

SatiriCorpus.Br: A Corpus of Satirical News in Brazilian Portuguese

DOI: 10.14393/LL63-v41s-2026-07

Gabriela Wick-Pedro^{1*}

RESUMO: A sátira constitui um gênero discursivo que recorre ao humor, à ironia e ao exagero como estratégias para criticar práticas e valores sociais, culturais e políticos. Dentre suas manifestações contemporâneas, destacam-se as notícias satíricas, que, embora fictícias e voltadas ao entretenimento, podem ser interpretadas como verídicas, contribuindo inadvertidamente para a circulação de desinformação. Diferentemente das chamadas *fake news*, sua intenção principal reside na provocação reflexiva e no riso. Este estudo propõe uma análise linguística de textos satíricos em língua portuguesa, com ênfase nas propriedades lexicais, sintáticas e semânticas. Para isso, foi construído um corpus com base em publicações do site *Sensacionalista*, visando identificar padrões linguísticos que subsidiem, em pesquisas futuras, a detecção automática desse tipo de conteúdo no âmbito do Processamento de Linguagem Natural (PLN).

PALAVRAS-CHAVE: Sátira. Notícias satíricas. Análise Linguística. Desinformação. PLN.

ABSTRACT: Satire constitutes a discursive genre that employs humor, irony, and exaggeration as strategies to critique social, cultural, and political practices and values. Among its contemporary manifestations, satirical news stands out although fictional and intended for entertainment; such texts may be interpreted as factual, inadvertently contributing to the spread of misinformation. Unlike so-called fake news, their primary purpose lies in provoking reflection and eliciting laughter. This study presents a linguistic analysis of satirical texts in Brazilian Portuguese, focusing on lexical, syntactic, and semantic features. To that end, a corpus was compiled from publications on the website *Sensacionalista*, aiming to identify linguistic patterns that may support future research on the automatic detection of this type of content within the scope of Natural Language Processing (NLP).

KEYWORDS: Satire. Satirical news. Linguistic analysis. Misinformation. NLP.

^{1*} Doutora. Universidade Federal de São Carlos. ORCID: 0000-0002-7332-4482. E-mail para contato: gwpedro(AT)estudante.ufscar.br.

1 Introdução

A sátira constitui um gênero literário e artístico amplamente reconhecido por seu potencial crítico e sua função social. Fundamentada em estratégias discursivas como o humor, a ironia, o exagero e a ridicularização, visa expor incoerências, contestar normas estabelecidas e provocar a reflexão sobre práticas, instituições e valores de ordem social, política, cultural ou individual (Kreuz; Roberts, 1993; Simpson, 2003; Attardo, 2014). Além disso, ao longo da história, tem se manifestado por meio de diferentes linguagens e suportes, desde obras da tradição clássica até produções midiáticas contemporâneas, como peças teatrais, charges, tiras cômicas, programas televisivos e, mais recentemente, veículos jornalísticos satíricos em ambientes digitais (Lamarre; Landreville; Beam, 2009; Singh, 2012). Embora o riso figure como marca distintiva do gênero, é sua dimensão crítica que o insere no debate público, ao desestabilizar discursos hegemônicos e evidenciar contradições frequentemente invisibilizadas no cotidiano.

No que diz respeito às notícias satíricas, observa-se um processo de apropriação do formato jornalístico tradicional para a construção de conteúdos fictícios inspirados em eventos reais. Dessa forma, tais textos empregam recursos como paródia, incongruência semântica e ironia para tensionar temas sociais, políticos e culturais contemporâneos, promovendo o riso e, simultaneamente, incentivando uma leitura crítica da realidade. Contudo, a sutileza com que tais estratégias são articuladas pode levar à ambiguidade interpretativa, fazendo com que parte do público intérprete tais notícias como verídicas. Esse problema se agrava no ambiente digital, especialmente nas redes sociais, onde o consumo fragmentado e veloz de informações muitas vezes inibe uma leitura atenta e contextualizada (Rubin et al., 2016).

Diante desse cenário, a desinformação torna-se um fenômeno ainda mais complexo. Embora as notícias satíricas não tenham como objetivo enganar deliberadamente, sua semelhança estrutural com o jornalismo factual pode contribuir, de modo involuntário, para a circulação de informações equivocadas. Por conseguinte, cresce o interesse de pesquisadores das áreas da linguística computacional e do jornalismo em desenvolver métodos capazes de identificar automaticamente esse tipo de conteúdo. Nesse sentido, tais esforços envolvem a aplicação de técnicas de Processamento de Linguagem Natural (PLN) e

Aprendizado de Máquina (AM), que buscam detectar padrões linguísticos e discursivos típicos da sátira (Burfoot; Baldwin, 2009). No entanto, a identificação automática da sátira não é uma tarefa trivial, pois sua interpretação depende frequentemente de conhecimentos contextuais, culturais e pragmáticos por parte do leitor. Estudos como o de Carvalho et al. (2020), sobre o português europeu, indicam que o efeito satírico frequentemente decorre da combinação inesperada de domínios semânticos distintos: uma característica que a distingue de outros gêneros irônicos, como o sarcasmo.

Com base nessas discussões, este estudo tem como objetivo investigar os traços linguísticos característicos da sátira jornalística no português brasileiro. Para tanto, a pesquisa parte da construção do SatiriCorpus.Br, um corpus composto por 5.048 notícias satíricas extraídas do site Sensacionalista, organizadas em categorias temáticas como política, comportamento e entretenimento. Além disso, a partir desse corpus geral, foi elaborado um *subcorpus* analítico, contendo 150 textos satíricos emparelhados com 150 notícias reais correspondentes, com vistas a uma análise comparativa. Por fim, a proposta é contribuir com subsídios teóricos e empíricos para o avanço de pesquisas voltadas à detecção automática da sátira, bem como para a compreensão de seus mecanismos linguísticos e discursivos no contexto da desinformação.

2 Conteúdo enganoso e notícias satíricas

O fenômeno da desinformação manifesta-se de diversas formas, incluindo boatos, conteúdo fabricado, manipulações informacionais e textos satíricos. Embora a sátira não tenha a intenção de enganar, sua natureza ambígua pode gerar mal-entendidos, contribuindo, de forma não intencional, para a disseminação de informações falsas. Diante dessa complexidade, diversos estudos têm buscado compreender as particularidades da sátira dentro do universo dos conteúdos enganosos, abordando os desafios envolvidos em sua detecção e classificação, além dos impactos sociocomunicativos decorrentes.

2.1 Natureza e função das notícias satíricas

As notícias satíricas são textos ficcionais que parodiam o gênero jornalístico e abrangem temas variados como política, cultura, entretenimento e atualidades, utilizando

estratégias como absurdo, incongruência e ironia para produzir humor e crítica social. Embora sejam cômicas, essas narrativas influenciam diretamente a percepção dos indivíduos sobre a sociedade e suas instituições, especialmente quando amplamente compartilhadas em ambientes digitais (Rubin et al., 2016). Nesse sentido, Ermida (2012) destaca que a sátira noticiosa ultrapassa fronteiras midiáticas, sendo difundida por revistas, programas televisivos, sites especializados e até personagens fictícios em redes sociais.

Partindo dessa concepção, autores como Brugman et al. (2022a) caracterizam as notícias satíricas como um gênero híbrido que mescla elementos de comédia, notícia e opinião política. Programas como *The Daily Show* e *The Onion* exemplificam essa combinação, oferecendo críticas bem-humoradas da atualidade. Conforme apontado pela meta-análise de Burgers e Brugman (2021), os efeitos da sátira variam conforme o estilo humorístico, o contexto temático e o perfil do público, sendo uma fonte de informação relevante especialmente entre jovens adultos (DROOG, BURGERS E STEEN, 2020). Baym (2005, p.273) argumenta que essa força cultural e política da sátira decorre justamente de sua natureza híbrida, que promove "notícias que divertem" e desafia os moldes tradicionais do jornalismo.

Skalicky et al. (2021) destacam que as notícias satíricas funcionam como uma forma influente de informação pública, ideia que se apoia em Baym (2005), para quem a força desse gênero deriva da integração entre crítica, humor e informação política em formatos acessíveis. A proliferação de portais satíricos internacionais, como *The Onion* (EUA), *The Beaverton* (Canadá), *The Civilian* (Nova Zelândia) e *The Daily Mash* (Reino Unido), evidencia sua presença marcante no ambiente digital (Brugman et al., 2022b). Nesse contexto, Burgers e Brugman (2021) ressaltam que esse tipo de conteúdo é uma importante fonte alternativa de informação sobre eventos atuais, com impactos tanto informativos quanto persuasivos, justificando o crescente interesse acadêmico em compreender suas características e diferenciações em relação às notícias convencionais, especialmente diante do uso de enquadramentos mais emocionais, negativos e exagerados em programas como *Last Week Tonight* e *Saturday Night Live*.

2.2 Sátira e desinformação: intersecções e ambivalências

Embora seja construída com propósitos humorísticos e críticos, a sátira jornalística pode, em certos contextos, ser interpretada equivocadamente como conteúdo factual. Como discutido, trata-se de um gênero híbrido que mescla elementos de informação, entretenimento e opinião política, construído a partir da paródia de eventos e formatos noticiosos. No entanto, quando esses textos circulam em ambientes digitais, marcados por dinâmicas de compartilhamento veloz, fragmentação de conteúdo e ausência de contexto, seu potencial interpretativo se torna mais ambíguo. A aparência de verossimilhança, o vocabulário jornalístico e o layout gráfico semelhante ao de portais de notícia convencionais tornam as notícias satíricas passíveis de confusão, sobretudo entre leitores menos familiarizados com os mecanismos da ironia e da paródia.

Nesse cenário, Rubin, Chen e Conroy (2015) investigam a dificuldade que leitores apresentam em distinguir entre sátira, notícia real e conteúdo fraudulento, especialmente quando o texto satírico adota um estilo jornalístico convencional. Para enfrentar esse desafio, os pesquisadores exploram padrões linguísticos e estruturais recorrentes em textos satíricos, como construções absurdas e exageros semânticos, e propõem algoritmos de aprendizado de máquina para detectar esses elementos de forma automática. Ao articular aspectos estilísticos e pragmáticos da linguagem com técnicas computacionais, o estudo contribui para o desenvolvimento de sistemas mais precisos de verificação de conteúdo.

Complementarmente, Wardle e Derakhshan (2018) oferecem uma das classificações mais influentes sobre o ecossistema da desinformação, distinguindo três categorias fundamentais: *misinformation* (informação falsa compartilhada sem intenção de causar dano), *disinformation* (informação falsa produzida intencionalmente para enganar) e *malinformation* (informação verdadeira divulgada fora de contexto ou de modo a causar prejuízo). Embora a sátira, em sua origem, não se encaixe na categoria de desinformação deliberada, sua circulação descontextualizada pode gerar efeitos semelhantes aos de conteúdos enganosos. Esse risco é especialmente preocupante nas redes sociais, onde a ausência de pistas contextuais e o alto grau de replicabilidade favorecem leituras equivocadas.

Por sua vez, Tandoc, Lim e Ling (2018) propõem uma matriz teórica que cruza dois eixos, veracidade do conteúdo e intenção de enganar, para classificar diferentes formas de

conteúdo informacional. Com base nesse modelo, a sátira é posicionada como um tipo de conteúdo geralmente baseado em fatos e com baixa intenção de engano, mas altamente suscetível à má interpretação. A principal contribuição desse enquadramento reside no reconhecimento da desinformação como um espectro contínuo, no qual determinados gêneros liminares, como a sátira, desafiam classificações rígidas e exigem abordagens interpretativas mais nuançadas.

Assim, os estudos contemporâneos convergem para uma compreensão mais refinada da sátira enquanto conteúdo ambíguo, que pode, mesmo sem intenção, alimentar o ecossistema da desinformação. Em contextos de déficit de letramento midiático e informacional, esse tipo de confusão se torna ainda mais recorrente. Diante disso, torna-se fundamental combinar o desenvolvimento de ferramentas computacionais com iniciativas educativas que promovam uma leitura crítica, contextualizada e sensível às marcas linguísticas da sátira. Essa articulação entre tecnologia e educação se apresenta como um caminho promissor para o fortalecimento da resiliência informacional da sociedade.

3 O SatiriCorpus.Br

O *SatiriCorpus.Br* constitui um *corpus* desenvolvido com o propósito de fornecer uma base empírica robusta para o estudo das notícias satíricas em português brasileiro, com ênfase particular no contexto político. Essa ferramenta foi concebida para viabilizar análises comparativas entre textos satíricos e suas correspondentes versões jornalísticas legítimas, possibilitando o exame aprofundado das especificidades linguísticas, estilísticas e discursivas desse gênero.

Nesse sentido, o *corpus* completo abarca um total de 5.048 notícias satíricas, organizadas em cinco categorias temáticas principais: Comportamento, Entretenimento, Esporte, Mundo e País. A partir desse repertório amplo, foi construído um *subcorpus* analítico, composto por 300 textos, sendo 150 notícias satíricas e 150 notícias reais, selecionados especificamente para o estudo da temática política, dada a proeminência desse tema na produção de sátiras jornalísticas brasileiras.

3.1 Construção do corpus

A criação do *corpus* iniciou-se com a seleção criteriosa de notícias satíricas em português brasileiro, com ênfase nos textos extraídos do site *Sensacionalista*. Este portal é amplamente reconhecido no Brasil por simular o formato típico das notícias jornalísticas, empregando o humor como recurso para criticar diversos temas, especialmente nas áreas política e de entretenimento. Portanto, a escolha do *Sensacionalista* como fonte principal justifica-se não apenas pela vasta produção de conteúdo satírico, mas também pela relevância cultural e social de suas publicações, que refletem e influenciam a percepção pública sobre questões contemporâneas.

Partindo dessas observações, optou-se por focalizar a categoria “País”, que concentra predominantemente temas políticos, alinhando-se diretamente ao objetivo desta pesquisa de investigar as características linguísticas e discursivas da sátira no âmbito político. Tal delimitação permite um aprofundamento mais preciso na análise, dada a importância e a frequência com que o tema político é abordado pelas notícias satíricas brasileiras.

Para viabilizar a coleta de dados, foi desenvolvido um *crawler* automatizado, responsável por extrair os textos diretamente das páginas do *Sensacionalista*. Durante esse processo, foram eliminados elementos não textuais, como etiquetas HTML, imagens e *scripts*, assegurando que os arquivos gerados, no formato JSON, contivessem exclusivamente o corpo textual das notícias. Importa destacar que a extração ocorreu em janeiro de 2019 e abarcou o período compreendido entre o início das postagens no site, em 2016, até o final de 2018, garantindo assim uma amostra representativa da produção satírica naquele intervalo temporal.

Como resultado dessa etapa, compilou-se o *SatiriCorpus.Br*, composto por 5.048 notícias satíricas distribuídas em cinco categorias temáticas: i) Comportamento, ii) Entretenimento, iii) Esporte, iv) Mundo e v) País. A categorização seguiu a estrutura organizacional original do portal, o que confere ao *corpus* uma fidelidade intrínseca à segmentação temática adotada pelos produtores de conteúdo. Para melhor compreensão, a Tabela 1 apresenta a distribuição quantitativa das notícias, bem como o número de *tokens* e sentenças para cada categoria, evidenciando a dimensão e a densidade textual do *corpus*.

Tabela 1 – Característica do *corpus*.

Categorias	Nº de notícias	Tokens	Sentenças
Comportamento	56	8.014	267
Entretenimento	1.406	255.972	10.720
Esporte	738	120.332	4.931
Mundo	1.001	178.185	8.250
País	1.847	316.588	12.347
Total	5.048	879.091	36.515

Fonte: elaborada pelo autor.

Esse detalhamento quantitativo é fundamental para situar a amplitude do *corpus* e fundamentar as análises subsequentes, sobretudo no que tange ao *subcorpus* político, que concentra a maioria do material textual, refletindo a centralidade do tema político na sátira jornalística brasileira contemporânea.

3.2 Criação do *subcorpus*

Seguindo as recomendações de Rubin, Chen e Conroy (2015) para a construção de *corpora* de notícias falsas, que ressaltam a importância de alinhar notícias satíricas com correspondentes verdadeiras para análise comparativa, o *corpus* inicial foi segmentado em um *subcorpus*. Esse subconjunto é composto por 150 notícias satíricas da categoria "País" e suas respectivas 150 notícias reais.

A seleção das notícias verdadeiras foi realizada por meio da extração de palavras-chave das sátiras, que serviram para buscar reportagens em portais jornalísticos tradicionais e confiáveis, garantindo a correspondência temática e a autenticidade das fontes. Essa abordagem evitou a inclusão de conteúdos duvidosos e permitiu o estabelecimento de pares textuais consistentes para análise.

As palavras-chave foram extraídas a partir dos elementos temáticos mais frequentes nos títulos e primeiras sentenças das notícias satíricas. Entre os termos mais recorrentes estiveram: *governo, Congresso, eleições, polícia federal, lava-jato, presidente, ministro, STF, partido, economia, protestos e corrupção*. Esses itens foram utilizados como filtros iniciais para

localizar notícias reais que abordassem os mesmos tópicos centrais, assegurando a correspondência temática entre cada par satírico-real.

As notícias reais foram obtidas em portais jornalísticos amplamente reconhecidos pela credibilidade editorial, incluindo *G1*, *Folha de S. Paulo*, *Estadão*, *UOL Notícias*, *BBC Brasil* e *CartaCapital*. A consulta a diferentes veículos buscou garantir diversidade editorial e confiabilidade das informações selecionadas

Assim, a escolha da categoria "País" justifica-se pelo foco em temas políticos brasileiros, que apresentam elevado potencial satírico e facilitam a identificação de notícias reais correspondentes. Esse enfoque permite comparar os discursos satírico e jornalístico de forma mais precisa, contribuindo para a investigação dos padrões linguísticos e discursivos próprios de cada gênero. A Tabela 2 apresenta as características quantitativas do subcorpus, incluindo o número de tokens e sentenças nas notícias satíricas e reais:

Tabela 2 – Característica do *subcorpus*.

Notícias	Tokens	Sentenças
Satíricas	22.963	1.212
Reais	107.133	5.721

Fonte: elaborada pelo autor.

Embora o número de tokens seja maior nas notícias reais, a análise linguística deste estudo foi realizada exclusivamente sobre as notícias satíricas do *subcorpus*. O emparelhamento em número de textos teve como objetivo garantir equivalência temática, não simetria textual. Assim, o desequilíbrio de extensão não afeta os resultados apresentados, uma vez que nenhuma das métricas quantitativas dependeu do volume total de tokens das notícias reais.

Por fim, os dados do *subcorpus* foram processados utilizando as bibliotecas NLTK e spaCy, ferramentas amplamente empregadas no campo do processamento de linguagem natural para a análise e manipulação eficiente de grandes volumes textuais.

4 Características linguísticas

Para compreender o processo de construção das notícias satíricas, torna-se fundamental analisar como o receptor identifica e interpreta os recursos linguísticos presentes na estrutura desses textos. Importa destacar que o objetivo desta pesquisa não é simplesmente apontar uma característica linguística pré-estabelecida, mas sim estabelecer um conjunto de características a partir de uma análise descritiva das notícias satíricas. Portanto, o estudo busca não apenas identificar elementos linguísticos, mas também observar como o efeito satírico se manifesta e pode ser classificado em categorias linguísticas específicas.

Nesse sentido, para alcançar essa compreensão, a análise concentrou-se no exame detalhado do sentido figurado, reconhecendo que ele se constrói por meio de diversas características linguísticas e retóricas, que se agrupam em diferentes categorias funcionais. Para isso, foi realizada uma análise minuciosa ao nível da sentença, considerando cada uma das 150 notícias satíricas do *subcorpus*.

Dessa forma, essa abordagem metodológica possibilita a identificação sistemática dos mecanismos linguísticos que sustentam o efeito satírico, contribuindo para um entendimento mais aprofundado dos aspectos discursivos e estilísticos que diferenciam a sátira de outros gêneros textuais.

4.1 Características lexicais

Inicialmente, é importante destacar que a escolha lexical desempenha um papel crucial na identificação de notícias satíricas. Como apontam Burfoot e Baldwin (2009) e Carvalho et al. (2020), o uso de uma linguagem mais informal pode indicar que o texto pertence ao gênero satírico, uma vez que os meios de comunicação tradicionais raramente adotam uma linguagem coloquial. Por exemplo, em (1), observamos a utilização de **palavras pejorativas** com a função de ofender ou ridicularizar alguém, o que não é característico do jornalismo convencional:

(1) *Coxinhas e mortadelas* foram vistos chorando com cara de patos.

Além disso, o emprego de **vulgarismos** ou palavras inadequadas pode reforçar essa suspeita. No exemplo (2), a expressão ‘foda’ sugere um desvio das normas formais de linguagem, indicando que o texto provavelmente não é uma notícia factual:

(2) Paulo é tão *foda* que Despacito é que fica com Paulo na cabeça.

Outra característica lexical relevante é a marca explícita de **autoria** satírica. Essa, por sua vez, é uma das evidências mais concretas para a identificação da sátira, já que a menção à fonte pode facilmente revelar a natureza ficcional do conteúdo. No exemplo (3), a referência ao *Sensacionalista* dificulta a interpretação do texto como uma notícia verdadeira:

(3) *Sensacionalista* pode encerrar atividades após Temer não provar que está vivo.

Adicionalmente, a utilização de **jogos de palavras** é uma estratégia comum em textos satíricos, pois cria efeitos de humor por meio da ambiguidade sonora e semântica. No exemplo (4), o duplo sentido gerado pelas palavras semelhantes produz um efeito cômico inesperado, típico do gênero:

(4) Ela revelou já ter comido *Al Caparras*, mas negou ter assistido ao filme *Al Capone*.

Complementando esses recursos, o uso de **emoticons** ou **emojis** surge como um elemento paralinguístico distintivo. Esses símbolos são raramente encontrados em textos jornalísticos tradicionais, mas aparecem com frequência em notícias satíricas, funcionando como pistas adicionais para o leitor. No exemplo (5), os emojis de riso e as risadas escritas indicam o tom humorístico do texto:

(5) Continuação do filme da Lava Jato vai se chamar “A Lei é para Todos 🤔🤔🤔 kkkkk”

Por fim, as **repetições lexicais** são empregadas para enfatizar determinados termos, contribuindo para o efeito de ironia ou sátira. O exemplo (6) ilustra esse uso, reforçando a crítica ao repetir o nome do senador de forma irônica:

(6) O senador *Aécio Neves* está animado para substituir *Aécio Neves* no PSDB.

Assim, esses diferentes aspectos lexicais atuam em conjunto para compor o efeito satírico, auxiliando tanto na construção do humor quanto na sinalização da não veracidade do texto.

4.2 Característica sintática

Para aprofundar a identificação de indícios que caracterizam notícias satíricas, esta pesquisa também dedicou atenção à análise da estrutura sintática das frases, com ênfase particular na **troca do sujeito**. Inicialmente, observa-se que, em notícias reais, o sujeito geralmente corresponde ao agente da ação de forma lógica e convencional. Por exemplo, no exemplo (7), extraído de uma notícia factual, Marcela Temer é o sujeito responsável pela ação de pular no lago para salvar seu cachorro:

(7) *Marcela Temer* pula em lago para salvar seu *cachorro* e afasta segurança que não ajudou.

Em contrapartida, nas notícias satíricas, essa relação tradicional entre sujeito e ação é subvertida para gerar efeito cômico e absurdo. No exemplo (8), extraído de uma notícia satírica, o sujeito da frase é inusitadamente atribuído ao cachorro, que se torna o agente da ação, criando uma situação improvável e humorística:

(8) *Cachorro de Marcela* se jogou no lago por ter que conviver com Temer.

Dessa forma, a inversão do sujeito não apenas desafia as expectativas do leitor, mas também contribui para o tom humorístico característico do gênero satírico. Essa manipulação sintática evidencia como a estrutura da frase pode ser utilizada para sinalizar a natureza ficcional do texto, ao mesmo tempo em que provoca o riso por meio do absurdo narrativo.

4.3 Características semânticas

As características semânticas desempenham um papel fundamental na criação do efeito humorístico presente nas notícias satíricas, uma vez que o significado vai além do sentido literal das palavras ou expressões utilizadas. Para ilustrar essa ideia, o exemplo (9) utiliza um **memé** para incorporar ideias, informações e expressões populares nas redes

sociais, configurando um elemento cultural digital amplamente difundido e improvável de aparecer em notícias reais. Nesse caso, a expressão “pedir música no Fantástico” remete a uma prática do programa *Fantástico*, da Rede Globo, em que jogadores de futebol podem solicitar uma canção especial ao marcar três gols na mesma partida. Assim, ao empregar essa expressão em um contexto político, o meme cria um efeito humorístico ao sugerir que uma determinada ação grave aconteceu repetidas vezes, de forma leve e irônica.

(9) Temer, se roubar mais uma faixa presidencial, poderá *pedir música no Fantástico*.

De forma análoga, as **expressões idiomáticas** representam outro recurso semântico recorrente, em que um conjunto de palavras adquire um significado distinto daquele que teriam isoladamente. No exemplo (10), a expressão “pagar o pato” é usada metaforicamente para atribuir culpa a alguém que, na verdade, não cometeu o ato em questão:

(10) E até o pato da Fiesp deve pagar o pato.

Além desses recursos, destaca-se o emprego de **palavras fora do contexto** esperado, o que contribui para o efeito satírico ao inserir termos que normalmente pertencem a domínios distintos do tema tratado. Essa estratégia é exemplificada no exemplo (11), onde os termos *Alckmin* e *Lava-Jato* pertencem ao campo lexical político, enquanto *drible* e *Tite* são referentes ao domínio do futebol. No léxico futebolístico, *drible* significa o movimento com a bola para se esquivar do adversário. Entretanto, na sentença satírica, o termo é usado de forma figurada, conferindo o sentido de tentativa de enganar ou ludibriar alguém.

(11) *Alckmin dá um drible na Lava-Jato e é convocado por Tite*.

De modo semelhante, o exemplo (12) apresenta palavras do campo lexical circense, como *Cirque du Soleil* e *malabarismo*, em contraste com *Waack*, que é um jornalista brasileiro. No contexto circense, *malabarismo* refere-se à técnica artística realizada pelo malabarista. Contudo, neste caso, a palavra é aplicada figurativamente para indicar a habilidade em contornar situações difíceis ou adversas, reforçando o efeito humorístico por meio da **ambiguidade** e da associação inesperada entre campos semânticos distintos.

- (12) *Cirque du Soleil* abre seleção no Brasil após exhibições de *malabarismo* em defesa de *Waack*.

Por fim, outro recurso semântico relevante para a construção do efeito humorístico na sátira é a **quebra de expectativa**, frequentemente realizada após conjunções em orações coordenadas ou subordinadas. Quando comparadas as sentenças dos exemplos (13) e (14), percebe-se que o humor emerge da surpresa causada pela continuação inesperada após a conjunção “e”:

- (13) Joaquim Barbosa desiste da disputa presidencial.

- (14) Joaquim Barbosa desiste da presidência e mira o comando do Caldeirão do Huck.

Assim, os recursos semânticos analisados colaboram para a construção do efeito satírico, por meio de estratégias que desviam do sentido convencional e literal, promovendo a ambiguidade, o absurdo e a surpresa, elementos essenciais para a eficácia do humor presente nas notícias satíricas.

4.4 Características estilísticas

As características estilísticas dizem respeito a um conjunto de recursos linguísticos, incluindo figuras de linguagem, empregados para intensificar o efeito humorístico nas notícias satíricas. Entre esses recursos, destaca-se o uso do absurdo como estratégia para expressar situações inesperadas ou inverossímeis, como exemplificado em (15):

- (15) O *papa* negou estar trabalhando pela *canonização* de Lula, como *afirmou o PT*.

Esse tipo de construção absurda, amplamente recorrente nesse gênero textual, costuma ser reforçada por figuras como a **hipérbole**, que intensifica desproporcionalmente os elementos da enunciação. No exemplo (16), observa-se o uso dessa figura para enfatizar de maneira exagerada um evento histórico:

- (16) Com isso, o país se livra *para sempre* da corrupção *endêmica* que marcou sua vida política *desde a chegada dos portugueses em 1500*.

Além da hipérbole, outras figuras de linguagem contribuem para a construção do humor satírico. Um exemplo disso é o **eufemismo**, que suaviza eventos graves ou impactantes com termos mais amenos, criando um contraste irônico, como no exemplo (17):

- (17) Após começar *desempregando Dilma*, Temer atinge a marca de 13,7 milhões sem emprego.

No exemplo (18), observa-se o uso da **metáfora** associada a uma construção irônica que reforça o efeito humorístico do texto.

- (18) Depois de soltar Paulo Preto, Gilmar *lembrou a ele de não se esquecer do casquinho porque a noite estava fria*.

A sentença faz alusão ao episódio em que o ministro do STF, Gilmar Mendes, concedeu habeas corpus a Paulo Preto, condenado na Lava-Jato, utilizando-se de uma metáfora cômica ao compará-lo a uma mãe que recomenda ao filho levar um casaco. Essa associação entre um gesto materno e uma decisão judicial relacionada a um político envolvido em corrupção produz um efeito satírico por meio da incongruência entre a autoridade institucional e o comportamento afetivo, suavizando a gravidade da situação jurídica e, ao mesmo tempo, promovendo uma crítica implícita ao sistema judiciário.

Outro mecanismo estilístico comum é a **personificação**, ou prosopopeia, em que objetos inanimados assumem características humanas. No exemplo (19), a “panela de pressão” é descrita como dotada de medo, o que reforça o tom cômico e surreal da narrativa:

- (19) A *panela de pressão* que tem *medo* de Paulo explodir.

Por fim, destaca-se o uso da **ironia**, figura que se baseia na contradição entre o sentido literal da enunciação e a intenção real do enunciador, frequentemente adotada para provocar efeito de crítica velada ou sarcasmo. No exemplo (20), a ironia está na analogia feita entre a honestidade do político e narrativas mitológicas:

- (20) Crivella diz que *honestidade* de Garotinho é *tão real* quanto *Adão e Eva* e *Arca de Noé*.

Dessa forma, observa-se que os textos satíricos recorrem sistematicamente a construções estilísticas que exploram o nível semântico-pragmático da linguagem, articulando diferentes figuras de linguagem para provocar o riso, gerar estranhamento e estimular a crítica.

A seguir, serão apresentados os resultados quantitativos da análise das características linguísticas descritas, evidenciando sua frequência no corpus e suas implicações para a detecção automatizada de sátira.

5 Resultados

Com base nas características linguísticas descritas na seção anterior, esta seção apresenta os principais achados da análise linguística aplicada ao subcorpus satírico. O objetivo é evidenciar a distribuição e frequência dos recursos utilizados na construção das notícias satíricas, demonstrando empiricamente quais estratégias contribuem de forma mais significativa para a produção do efeito humorístico.

Conforme sistematizado no Quadro 1, as características identificadas foram organizadas em quatro categorias principais (lexicais, sintáticas, semânticas e estilísticas), totalizando 17 características específicas. Essa categorização visa permitir uma leitura mais precisa das estratégias linguísticas envolvidas na criação da sátira jornalística.

Quadro 1 – Características linguísticas das notícias satíricas.

Características linguísticas	Características específicas
Lexicais	Palavras pejorativas (Pej); Vulgarismos (Vul); Autoria (Aut); Jogo de Palavras (JPa); Emoticons/emojis (Emo); Repetição (Rep)
Sintática	Troca do Sujeito (TSuj)
Semânticas	Expressões Idiomáticas (Eld); Palavras fora do contexto de domínio (FdC); Ambiguidade (Amb); Meme (Meme); Quebra de Expectativa (QExp)
Estilísticas	Eufemismo (Euf); Metáfora (Met); Exagero (Exa); Personificação (Pers); Ironia

Fonte: elaborada pelo autor.

A Tabela 3 sintetiza a frequência relativa dos grupos de características linguísticas analisadas. Observa-se que a maior parte dos traços identificados (60,72%) se concentra na categoria semântica, o que reforça sua centralidade na construção do efeito satírico. Em contraste, os recursos de ordem sintática apresentaram ocorrência significativamente menor (0,42%), sugerindo uma participação mais limitada desse nível linguístico na geração do humor característico do gênero.

Para o cálculo das frequências relativas, cada característica específica apresentada no Quadro 1 foi anotada manualmente em nível de sentença para todas as 150 notícias satíricas do subcorpus. Em seguida, contabilizaram-se todas as ocorrências por categoria (lexical, sintática, semântica e estilística). A frequência relativa foi obtida dividindo-se o total de ocorrências de cada grupo pelo número total de ocorrências registradas no conjunto de dados, expressando o resultado em percentual. Esse procedimento permitiu comparar a distribuição das características independentemente do comprimento dos textos.

Tabela 3 – Frequência das características linguísticas

Características	Características específicas
Lexicais	13,58%
Sintática	0,42%
Semânticas	60,72%
Estilísticas	25,26%

Fonte: elaborada pelo autor.

Esses resultados quantitativos corroboram a hipótese de que a sátira, enquanto prática discursiva, se estrutura principalmente por meio de mecanismos *semânticos e pragmáticos*, os quais possibilitam interpretações múltiplas, implícitos irônicos e deslocamentos de sentido. As categorias *estilísticas* e *lexicais* também demonstram importância significativa, sobretudo por seu potencial de acentuar o tom humorístico e crítico. Por outro lado, a baixa incidência de marcas *sintáticas* reforça a ideia de que tais recursos não são os principais vetores da comicidade satírica, ao menos no nível da sentença.

Dessa forma, os dados obtidos indicam que a sátira noticiosa depende fortemente de escolhas linguísticas que transgridem convenções de literalidade, promovendo *incongruências semânticas, hibridismos de registro, e figuras de linguagem* que estimulam a reflexão crítica por meio do riso. A seguir, essas tendências são retomadas em diálogo com a literatura da área, de modo a integrar os achados empíricos aos estudos já consolidados sobre humor, ironia e desinformação.

A literatura sobre sátira a caracteriza como um gênero híbrido capaz de articular crítica e entretenimento. Baym (2005) argumenta que sua força política decorre justamente dessa combinação entre formatos jornalísticos e estratégias humorísticas, o que se alinha às incongruências e aos hibridismos identificados nos dados. De forma semelhante, Skalicky et al. (2021) mostram que a sátira informativa opera por meio de escolhas linguísticas que rompem expectativas de literalidade, produzindo efeitos interpretativos dependentes do contraste entre o dito e o pretendido. Esses achados também convergem com modelos pragmáticos de humor, como a Teoria da Incongruência (Attardo, 1994) e os estudos sobre ironia como prática avaliativa e crítica (Simpson, 2003), reforçando que a eficácia comunicativa da sátira reside na manipulação estratégica de ambiguidades e deslocamentos de registro.

6 Considerações finais

A partir da análise das características linguísticas de notícias satíricas em português brasileiro, este estudo evidenciou a relevância de uma compreensão aprofundada das especificidades desse gênero textual para a construção de mecanismos eficazes de detecção automática de conteúdo enganoso. Nesse sentido, os resultados indicam que recursos lexicais, sintáticos, semânticos e estilísticos desempenham papel central na produção do efeito humorístico e crítico característico da sátira, ao mesmo tempo em que evidenciam os desafios envolvidos na distinção entre sátira e notícia factual.

Entende-se que é importante destacar que a descrição das estratégias linguísticas utilizadas na sátira e, de modo mais amplo, nas formas de linguagem figurada constitui uma tarefa complexa, sobretudo em contextos de aplicação computacional. Isso se deve, na maioria, ao fato de tais recursos estarem ancorados no conhecimento de mundo do leitor e

nos níveis semântico e pragmático da língua. Ainda assim, os achados desta pesquisa apontam para a viabilidade de se desenvolver ferramentas linguístico-computacionais que, ao incorporar tais mecanismos, contribuam para a identificação mais precisa de conteúdos satíricos e potencialmente enganosos.

Com vistas a aprofundar essa linha de investigação, sugerem-se, para pesquisas futuras, a ampliação do SatiriCorpus.Br com a inclusão de notícias satíricas provenientes de outras fontes, bem como a análise de diferentes estilos de escrita e seus impactos na compreensão da sátira. Ademais, propõe-se o fortalecimento do diálogo entre a linguística e a ciência de dados, especialmente por meio da integração entre análises linguísticas e técnicas de aprendizado de máquina e redes neurais, o que poderá aprimorar significativamente os modelos de detecção automática.

Por fim, considerando a escassez de estudos dedicados à análise descritiva do conteúdo satírico no contexto do português brasileiro, este trabalho busca preencher uma lacuna relevante na literatura e, ao mesmo tempo, abrir novos caminhos para pesquisas interdisciplinares que investiguem os usos, os efeitos e os desafios comunicacionais da sátira no cenário contemporâneo.

Agradecimentos

Este trabalho foi realizado no âmbito do Centro de Inteligência Artificial da Universidade de São Paulo (C4AI - <http://c4ai.inova.usp.br/>), com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM. Este projeto também foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei N. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

Referências

ATTARDO, S. (Ed.). **Encyclopedia of humor studies**. Thousand Oaks, CA: Sage Publications, 2014. <https://doi.org/10.4135/9781483346182>.

BAYM, G. The Daily Show: Discursive integration and the reinvention of political journalism. **Political Communication**, v. 22, n. 3, p. 259-276, 2005. <https://doi.org/10.1080/10584600591006492>.

BURGERS, C.; BRUGMAN, B. C. Effects of satirical news on learning and persuasion: A three-level random-effects meta-analysis. **Communication Research**, 2021. <https://doi.org/10.1177/00936502211032100>.

BRUGMAN, B. C. et al. Humor in satirical news headlines: Analyzing humor form and content, and their relations with audience engagement. **Mass Communication and Society**, p. 1-28, 2022a. <https://doi.org/10.1080/15205436.2022.2144747>.

BRUGMAN, B. C. et al. Satirical news from left to right: Discursive integration in written online satire. **Journalism**, v. 23, n. 8, p. 1626-1644, 2022b. <https://doi.org/10.1177/1464884920979090>.

BURFOOT, C.; BALDWIN, T. Automatic satire detection: Are you having a laugh?. In: *Proceedings of the ACL-IJCNLP 2009 conference short papers*. 2009. p. 161-164. <https://doi.org/10.3115/1667583.1667633>.

CARVALHO, P., MARTINS, B., ROSA, H., AMIR, S., BAPTISTA, J.; SILVA, M. J. Situational irony in farcical news headlines. In: **Lecture Notes in Computer Science**. Springer International Publishing, 2020. p. 65–75. https://doi.org/10.1007/978-3-030-41505-1_7.

DROOG, E.; BURGERS, C.; STEEN, GERARD J. How satirists alternate between discursive modes: An introduction of the humoristic metaphors in satirical news (HMSN) typology. **International Journal of Communication**, v. 14, p. 22, 2020.

ERMIDA, I. News satire in the press: Linguistic construction of humour. **Language and humour in the media**, p. 185-210, 2012.

KREUZ, R. J.; ROBERTS, R. M. On satire and parody: The importance of being ironic. **Metaphor and Symbol**, v. 8, n. 2, p. 97-109, 1993. https://doi.org/10.1207/s15327868ms0802_2.

LAMARRE, H. L.; LANDREVILLE, K. D.; BEAM, M. A. The irony of satire: Political ideology and the motivation to see what you want to see in The Colbert Report. **The International Journal of Press/Politics**, v. 14, n. 2, p. 212-231, 2009. <https://doi.org/10.1177/1940161208330904>.

RUBIN, V. L.; CHEN, Y.; CONROY, N. K. Deception detection for news: three types of fakes. **Proceedings of the Association for Information Science and Technology**, v. 52, n. 1, p. 1-4, 2015. <https://doi.org/10.1002/pa2.2015.145052010083>.

RUBIN, V. L., CONROY, N., CHEN, Y.; CORNWELL, S. Fake news or truth? Using satirical cues to detect potentially misleading news. In: **Proceedings of the second workshop on computational approaches to deception detection**. 2016. p. 7-17. <https://doi.org/10.18653/v1/W16-0802>.

SINGH, R. K. Humour, irony and satire in literature. **International Journal of English and Literature**, v. 3, n. 4, p. 63-72, 2012.

SIMPSON, P. **On the discourse of satire: Towards a stylistic model of satirical humour**. Amsterdam/Philadelphia: John Benjamins Publishing, 2003. <https://doi.org/10.1075/lal.2>.

SKALICKY, S., BRUGMAN, B. C., DROOG, E.; BURGERS, C. Satire from a far-away land: Psychological distance and satirical news. **Information, Communication & Society**, p. 1-18, 2021. <https://doi.org/10.1080/1369118X.2021.2014545>.

TANDOC, E. C.; LIM, Z. W.; LING, R. Defining “fake news” A typology of scholarly definitions. **Digital journalism**, v. 6, n. 2, p. 137-153, 2018. <https://doi.org/10.1080/21670811.2017.1360143>.

WARDLE, C; DERAKHSHAN, H. Thinking about ‘information disorder’: formats of misinformation, disinformation, and mal-information. In: IRETTON, Cherilyn; POSETTI, Julie (org.). **Journalism, ‘fake news’ & disinformation**. Paris: UNESCO, 2018. p. 43-54.