inativações; inativado; inativando; inativar; inatividade; incapacidade; incapacitado; incapaz; incerto; incidência; incidências; incidente; incidentes; ; incobrável; incomodado; incômodo; incômodos; incompatibilidade; incompatibilidades; incompatível; incompetência; incompetente; incompetente; incompetentemente; incompetentes; incompletamente; incompleto; incompletude; inconcebível; inconclusivo; inconsistência; inconsistências; inconsistente; inconsistentemente; incontrolável; inconveniência; inconveniências; inconveniente; incorretamente; incorreto; indecência; indecente; indeferimento; indesejado; indesejável; indeterminado; indevidamente; indevido; indiciado; indiciando; indicar; indisponibilidade; indisponível; infecácia; infeciz; infecizamente; inficiência; inficiências; inficiente; inficientemente; inegociável; inelegibilidade; inelegível; inesperadamente; inesperado; inevitável; inevitavelmente; inexperiência; inexperiente; iníquo; infeliz; infelizmente; inferior; infligido; infração; infrações; infrator; infratores; infringir; inibido; inimigo; injunção; injunções; injustamente; injustificado; injustificável; injustificavelmente; injusto; insatisfação; insatisfatório; insatisfeito; inseguro; insensível; insignificante; ;insólido; insolvência; insolvências; insolvente; instabilidade; instabilizado; instável; instigando; insubordinação; insuficiência; insuficiente; insuficientemente; insurreição; insurreições; insustentável; intempestivo; intencional; interfere;

interferência; interferências; interferido; interferindo; interferir; intermitente; intermitentemente; interpretação errada; interpretações erradas; interpretando errado; interpretar mal; interrompe; interrompendo; interromper; interrompido; interrupção; interrupções; intimação; intimidação; intrusão; inútil; inutilizável; invadido; invadindo; invadir; invalida; invalidação; invalidado; invalidando; invalidar; invalidez; inválido; invasão; invasões; invendável; invendível; inveracidade; inverdade; inverdades; investiga; investigação; investigações; investigado; investigando; investigar; inviabilidade econômica; inviável; involuntariamente; voluntário; irracional; irracionalidade; irracionalmente; irrealista; irreconciliável; irreconciliavelmente; Irrecuperavelmente; irrecuperável; irregular; irregularidade; irregularidades; irregularmente; irreparável; irreparavelmente; irresponsável; irreversível; irrevogável; irrevogavelmente; irrita; irritante; irritar; isenção de responsabilidade; isenções; isenções de responsabilidade; isento de responsabilidade; itens não colecionáveis; julga mal; julgamento errado; julgando mal; julgar mal; justificável; lapsó; lapsos; lavagem de dinheiro; lentamente; lentidão; lento; levianamente; lesões; liberar; limitação; limitações; limitada; limitar; limites; liquida; liquidação; liquidações; liquidado; liquidadores; liquidando; liquidante; liquidar; litiga; litigado; litigando; litigante; litigantes; litigar; litígio; litígios; má conduta; má gestão; má interpretação; machucando; mácula; mais arriscado; mais devagar; mais duro; mais fraco; mais lento; mais perigoso; mais rigoroso; mais severo; mal; mal administrado; mal aplicadas; mal direcionado; mal interpretado; mal julgado; mal pago; mal-entendido; mal-entendidos; malícia; maliciosamente; malicioso; malversador; malvisto; manchas; manifestante; manifestantes; manipula; manipulação; manipulações; manipulado; manipulando; manipular; manipulativo; manuseado mal; manuseia incorretamente; manuseio incorreto; marcha lenta; mau funcionamento; medos; menosprezar; menosprezos; mentindo; mentira; minam; minar; minimização; molestar; monopólio; monopolistas; monopolístico; monopoliza; monopolização; monopolizado; monopolizando; monopolizar; moratória; moratórias; morto; multado; multas; não aprovado; não atingimento; não atraente; não autorizado; não colecionável; não coletado; não comercializável; não competitivo; não confiável; não conformidade; não conformidades; não compatível; não contabilizado; não corrigido; não cumprido; não detectado; não divulgação; não divulgado; não documentado; não entregue; não funcional; não intencional; não liquidado; não lucrativo; não obtido; não operante; não pagamento; não pagamentos; não pago; não permite; não permitido; não permitindo; não permitir; não planejado; não produtivo; não produzindo; não programado; não qualificado; não realizado; não receptivo; não recuperado; não recuperável; não reembolsado; não remediado; não renovação; não reportado; não resolvido; não vendido; nega; negação; negações; negado; negando; negar; negativamente; negativo; negativos; negligencia; negligência; negligência médica; negligenciado; negligenciar; negligências; negligente; negligentemente; objeção. objecionável; objeções; ;objetando; obscenidade; obsceno; obscuro; obsolescência; obsoleto; obstáculo; obstáculos; obstrução; obstruções; obstruído; obstruindo; obstruir; ocioso; ocultando; oculto; ofende; ofendendo; ofender; ofendido; ofensa; ofensas; offensivo; ofusca; ofuscando; ofuscar; omissão; omissões; omite; omitido; omitindo; omitir; oneroso; ônus; opõe-se; opor; oportunista; oportunistamente; oposição; oposições; oposto; padrão; padrões; pagamento em excesso; pagamento inferior; pagamentos em excesso; pagamentos inferiores; paga insuficientemente; pago em excesso; pane; panes; pânico; pânicos; parada; paradas; parado; paralisações; parar; parou; passível de acusação; passo em falso; passos em falso; pausas; pejorativo; pena; penalidades; penaliza; penalizado; penalizando; penalizar; penoso; perda; perdas; perde; perdendo; perder; perdido; perdoado; perdoar; pergunta; perigo; perigos; perigosamente; perigoso; perjúrio; perpetra; perpetração; perpetrado; perpetrando; perpetrando; perpetrar; persiste; persistência; persistente; persistentemente; persistindo; persistir; perturbação; perturbado; perturbador; perturbar; pervasivamente; pervasivo; peso morto; pesos mortos; pior; piora; piorando; piorar; piorou; piquete; piquetes; pobre; ponte; por engano; por favor; posando; poses; postergado; precipitadamente; precipitado; precipitoso; precluído; precluindo; precluir; preço errado; preconceito; preconceitos; preconceituoso; preços errados; predatório; prejudica; prejudicado; prejudicando; prejudicar; prejudicial; prejudicialmente; pré-julgamento; prematuramente; prematuro; prender; preocupação; preocupações; preocupado; preocupante; preocupar; preso; pressão; pressionante; pretender; pretendido; prevaricação; prevenção; prevenido; prevenir; previne; prisão; prisões; priva;

privação; privado; privando; privar; problema; problemas; problemático; processado; processamento; processando; processar; processos; proíbe; proibições; proibido; proibindo; proibir; prolonga; prolongado; prolongamento; prolongamento excessivo; prolongamentos; prolongando; prolongar; propenso; propina; propinas; protestado; protestando; protesto; protestos; provoca; provocado; provocando; provocar; punição; punições; punido; punitivo; quebra; quebrado; quebrar; quebras; queimado; questionado; questionando; questionável; questionavelmente; ; questões; racionaliza; racionalização; racionalizações; racionalizado; racionalizando; racionalizar; reafirmar; reapropriado; reatribuição; reatribuições; reatribuir; reavaliação; reavaliações; rebaixa; rebaixado; rebaixamento; rebaixamentos; rebaixando; rebaixar; recaída na inadimplência; recall; recalls; recessão; recessionário; recessões; reclama; reclamação; reclamações; reclamando; reclamar; reclamou; recolhendo; recolhido; recolhimentos; reconvenção; reconvencionado; reconvenções; reconvindo; recuperar; recuperar posse; recusa; recusado; recusando; recusar; recusas; redação; redações; reapresentação; redefinição; redefinida; redesignado; redesignando; redigido; redigir; redução; redução de tamanho; reduções; reduções de tamanho; reduz; reduzido; reduzindo; reduzir; reduzir o tamanho; reestruturação; reestruturações; reestruturado; reformula; reformulações; reformulado; reiniciando; reintegração de posse; reintegrações de posse; reivindicações; reivindicando; rejeição; rejeições; rejeita; rejeitado; rejeitando; rejeitar; rejeitou; relutância; relutante; remanejamentos; renegocia; renegociação; renegociações; renegociado; renegociar; renuncia; renúncia; renunciado; renunciar; renúncias; reparação; reparações; reparado; representação incorreta; representações erradas; representado mal; reprovado; repudiado; repudiados; repudiando; repudiar; repúdio; repúdios; rescisão; rescisões; resgate; resignado; restituições; restringir; retalia; retaliação; retaliações; retaliado; retaliando; retaliar; retomar posse; retribuição; retribuições; retrodatação; réu; réus; revés; revirando; reviravolta; reviravoltas; revoga; revogação; revogações; revogado; revogando; revogar; ridicularizado; ridicularizando; ridículo; ridículos; rigoroso; rotulado errado; rotulado erroneamente; rotulado incorretamente; rotulagem errada; rótulo errado; rótulos errados; roubado; ruim; sabotar; sacrificado; sacrificando; sacrificial; sacrifício; sacrifícios; seca; secas; segredo; sem anúncio; sem brilho; sem desculpa; sem financiamento; sem fundamento; sem fundamentação; sem fundamento; sem intenção; sem licença; sem mérito; sem saber; sem seguro; sem sucesso; sendo vencido; seriamente; sério; severamente; severo; significados; sobrecarga; sobrecargas; sobrecarrega; sobrecarregado; sobrecarregando; sobrecarregar; sobrevalorização; sofre; sofrer; sofrido; sofrimento; sombria; sombrio; spam; spammers; subcapitalizado; subcotar; subestima; subestimação; subestimado; subestimando; subestimar; subfinanciado; submetendo; subnotificação; subornado; subornando; suborno; subornos; subprodução; subproduzido; subsegurado; subutilização; subutilizado; sujeição; sujeito; superestima; superestimação; superestimações; superestimado; superestimar; superficialmente; superprodução; superproduz; superproduzido; supervalorização; supervalorizado; súplica; supostamente; suscetibilidade; suscetível; suspeita; suspeitamente; suspeitas; suspeito; suspeitos; suspender; suspensão; suspenso; suspensões; tarde; temer; tempo de inatividade; tempos de inatividade; tendencioso; tensão; tenso; termina; terminado; terminando; terminar; testemunhar; tolera; tolerado; tolerância; tolerante; tolerar; tortuosamente; tortuoso; tragédia; tragédias; tragicamente; trágico; transbordando; traumático; tumulto; turbulência; única prevalência; urgência; urgente; usado mal; uso indevido; usos indevidos; usura; usurpa; usurpado; usurpando; usurpar; vandalismo; vencido; veredito; vereditos; vergonhosamente; vergonhoso; vetado; viciação; vigor; viola; violação; violações; violado; violadores; violando; violar; violador; violência; violentamente; violento; virado; vítimas; volátil; volatilidade; vulnerabilidade; vulnerabilidades; vulnerável; vulneravelmente.

Apêndice G – Listas das Palavras Incerteza

acredita, aleatória, aleatoriamente, alteração, ambiguidade, ambíguo, anomalia, antecipa, aparece, aparente, aproxima, arbitrariamente, arbitrariedade, arbitrário, arriscado, assume, caprichos, cauteloso, com cautela, concebível, condicional, confunde, confusão, confusamente, contingência, cruzamento, cuidado, de alguma forma, de repente, depende, dependência, desconhecido, desestabilizando, desvio, diferencia, diferente, diferido, dobradiças, dúvida, duvidado, em algum lugar, encruzilhada, esclarecimento, escondido, especula, esporadicamente, esporádico, exposição, exposições, flutua, flutuação, flutuações, flutuante, fora de época, fora de temporada, imprecisão, impreciso, imprecisões, imprevisibilidade, imprevisível, imprevisto, improbabilidade, improvável, incertamente, incompletude, incomum, indeciso, indefinição, indefinido, indefinitivamente, indetectável, indeterminado, inesperadamente, inexatidão, inexato, inquieto, instabilidade, intangíveis, intangível, mais arriscado, mais riscado, não avaliado, não claro, não comprovado, não confirmado, não designado, não documentado, não escrito, não especificado, não específico, não identificado, não observável, não planejado, não quantificado, não reconciliado, não testado, normalmente, ocasionalmente, parece, pendente, poder, possibilidade, possível, possivelmente, precaução, precauções, preditivo, preditor, preliminares, preliminarmente, presumido, presumivelmente, presunção, presunções, previsão, previsibilidade, previsões, previsto, probabilidade, probabilístico, provável, provavelmente, provisória, quase, randomizado, raramente, reavalia, reavaliação, recalcula, recálculo, reconsidera, reconsiderado, reexame, reexaminando, reinterpreta, repentino, rever, revisado, risco, rumores, sem cobertura, sem garantia, sugerido, sugestões, sugira, suposição, suposições, suscetibilidade, suspensão, suspensões, talvez, turbulência, vagamente, varia, variabilidade, variação, variações, variado, variância, variante, variável, variavelmente, volátil, volatilidade.

Apêndice H – Exemplificação Cálculos de Legibilidade, Tom e Incerteza

Abaixo, apresento o cálculo de legibilidade, tom e incerteza em um relatório da administração de uma empresa selecionada aleatoriamente. Na **Tabela 19**, são detalhados os parágrafos do relatório da administração, com a contagem de frases, sílabas, e a frequência de palavras classificadas como “positivas”, “negativas” e de “incerteza” em cada frase.

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
[Início da Frase 1] “Senhores Acionistas, Em cumprimento às disposições legais, a Z S.A., maior empresa da América Latina no desenvolvimento de software aplicativos, a 7ª maior desenvolvedora de sistemas de gestão integrada (ERP) do mundo, 1ª em países emergentes e líder absoluta no Brasil, submete à apreciação de seus acionistas o Relatório da Administração e as correspondentes Demonstrações Financeiras Societárias, acompanhadas do parecer dos auditores independentes, referentes aos exercícios findos em 31 de dezembro de 2010 e 2009, elaboradas de acordo com as práticas contábeis adotadas no Brasil.”	Fr 1: 85	207	5	1	1
*Integridade					
[Início da Frase 2] “MENSAGEM DA ADMINISTRAÇÃO Durante o ano de 2010, a economia brasileira experimentou forte crescimento, dando sequencia à rápida recuperação ocorrida depois da crise dos mercados financeiros mundiais e que afetaram a economia global, de maneira generalizada em meados de 2008 e 2009.” [Início da Frase 3] “No caso da Z, os investimentos na consolidação da marca e na unificação dos canais de distribuição, e a melhora de processos internos e das estruturas de desenvolvimento, foram os principais focos de gestão da empresa em 2010.”	Fr 2: 42	100	2	2	0
[Início da Frase 4] “Os resultados desse trabalho já estão refletidos, em mais um ano com indicadores favoráveis, com o atingimento de novos patamares históricos em vendas, número de novos clientes, rentabilidade e número de participantes.”	Fr 3: 38	77	1	1	0
[Início da Frase 5] “Dentro da visão de longo prazo da empresa, mais uma vez a Z reforçou sua capacidade de inovação, com gastos próximos a R\$150,0 milhões em pesquisa e desenvolvimento, o que equivale a 13,3% da receita líquida da empresa, e que permanece como um dos maiores montantes do mercado brasileiro.” [Início da Frase 6] “De acordo com estudo da Booz and Co. publicado no Jornal Valor Econômico, a Z é considerada hoje uma das 1.000 empresas que mais investem em pesquisa e desenvolvimento no mundo.” [Início da Frase 7] “Ainda no tocante à visão de longo prazo, o quadro da Companhia foi reforçado com aproximadamente 870 novos participantes, sendo que o principal acréscimo ocorreu nas áreas de desenvolvimento e implementação de software.” [Início da	Fr 4: 32	78	2	1	0
	Fr 5: 49	101	3	0	0
	Fr 6: 31	63	0	0	0
	Fr 7: 33	79	1	0	1
	Fr 8: 57	129	5	1	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
[Frase 8] “O crescimento observado fez com que a Companhia atingisse faturamento superior à marca de R\$1 bilhão e, como consequência da rápida expansão de seus negócios, a participação do mercado de software de gestão, segundo o estudo “Latin America Semiannual ERM Tracker” do instituto IDC, no final do primeiro semestre de 2010, aumentou para 45,3% do mercado brasileiro.”					
[Início da Frase 9] “A Administração avalia que a Z encerrou o ano de 2010 mais preparada, em temos de estrutura, equipe e tecnologia, para dar respostas positivas e contínuas aos principais responsáveis pelo sucesso da companhia, ou seja, clientes, acionistas, participantes, fornecedores e demais parceiros do mercado.”	Fr 9: 44	107	2	1	0
[Início da Frase 10] “CENÁRIO ECONÔMICO					
Em 2010, os reflexos da crise financeira, oriunda dos Estados Unidos de 2008 e que se alastrou às demais economias do mundo, permaneceram afetando o cenário global, notadamente nos países centrais, como Estados Unidos e Inglaterra, e em economias européias como a da Espanha, na qual o desemprego permaneceu em patamar de dois dígitos.”	Fr 10: 56	119	0	3	0
[Início da Frase 11] “Ainda na Europa, economias como a da Grécia, tiveram que ser socorridas com financiamentos de outros governos, a fim de evitar contágio na região.”	Fr 11: 24	55	0	2	0
[Início da Frase 12] “Por outro lado, as economias emergentes, como a do Brasil e algumas asiáticas, com destaque para a China, experimentaram um momento distinto com forte crescimento do PIB, motivado principalmente pela inserção de uma nova camada de consumidores na economia interna de cada um desses países.”	Fr 12: 45	109	4	0	0
[Início da Frase 13] “Outra característica relevante do cenário econômico no ano de 2010 foi o acirramento da “guerra cambial” entre os vários países.”	Fr 13: 20	50	0	0	0
[Início da Frase 14] “Nesse movimento, os países buscam resolver seus problemas internos de demanda desvalorizando suas moedas a fim de fazer com que seus produtos fiquem mais baratos no mercado internacional.”	Fr 14: 28	67	1	3	0
[Início da Frase 15] “O efeito no Brasil dessa “guerra” foi a apreciação de 4,3% do Real em relação ao dólar americano, e vários setores da economia brasileira passaram a apresentar perda de competitividade no mercado externo, e também no mercado interno pela competição com os produtos importados em praticamente todos os setores da economia.”	Fr 15: 51	122	1	1	0
[Início da Frase 16] “De qualquer forma, o que se viu no Brasil foi um período de crescimento e ampliação de consumo, com forte redução do número de brasileiros nas classes mais baixas e uma concentração maior de poder aquisitivo na classe média (classe C).”	Fr 16: 41	83	3	2	1
[Início da Frase 17] “Os recordes sucessivos de vendas em setores como o imobiliário e de automóveis, se de um lado trouxeram os benefícios de uma economia aquecida por outro lado fizeram aumentar a preocupação com a aceleração inflacionária.	Fr 17: 35	89	1	2	0
[Início da Frase 18] “A taxa de inflação dada pelo IPCA, que fora em 2009 de 4,31%, acabou 2010 com alta de 5,91%, portanto ligeiramente acima da meta estabelecida pelo Governo.”	Fr 18: 27	52	0	0	0
[Início da Frase 19] “Já o IGPM alcançou 11,32%, sensivelmente superior à deflação	Fr 19: 14	28	3	0	0
	Fr 20: 44	87	0	3	0
	Fr 21: 46	95	2	1	0
	Fr 22: 41	86	1	1	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
de 1,71% alcançada em 2009.” [Início da Frase 20] “A fim de atacar esse processo inflacionário, buscando limitar a demanda, a Selic, taxa de juro básica estabelecida pelo Banco Central, que havia iniciado o ano em 8,75% a.a., a menor desde que foi criada em 1999, chegou no final de 2010 a 10,75% a.a.” [Início da Frase 21] “Importante citar que o crescimento do PIB de 2010 – de cerca de 7,6%, segundo estimativas, já que até a data deste relatório, o número oficial não havia sido informado, – ocorreu mesmo com o aumento dos juros, bem como em um período de sucessão presidencial.” [Início da Frase 22] “Depois de ser reeleito em 2006, o presidente Luiz Inácio Lula da Silva terminou oito anos de governo com elevada popularidade e fez como sua sucessora, eleita no mês de Novembro, a presidente Dilma Rousseff, primeira mulher a comandar o Brasil.” [Início da Frase 23] “O período eleitoral transcorreu sem sobressaltos, com transição democrática sem efeito relevante na taxa de juros, câmbio ou qualquer outro índice de mercado.”	Fr 23: 23	58	0	0	0
[Início da Frase 24] “O Brasil permaneceu em 2010 com destaque nas pautas da mídia e dos investidores internacionais, não só pela continuidade de seu bom desempenho econômico, mas também pela expectativa dos investimentos adicionais com vistas ao avanço de infraestrutura necessários para dar continuidade ao crescimento do país e para dar suporte às necessidades como país-sede da Copa do Mundo de Futebol de 2014 e os Jogos Olímpicos de 2016, que serão sediados na cidade do Rio de Janeiro.”	Fr 24: 77	172	3	0	0
[Início da Frase 25] “Beneficiada por esse cenário, a indústria de Tecnologia de Informação (TI) apresentou crescimento em 2010.” [Início da Frase 26] “No mesmo estudo, já citado, do instituto IDC, datado de dezembro de 2010, a estimativa era de que o mercado de software de gestão no Brasil tivesse crescido mais de 14% na comparação com 2009 e na América Latina acima de 11% em 2010, com uma expectativa de taxa composta de crescimento anual acima de 12% a.a. no período 2008-2014.” [Início da Frase 27] “O ciclo de crescimento da economia, a maior utilização dos instrumentos oriundos da internet, a precificação mais acessível dos computadores e a obrigatoriedade de que mais empresas tenham que aderir ao Sistema Público de Escrituração Digital (SPED) e emitir Nota Fiscal Eletrônica (NF-e) tem ajudado no crescimento de vendas de software de gestão.”	Fr 25: 15	40	3	0	0
[Início da Frase 28] “DESEMPENHO FINANCEIRO E OPERACIONAL CONSOLIDADO	Fr 26: 59	114	2	0	0
Mais uma vez, suportada por fortes fundamentos, a Companhia encerrou um ciclo de 12 meses com indicadores positivos como o aumento do número de novos clientes, crescimento em todas as linhas de receita e lucro recorde de R\$137.863 mil, 14,6% superior ao ano anterior.” [Início da Frase 29] “A Z alcançou EBITDA(*) (“Earnings before interest, taxes, depreciation and amortization”; lucro antes de juros, impostos, depreciação e amortização) de R\$289.552 mil em 2010, com crescimento de 16,2% sobre	Fr 27: 53	125	3	0	0
	Fr 28: 49	96	5	2	0
	Fr 29: 48	96	3	1	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
o ano de 2009 e a margem EBITDA(**) atingiu 25,6%, acima 40 pontos-base quando comparada à margem de 2009.”					
[Início da Frase 30] “A Companhia tem conseguido crescer e ao mesmo tempo ampliar sua lucratividade, sem afetar os investimentos em pesquisa e desenvolvimento, fundamentais para a construção do futuro da Z.”	Fr 30: 28	67	2	1	0
[Início da Frase 31] “O ganho de margem tem sido fruto do crescimento da receita combinado a uma forte captura de sinergias oriundas das empresas adquiridas nos últimos anos, e de ganho de eficiência no sistema interno de gestão, com implantação de novos processos e de indicadores de qualidade.”	Fr 31: 45	99	5	2	0
[Início da Frase 32] “A Z encerra o ano de 2010 com receita líquida de R\$1.129.475 mil, com crescimento de 14,2% sobre a receita líquida de 2009.”	Fr 32: 23	37	1	1	0
[Início da Frase 33] “A reorganização no sistema de distribuição, iniciada no segundo semestre de 2009, e a expansão econômica do Brasil e da região, colaboraram significativamente com esse desempenho.”	Fr 33: 26	68	2	0	0
[Início da Frase 34] “A receita de taxas de licenciamento totalizou R\$278.845 mil e cresceu 22,1% ante os R\$228.270 mil de 2009.”	Fr 34: 18	31	1	0	0
[Início da Frase 35] “O incremento é justificado pela adição de 2.840 novos clientes de software, e pelo significativo número de clientes atendidos na atual base da Z.”	Fr 35: 24	55	0	1	0
[Início da Frase 36] “A receita de serviços foi superior em 10,7% à de 2009, alcançando R\$367.698 mil.”	Fr 36: 14	24	2	0	0
[Início da Frase 37] “Esse crescimento deve-se principalmente ao aumento no volume de implementações, em função de novas vendas de software para clientes da base e novos e a ampliação das operações dos Serviços de Valor Agregado e de Consultoria.”	Fr 37: 36	79	2	0	0
[Início da Frase 38] “A receita de manutenção encerrou 2010 com R\$513.860 mil, montante 12,6% acima dos R\$456.515 mil de 2009.”	Fr 38: 17	27	0	1	0
[Início da Frase 39] “O aumento da base de clientes ativos da Companhia, isto é, que pagam mensalmente por serviços de atualização do software e atendimento, afeta diretamente essa linha de receita, bem como a baixa perda de clientes ao longo do ano foram causas importantes do crescimento dessa rubrica.”	Fr 39: 46	104	2	3	0
[Início da Frase 40] “A Z encerra 2010 com mais de 26 mil clientes ativos.”	Fr 40: 11	15	0	1	0
[Início da Frase 41] “O custo das taxas de licenciamento apresentou crescimento de 5,3%, saindo de R\$22.909 mil, em 2009, para R\$24.118 mil, em 2010.”	Fr 41: 21	36	1	0	0
[Início da Frase 42] “Essa linha de custo representa o custo com soluções de terceiros vendidas pela Companhia.”	Fr 42: 14	30	0	0	0
[Início da Frase 43] “A soma dos custos de serviços e vendas totalizou R\$353.332 mil em 2010, crescimento de 19,1% sobre 2009.”	Fr 43: 18	30	1	0	0
[Início da Frase 44] “Esse crescimento de custos, superior ao da receita de serviços, deve-se principalmente ao processo de reestruturação das franquias, principais canais de distribuição da Companhia, que, embora tenha ocorrido no final de 2009, ainda apresentou reflexo de perda de eficiência em 2010.”	Fr 44: 41	98	3	3	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
[Início da Frase 45] “As despesas com pesquisa e desenvolvimento atingiram R\$149.924 mil em 2010, e representaram 13,3% da receita líquida da Companhia.”	Fr 45: 19	42	1	0	0
[Início da Frase 46] “A ampliação dos investimentos na tecnologia proprietária da Z, a fim de torná-la mais versátil, e investimentos em inovações como no segmento de TV digital, foram os principais itens que influenciaram essa linha de despesa, que, conforme já descrito anteriormente, são fundamentais para o futuro dentro do setor de tecnologia da informação.”	Fr 46: 52	123	1	2	0
[Início da Frase 47] “As despesas com propaganda totalizaram R\$32.335 mil em 2010, o equivalente a 2,9% da receita líquida.” [Início da Frase 48] “O foco dos investimentos nessa rubrica foi de reforço da marca Z, através das campanhas publicitárias e de marketing no Brasil e exterior.”	Fr 47: 16	33	0	0	0
[Início da Frase 49] “As despesas com vendas cresceram 24,1% entre 2010 e 2009, alcançando R\$71.225 mil, impactadas por reorganizações internas, a partir da incorporação de equipes de vendas que anteriormente estavam nas franquias.” [Início da Frase 50] “Em contrapartida, conforme explicação no próximo parágrafo, houve redução no volume de comissões pagas, em função da internalização de equipes de vendas.”	Fr 49: 30	66	2	0	0
[Início da Frase 51] “As despesas com comissões passaram de R\$102.970 mil, em 2010, para R\$108.508 mil, em 2010, crescimento de 5,4%.” [Início da Frase 52] “Esse crescimento menos acelerado do que o aumento de vendas foi influenciado, conforme o parágrafo acima, pela internalização de estruturas de vendas.”	Fr 51: 18	28	1	0	0
[Início da Frase 53] “As despesas gerais e administrativas totalizaram em 2010 R\$71.819 mil, redução de 5,8% sobre os R\$76.268 mil de 2009.” [Início da Frase 54] “A representatividade dessa linha de despesa em relação à receita líquida passou de 7,7%, em 2009 para 6,4%, em 2010, refletindo os ganhos de escala da otimização de estrutura.”	Fr 52: 22	56	2	1	0
[Início da Frase 55] “As despesas com honorários da administração atingiram em 2010 R\$27.967 mil, redução de 11,9% sobre o valor do ano de 2009, influenciado pela provisão do atingimento proporcional das metas financeiras da empresa.”	Fr 53: 19	33	0	1	0
[Início da Frase 56] “A despesa com provisão para créditos de liquidação duvidosa em 2009 acumulou R\$10.595 mil, 37,2% superior a 2009.” [Início da Frase 57] “O incremento é decorrente da análise da carteira de recebíveis da empresa, e o julgamento sobre potenciais perdas que a empresa possa ter, e que corresponde a 0,9% da receita líquida 2010 (0,8% em 2009).”	Fr 54: 29	60	1	1	0
[Início da Frase 58] “As despesas com depreciação e amortização totalizaram em 2010 R\$77.883 mil, ante R\$69.918 mil em 2009, aumento de 11,4%.” [Início da Frase 59] “Essa elevação do patamar de despesas com depreciação e amortização traduz o crescimento da infraestrutura da Companhia e o aumento do volume de	Fr 55: 32	71	2	1	0
	Fr 56: 18	36	1	2	1
	Fr 57: 35	70	0	1	0
	Fr 58: 19	35	1	1	0
	Fr 59: 31	77	2	1	1

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
amortização de intangíveis oriundos de aquisições de operações.”					
[Início da Frase 60] “As despesas financeiras alcançaram R\$68.172 mil, um aumento de 35% em comparação às despesas de R\$50.481 mil em 2009.” [Início da Frase 61] “Esse aumento decorre, principalmente, da maior despesa de juros sobre financiamentos BNDES e PROSOFT, (R\$37.023 mil em 2010 vs. R\$33.480 mil em 2009), marcação a mercado do prêmio de não conversão das debêntures (R\$11.801 mil em 2010 vs. R\$9.442 mil em 2009), e deságio na emissão de ações aos debenturistas em 19 de agosto de 2010 quando da conversão obrigatória a preço inferior ao de mercado (R\$7.442 mil em 2010).”	Fr 60: 19	34	2	0	0
[Início da Frase 62] “O caixa da Z em 31 de Dezembro de 2010 é de R\$232.508 mil, representando um aumento de 11,9% sobre a posição de R\$ 207.721 mil em 31/12/2009.”	Fr 61: 70	124	2	1	0
[Início da Frase 63] “A empresa reitera que tem política conservadora na gestão de suas atividades financeiras, ativas e passivas, e que não realizou nenhuma operação especulativa no mercado financeiro.”	Fr 62: 28	38	1	0	0
[Início da Frase 64] “O lucro líquido alcançou o recorde de R\$137.863 mil em 2010, crescimento de 14,6% na comparação com 2009.”	Fr 63: 26	70	1	0	0
[Início da Frase 65] “(*) O EBITDA é uma medição não contábil elaborada pela Companhia, calculada observando uma das sugestões do Ofício Circular CVM nº 01/2007, consistindo no lucro antes dos impostos e contribuições, das participações de minoritários, despesa/receita financeira líquida (receita e despesas financeiras), depreciação e amortização e resultado não operacional.”	Fr 64: 18	31	2	0	0
[Início da Frase 66] “(**) A margem de EBITDA representa a relação percentual entre o EBITDA e receita líquida.”	Fr 65: 48	126	0	1	1
[Início da Frase 67] “MERCADO DE CAPITAIS Em 31 de dezembro de 2010, as ações em circulação (free float) representavam 76,4% do total de ações da Companhia, ou 24.035.873 de um total de 31.459.272 ações.” [Início da Frase 68] “O percentual que não faz parte do free float corresponde às participações dos administradores da Companhia, incluindo as pessoas vinculadas e empresas controladas por estes (18,33%), assim como a participação da BNDES Participações S/A (5,27%).”	Fr 66: 15	29	0	0	0
[Início da Frase 69] “Em 2010 as ações da Z (Bovespa: Z) apresentaram valorização de 43,2% frente a valorização de 1,0% do IBOVESPA.”	Fr 67: 31	49	0	0	0
[Início da Frase 70] “O volume financeiro médio no ano de 2010 registrou R\$15,7 milhões/dia, contra R\$4,9 milhões/dia no ano de 2009, crescimento de 217,7%.”	Fr 68: 35	80	0	0	0
[Início da Frase 71] “Juros sobre o capital próprio referentes ao exercício 2009: aprovados e pagos respectivamente em 10 de dezembro de 2009 e 20 de janeiro de 2010, o montante de R\$22.500 mil é relativo ao exercício 2009 e foi imputado ao dividendo mínimo obrigatório nos termos do artigo 37 do Estatuto Social da Z, resultando em um valor por ação de R\$0,7223 (setenta e dois centavos).”	Fr 69: 19	34	0	0	0
	Fr 70: 21	41	1	1	0
	Fr 71: 64	125	0	1	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
[Início da Frase 72] “Dividendos referentes ao exercício 2009: Em 14 de abril de 2010, conforme aprovado na Assembléia Geral Ordinária de 24 de março de 2010, a Companhia distribuiu dividendos referentes ao exercício findo em 31 de dezembro de 2009, no montante de R\$49.598 mil), resultando em um valor por ação de R\$1,5921 (um real e cinqüenta e nove centavos).”	Fr 72: 57	111	0	1	0
[Início da Frase 73] “Juros sobre o capital próprio referentes ao exercício 2010: No dia 21 de dezembro de 2010, a Companhia anunciou aos seus acionistas a deliberação de juros sobre o capital próprio no montante total de R\$27.000 mil, relativos ao exercício de 2010, correspondente a R\$ 0,858252 por ação, tendo feito jus ao provendo os acionistas detentores de ações de emissão da Z no dia 21/12/2010 pagos no dia 19/01/2011.”	Fr 73: 68	139	0	1	0
[Início da Frase 74] “A distribuição é relativa ao exercício 2010 e o montante será imputado ao dividendo mínimo obrigatório nos termos do artigo 37 do Estatuto Social da Z.	Fr 74: 26	59	0	1	0
[Início da Frase 75] “Dividendos referentes ao exercício 2010: propostos pelo Conselho de Administração em 27 de janeiro de 2011, o montante totaliza R\$83.350 mil e será submetido à deliberação da Assembléia Geral Ordinária.”	Fr 75: 30	70	0	2	0
[Início da Frase 76] “GOVERNANÇA CORPORATIVA Novo Mercado: a Z foi a primeira Companhia de software brasileira a aderir a modalidade que atende as melhores práticas de governança corporativa da Bolsa de Valores, Mercadorias e Futuros – (BM&FBOVESPA).”	Fr 76: 34	76	1	0	0
[Início da Frase 77] “Conselho Administrativo: o Conselho de Administração da Z é composto por 5 conselheiros, sendo 4 externos e independentes, segundo as definições do Novo Mercado.” [Início da Frase 78] “A diretoria executiva da Companhia é composta por 29 diretores.”	Fr 77: 24	53	0	0	0
[Início da Frase 79] “A lista com o nome, descrição de cargo e breve currículo dos conselheiros e diretores pode ser encontrada no website www.z.com/ri.”	Fr 78: 10	25	0	1	0
[Início da Frase 80] “Comitê de auditoria: criado em 2007, o grupo auxilia o Conselho na tarefa de conduzir da melhor maneira, e seguindo as melhores práticas de mercado, a execução da auditoria interna e externa da Companhia.”	Fr 79: 21	43	0	0	1
[Início da Frase 81] “Os membros são eleitos pelo Conselho de Administração, encontrando-se pelo menos 8 vezes ao ano e possui, atualmente, 3 membros.”	Fr 80: 34	77	2	0	0
[Início da Frase 82] “Comitê de Gente: instituído na Z com o objetivo de auxiliar o Conselho de Administração na definição da remuneração e benefícios dos diretores e conselheiros.” [Início da Frase 83] “Importante ressaltar que hoje a empresa conta com 4 membros, sendo 2 externos.”	Fr 81: 20	42	0	0	0
[Início da Frase 84] “Arbitragem: Pelo Regulamento do Novo Mercado, e pelo Estatuto Social da Companhia, o acionista controlador, os administradores, a própria Companhia e os membros do Conselho Fiscal devem comprometer-se a resolver toda e qualquer disputa ou controvérsia relacionada ou oriunda a estas regras do Regulamento do Novo Mercado, do Contrato de Participação no Novo Mercado, das Cláusulas	Fr 82: 25	60	0	1	0
	Fr 83: 13	24	0	0	0
	Fr 84: 73	192	1	3	0
	Fr 85: 14	37	1	1	0

Tabela 19

Exemplo de Categorização de Relatório da Administração.

Relatório da Administração de Empresa Aleatória	Palavras	Sílabas	P	N	I
Compromissórias, em especial, quanto à sua aplicação, validade, eficácia, interpretação, violação e seus efeitos, através da arbitragem.” [Início da Frase 85] “Também serão resolvidas por arbitragem as divergências quanto à alienação de Controle da Companhia.”	Fr 86: 52	114	0	2	2
[Início da Frase 86] “Declaração da Diretoria: em conformidade com o inciso VI do Artigo 25 da Instrução CVM nº 480/09, os diretores da Z declararam que discutiram, reviram e concordaram com as opiniões expressas no parecer dos auditores independentes e com as demonstrações financeiras referentes ao exercício social encerrado em 31 de dezembro de 2010.”	Fr 87: 30	74	0	0	0
[Início da Frase 87] “RELACIONAMENTO COM OS AUDITORES INDEPENDENTES	Fr 88: 42	88	0	2	0
A política da Companhia na contratação de serviços não relacionados à auditoria externa com os auditores independentes se fundamenta nos princípios que preservam sua independência.” [Início da Frase 85] “Esses princípios consistem nos padrões internacionalmente aceitos, em: (a) o auditor não deve auditar seu próprio trabalho; (b) o auditor não deve exercer função de gerência no seu cliente; e (c) o auditor não deve gerar conflitos de interesses de seus clientes.”	Fr 89: 77	205	1	2	0
[Início da Frase 89] “Procedimentos adotados pela Companhia, conforme inciso III, art. 2º Instrução CVM nº381/03: A Companhia e suas controladas adotam como procedimento formal, previamente à contratação de outros serviços profissionais que não os relacionados à auditoria contábil externa, consultar os auditores independentes, no sentido de assegurar-se que a realização da prestação destes outros serviços não venha a afetar sua independência e objetividade, necessárias ao desempenho dos serviços de auditoria independente, bem como obter aprovação de seu Comitê de Auditoria.” [Início da Frase 90] “Adicionalmente são requeridas declarações formais destes mesmos auditores quanto à sua independência na realização dos serviços de não auditoria.”	Fr 90: 19	55	1	0	0
[Início da Frase 91] “Além dos serviços de auditoria das demonstrações financeiras anuais foram prestados serviços relacionados com auditoria.” [Início da Frase 92] “Os honorários desses serviços totalizaram R\$363,7 mil, representando 45,7% do total dos honorários relacionados à auditoria externa.”	Fr 91: 15	42	0	0	0
[Início da Frase 93] “AGRADECIMENTOS	Fr 92: 17	48	0	0	0
Agradecemos todos aqueles que de alguma forma fizeram contribuições à Companhia, possibilitando assim o sucesso da Z no mercado brasileiro e latino-americano, em especial reconhecemos as contribuições de nossos participantes, nossos acionistas, nossos clientes e nossos parceiros.	Fr 93: 38	103	1	0	0
TOTAL >>	3.107	6.835	111	74	10

Nota:

a) O nome da empresa foi substituído pela letra “Z”.

Fonte: Dados da Pesquisa (2024).

A seguir, exemplifico o cálculo de legibilidade, por meio do índice Flesch-Br. Primeiramente, se faz necessário identificar: número total de palavras, número total de frases e número total de sílabas, conforme exposto na Tabela XXXX acima e o resumido abaixo:

- Total de palavras: 3.107
- Total de frases: 93
- Total de sílabas: 6.835

O índice Flesch adaptado ao português é calculado da seguinte forma:

$$\text{Índice Flesch Adaptado} = 248,835 - 1,015 \left(\frac{\text{Total de Palavras}}{\text{Total de Frases}} \right) - 84,6 \left(\frac{\text{Total de Sílabas}}{\text{Total de Palavras}} \right)$$

$$\text{Índice Flesch Adaptado} = 248,835 - 1,015 \left(\frac{3.107}{93} \right) - 84,6 \left(\frac{6.835}{3.107} \right)$$

$$\text{Índice Flesch Adaptado} = 248,835 - 1,015 \times 33,4086022 - 84,6 \times 2,19987126 \approx 28,82$$

Com um valor de 28,82, o texto em análise é classificado como de difícil legibilidade, adequado a leitores com nível de Ensino Superior. Abaixo, segue a interpretação automática realizada no R para o índice:

```
> # 9. Exibir resumo geral
> total_frases <- nrow(resultados)
> total_palavras <- sum(resultados$Palavras)
> total_silabas <- sum(resultados$Silabas)
> cat("Total de Frases:", total_frases, "\n")
Total de Frases: 93
> cat("Total de Palavras:", total_palavras, "\n")
Total de Palavras: 3107
> cat("Total de Sílabas:", total_silabas, "\n")
Total de Sílabas: 6835
> # 10. Cálculo do Índice Flesch Português
(...)
> # 12. Exibir Índice Flesch e Interpretação
> cat("Índice Flesch:", round(indice_flesch, 2), "\n")
Índice Flesch: 28.82
> cat("Interpretação:", interpretacao_flesch, "\n")
Interpretação: Texto difícil de ler (Ensino Superior).
```

Para o cálculo do tom no texto, foi adotada a classificação de palavras em categorias, como “positivas” e “negativas”, conforme o dicionário de Loughran-McDonald, adaptado para o português.

Além disso, aplicou-se o procedimento de *stemming*, com o objetivo de uniformizar variações lexicais por meio da redução das palavras aos seus respectivos radicais. A análise contextual foi complementada pela geração de n-gramas

(*unigramas*, *bigramas* e *trigramas*), possibilitando a captura de expressões compostas e do contexto em que os termos relevantes estão inseridos. O texto também foi previamente tratado com a remoção de *stopwords*, a fim de eliminar palavras de baixa carga informacional. Por fim, procedeu-se à contagem das ocorrências de cada categoria semântica e ao cálculo do índice de tom com base no modelo proposto por Henry, conforme a seguir:

$$TOM = \frac{Nº\ palavras\ positividas - Nº\ palavras\ negativas}{Nº\ total\ das\ palavras} = \frac{111 - 74}{3.105} = 0,0119085$$