Linguistic and emotional dynamics in satirical vs. real news: a psycholinguistic analysis

Gabriela Wick-Pedro[1]**, Roney Lira de Sales Santos**[2]**, Oto Araújo Vale**[3]

[1]Instituto de Ciência, Tecnologia e Inovação Ciência e Tecnologia (IBICT)
Brasília – DF – Brazil

[2]Instituto de Ciência, Tecnologia e Inovação – Universidade Federal da Bahia (UFBA)
Camaçari – BA – Brazil

[3]Programa de Pós-graduação em Linguística – Universidade Federal de São Carlos (UFSCar)
São Carlos – SP – Brazil

`gabrielawick@ibict.br, roneysantos@ufba.br, otovale@ufscar.br`

Abstract. *This study compares the psycholinguistic differences between satirical and real news using data from LIWC (Linguistic Inquiry and Word Count). We found that satirical news utilizes a broader range of emotional and rhetorical resources, often exaggerating or subverting reality, while real news maintains a more factual and objective tone. This highlights the critical and humorous role of satire in social communication. Furthermore, the research advances the field of NLP by improving satire detection through a psycholinguistic lens, contributing to the development of algorithms that effectively differentiate satirical news from fake news and help combat misinformation.*

1. Introduction

The complexity of satire arises from its potential to "reinterpret" factual news, thereby blending real information with fictional elements to convey a subjective message. This creates a significant challenge in distinguishing satirical content from genuine news, especially as satire may incorporate authentic events and details, causing overlap between factual and fictional elements [Burfoot and Baldwin 2009]. The ability to automatically identify satirical news is crucial for developing effective tools to combat misinformation and deceptive content.

Satirical news and fake news are interconnected phenomena that emerge in the context of contemporary communication, particularly on digital platforms. While fake news, or false news, consists of misleading or fictitious information presented as true, with the deliberate intention of deceiving the public for purposes of propaganda, financial gain, or confusion [Rubin et al. 2015], [Wardle and Derakhshan 2017], [Tandoc Jr et al. 2018], satirical news employs humor and irony as tools to provoke critical reflection on social and political issues. This difference in intent is fundamental to understanding how each type of content is perceived and interpreted by readers.

However, the humorous nature of satirical news can easily be lost, leading to confusion between what is real and what is fictional. The widespread sharing of satirical content on social media often obscures the lines between satire and reality, resulting in the inadvertent dissemination of false information when the satire is not correctly identified. This phenomenon highlights the challenge of reader interpretation, especially in

digital environments where the context and intent of the message may not be immediately clear. As noted by [Friggeri et al. 2014], 'the rapid spread of satirical content on social media often blurs the lines between fact and fiction, contributing to the confusion experienced by audiences.' Thus, analyzing the differences between these two types of content is crucial for developing effective strategies for identifying and mitigating misinformation in the digital age.

To address these challenges, this study explores psycholinguistic relationships using data extracted from the LIWC (Linguistic Inquiry and Word Count) [Pennebaker 2001], [Balage Filho et al. 2013]. The LIWC is a computational text analysis tool designed to read each word in a text (or group of words) and match it with reference dictionaries loaded into the program's memory. [Yang et al. 2017] emphasized the importance of psychological characteristics in this context, noting that real news tends to be more conservative, while satirical news is often more aggressive. [Salas-Zárate et al. 2014] identified five significant psychological characteristics for detecting satirical news: social processes, affective processes, positive emotions, cognitive processes, and certainty.

The present research utilizes LIWC data to explore the psycholinguistic differences between satirical and real news content. A corpus was constructed with satirical news articles sourced from the Sensacionalista website, and corresponding real news articles were manually selected from reliable sources. The analysis focuses on key psycholinguistic features within the LIWC framework, examining categories such as family, affect, anger, and anxiety, among others, to differentiate between real and satirical news. By identifying these features, the research seeks to enhance the understanding of satire and improve automated methods for distinguishing satirical content from genuine news.

The article is organized as follows: Section 2 describes the corpus used in the analysis, outlining the collection process and characteristics of both satirical and real news texts. Section 3 presents the main psycholinguistic features extracted by LIWC and discusses their implications. Section 4 provides a detailed analysis of the results, highlighting the key differences between the two types of news. Finally, Section 5 offers the conclusions and discusses potential directions for future research.

2. *Corpus*

Considering the criteria proposed by [Rubin et al. 2015] for a fake news *corpus*, which emphasizes the importance of aligning fake news with true news to validate linguistic patterns through the identification of positive and negative instances, a *corpus* was created consisting of 300 news articles in the political domain, with 150 satirical news articles and 150 true news articles. The satirical news articles were automatically extracted from the Sensacionalista website, an online news outlet that uses the journalistic format to humorously address various current topics in Brazilian politics and entertainment. For the true news articles, the collection was conducted manually, initially identifying keywords from the satirical news articles and subsequently performing a manual search for each true news article equivalent to the satirical one. Table 1 details the characteristics of the *corpus*.

The data used to describe the corpus characteristics mentioned above were gen-

Table 1. Characteristics of the *corpus*

News	Tokens	Types	Sentences
Satirical	22.963	4.843	1.212
Real	107.133	11.304	5.721

erated using NLTK[1] and spaCy[2], two open-source libraries that offer valuable tools for natural language processing in Python. These libraries are widely employed both in production environments and in the development of applications capable of processing and understanding large volumes of text. To avoid losing information during analysis, the decision was made not to balance the corpus, as the number of words, sentences, and lexical diversity may serve as important descriptive features of this type of content. However, in future iterations of this work, a balance between the number of real and satirical texts will be sought, ensuring the integrity of the linguistic data and the representativeness of the categories analyzed."

3. Features Extracted by LIWC

Each category represented in the LIWC reflects a specific aspect of language that can illustrate different dimensions of human experience and communication. For example, categories such as "affect" and "anger" help in understanding how emotions are expressed and manipulated in texts, while categories such as "discrepancy" and "certainty" reveal how texts address and represent controversial issues or absolute certainties.

The analysis of percentage differences between the two news categories indicates patterns that highlight how satirical news often amplifies or subverts aspects of reality to provoke reflection, humor, or social critique. In contrast, real news tends to maintain a more direct and factual approach, reflecting the informative and objective nature of traditional journalism.

Table 2 presents data extracted from the LIWC dictionary related to various psycholinguistic features in real and satirical news. The displayed percentages reflect the relative frequency of terms associated with emotions, mental states, and specific topics in both types of news. This analysis provides insights into how the language used in satirical texts differs from that in conventional news, highlighting the rhetorical and emotional strategies employed to engage the reader and provoke critical reflection.

Table tab:emoliwc presents the frequency of emotions identified in real and satirical news, as categorized by the LIWC dictionary. This table illustrates the distribution of words associated with three emotional categories: positive, negative, and neutral. The provided information enables a comparative analysis of the use of emotional terms in each type of news.

4. Analysis of results

The analysis of the data reveals significant differences in the psycholinguistic relationships between satirical and real news.

---

[1]Available at: `https://www.nltk.org/index.html`.
[2]Available at: `https://spacy.io/`.

Table 2. Psycholinguistic relationships from LIWC extracted data

Features	Real	Satirical
achieve	5,70907483	5,80581862
anger	1,27298184	1,49304667
anx	0,59711395	0,93469965
cause	4,75479901	5,04649774
certain	2,9125677	3,15672724
death	0,33252275	0,39672383
discrep	5,1703802	5,81400334
family	0,595271	0,51880118
feel	1,65985041	1,98788499
friend	0,55472623	0,58740299
hear	1,80262334	2,23103831
home	0,91558896	0,95128402
inhib	6,05080874	6,24946677
insight	5,80394974	5,86980633
leisure	1,306142	1,98788499
money	2,86835415	2,84105452
relig	0,81242401	0,93848648
sad	0,71478575	0,86170122
see	1,17442246	1,04086682
swear	7,02166464	7,07277536
tentat	7,36431966	8,51036601
work	4,05198777	3,50652675

Table 3. Emotion scores of news articles based on the LIWC dictionary

Emotions	Real	Satirical
positive	0.109	0.392
negative	0.024	0.174
neutral	0.003	0,072

Satirical news employs terms related to achievements (5.8%) more frequently than real news (5.7%), suggesting a greater emphasis on exaggerating or highlighting accomplishments. Additionally, 5.046% of words in satirical news and 4.7% in real news fall under the category of "cause".

The presence of words related to affect is higher in satirical news (3.9%) compared to real news (2.7%), indicating an amplification of emotional expressions to create contrast or emphasize absurdity.

The frequency of words associated with anger is greater in satirical news (1.4%)

than in real news (1.2%), which may reflect a tendency to underscore feelings of frustration or indignation.

Satirical news also shows higher levels of words related to anxiety (0.9%) compared to real news (0.5%), possibly to exaggerate concerns.

The use of terms related to "work" is lower in satirical news (3.5%) than in real news (4.0%), indicating a reduced focus on work-related issues. It is observed that satirical texts have higher levels of words related to attempts (8.5%) and certainty (3.1%), but lower levels of words related to work (3.5%). This may suggest that, due to their political domain, real news covers more topics related to political or legal work, while satirical news may be more focused on creating a sense of meaning and expressing possibility and uncertainty. As noted by [Salas-Zárate et al. 2017], the higher levels of certainty words (such as "never" or "always") in satirical news (3.1%) compared to real news (2.9%) may be associated with hyperbole.

The discrepancy is more pronounced in satirical news (5.8%) than in real news (5.1%), suggesting a greater engagement in contesting or ridiculing ideas. Mentions of family are slightly less frequent in satirical news (0.51%) compared to real news (0.59%), possibly due to a greater focus on social or political issues.

Satirical news exhibits a higher frequency of words related to feelings (1.9%) compared to real news (1.6%), indicating a more intense exploration of emotions. The difference in mentions of friendships is minimal between the two types of news, suggesting a similar approach to interpersonal relationships, though satire may include critical nuances.

Words associated with "hearing" are more frequent in satirical news (2.2%) than in real news (1.8%), potentially highlighting aspects of communication or its absence.

Mentions of home are slightly more common in satirical news (0.95%) than in real news (0.91%), with no significant difference in the treatment of the theme.

Inhibition is somewhat more frequent in satirical news (6.2%) compared to real news (6.0%), which may reflect a more intense exploration of limitations. The presence of insights is slightly higher in satirical news (5.86%) compared to real news (5.80%), suggesting a more frequent use of insights to reinforce arguments or jokes.

The higher presence of terms related to leisure in satirical news (1.9%) may reflect a greater focus on trivial or everyday aspects, possibly to critique consumer society. Mentions of money are similar in both types of news, indicating that financial issues are equally relevant, although satire may address them critically.

Religion is mentioned more frequently in satirical news (0.9%) than in real news (0.8%), possibly to explore themes of faith or morality in a critical manner. Sadness is more prevalent in satirical news (0.8%) than in real news (0.7%), indicating a more frequent use of emotional themes to create impact.

Mentions of "seeing" are slightly less common in satirical news (1.0%) compared to real news (1.1%), suggesting a lower focus on direct visual observations. The frequency of swear words is similar in both types of news, suggesting that strong language is a common feature, with satire utilizing it more strategically.

Attempts are more common in satirical news (8.5%) than in real news (7.3%), reflecting a greater use of expressions of uncertainty or speculation. It is also noted that there are no completely divergent categories between real and satirical news. Satirical news features higher levels of words related to attempts, causes, and certainty, but lower levels of words related to work. This may indicate that, due to its political domain, real news covers more topics related to political or legal work, while satirical news may be more associated with creating a sense of meaning and expressing possibility and uncertainty. Additionally, as [Salas-Zárate et al. 2017] point out, the higher levels of certainty words in satirical news (3.1%) compared to real news (2.9%) may be associated with hyperbole. It is also noteworthy that 8.510% of words in satirical news and 7.3% of words in real news fall under the category of attempts, indicating that satirical texts make more use of modalizations, producing a sense of meaning and expressing a possibility and uncertainty.

In addition to the evident presence of affective terms in satirical news, validating that satirical texts are more emotional, the data shows that satirical news employs significantly more positive (0.392) and negative (0.174) words compared to real news. Thus, it can be asserted that the use of emotions (positive, negative, and neutral) is more pronounced in satirical news. It is interesting to consider that the use of emotion may be related to the fact that (1) real news tends to use more impartial language compared to satirical news and (2) the excessive use of positive words may, in fact, reflect the use of resources such as irony, which employs words or sentences with opposite polarity to the literal meaning, as also noted by [Salas-Zárate et al. 2017]. Additionally, the data show that satirical news contains a significantly higher proportion of words associated with positive (0.392) and negative (0.174) emotions compared to real news, which present 0.10% and 0.02%, respectively. Furthermore, the frequency of neutral words is also higher in satirical news (0.07%) compared to real news (0.003%). This difference in proportions may reflect the more emotional and expressive style of satirical news in contrast to the more impartial and factual approach of real news. Analyzing these data can provide insights into how different news genres use emotional language to create impact and engage readers.

5. Conclusions

The analysis of the data reveals striking differences in the psycholinguistic characteristics between satirical and real news. Satirical news employs a broader array of linguistic and emotional resources to create impact and convey social critique. The higher frequency of terms associated with achievements, affect, and attempts, as well as the more intense use of positive and negative emotions, suggests that satire employs exaggeration and subversion of reality to provoke reflection or humor. The pronounced presence of words related to uncertainty, certainty, and discrepancy reflects the nature of satire as a tool for challenging perceptions and questioning social norms. These differences indicate that satirical news not only highlights and amplifies aspects of reality but also utilizes linguistic resources to critique and satirize, offering a perspective that is often more emotive and critical compared to the more impartial approach of real news. Thus, satire emerges as an effective means for exploring and commenting on social and political issues, employing a wider range of discursive strategies to engage the audience and stimulate critical reflection.

Acknowledgements

References

Balage Filho, P. P., Pardo, T. A. S., and Aluísio, S. M. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 215–219.

Burfoot, C. and Baldwin, T. (2009). Automatic satire detection: Are you having a laugh? In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 161–164, Suntec, Singapore. Association for Computational Linguistics.

Friggeri, A., Adamic, L., Eckles, D., and Cheng, J. (2014). Rumor cascades. In *proceedings of the international AAAI conference on web and social media*, volume 8, pages 101–110.

Pennebaker, J. W. (2001). Linguistic inquiry and word count: Liwc 2001.

Rubin, V. L., Chen, Y., and Conroy, N. K. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.

Salas-Zárate, M. d. P., López-López, E., Valencia-García, R., Aussenac-Gilles, N., Almela, Á., and Alor-Hernández, G. (2014). A study on liwc categories for opinion mining in spanish reviews. *Journal of Information Science*, 40(6):749–760.

Salas-Zárate, M. d. P., Paredes-Valverde, M. A., Rodriguez-García, M. Á., Valencia-García, R., and Alor-Hernández, G. (2017). Automatic detection of satire in twitter: A psycholinguistic-based approach. *Knowledge-Based Systems*, 128:20–33.

Tandoc Jr, E. C., Lim, Z. W., and Ling, R. (2018). Defining "fake news" a typology of scholarly definitions. *Digital journalism*, 6(2):137–153.

Wardle, C. and Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe*, 27.

Yang, F., Mukherjee, A., and Dragut, E. (2017). Satirical news detection and analysis using attention mechanism and linguistic features. *arXiv preprint arXiv:1709.01189*.