



DOI: 10.58734/plc-2025-0012



Florencia Reali 

Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Uruguay

Fake News are Easier, More Emotional and Less Complex: Some Evidence from Spanish

Characterizing false information based on linguistic analysis is important to understand the factors that affect the proliferation of fake news in the media. Previous work has identified some linguistic regularities that suggest a trend towards decreased complexity, polarization and sentiment in false information. This study is aimed at identifying linguistic differences between real and fake news using a corpus of annotated media news in Spanish via the automatic analysis of linguistic cues using dictionaries of lexical norms. We focus on lexical aspects of complexity, familiarity and sentiment. Consistent with previous results, we found that fake news are associated with lower cognitive loads, reflected by reduced sentence complexity, and increased lexical familiarity and imaginability. Moreover, and consistently with previous results, the analysis revealed that fake news are associated with more polarized emotional content.

Key words: fake news, news replication, linguistic cues, complexity, emotional content

Address for correspondence: Florencia Reali

Neurociencia y Aprendizaje, Universidad Católica del Uruguay, Uruguay

E-mail: florencia.reali@gmail.com

This is an open access article licensed under the CC BY NC ND 4.0 License.

The term post-truth refers to the effect of messages where the objective facts behind the statements are less important than the subjective reactions awoken in public opinion (McIntyre, 2018). The dissemination of fake news, particularly in the context of social media use, have played a fundamental role in the generalization of post-truth practices. Digital forms of information spreading, distort the messages in echo chambers and filtered bubbles, simply reinforcing the impact of fake news (Graham et al., 2013; Cinelli et al., 2021). However, social media is not the only culprit to blame. Powerful yet natural psychological forces underlie audiences' biases to believe and spread news to which there is no evidence, in an enthusiastic and uncritical manner. In fact, over a century ahead of the phenomenon of digital communication, in a work entitled *Group psychology and the analysis of the ego*, S. Freud (1921), put forward an analysis that remains valid today. Based on the idea that certain behaviors are reinforced or appear when individuals constitute a multitude, Freud examined what he referred to as "the collective soul", proposing that the individual will blurs in collective conditions, becoming an automaton full of enthusiasm and lacking a critical spirit. He described the crowd as extraordinarily impressionable and credulous, lacking critical sense. Rationality, he argued, is diminished because of the mass effect. Taking ideas from sociologist Le Bon, he argued that collectives are excited when exposed to excessive, exaggerated, colorful and repeated stimuli, being unmoved by logical arguments. These ideas remain current over a hundred years later, as simplicity and emotionality are the key ingredients for high impact on hypermodern news consumers.

The Fake News Phenomenon

The phenomenon of fake news has become a matter of permanent discussion and analysis in the media in recent years (D'Ulizia et al., 2021; Moscadelli et al., 2020; Paslawska & Papielska-Borys, 2018; Wu et al., 2022). The issue has been examined from different angles and disciplines including psychology, sociology, media and digital technologies. One of the focuses of analysis involves the importance of the ability to evaluate information and sources critically (Harari, 2019; Lewandosky et al., 2021). Misinformation, particularly in political discourse, constitutes a threat to democracy (Ecker et al., 2024). The spread of false information, not only distorts the nature of opinions and event (defying all methods of truth-checking methods) but also boosts political polarization, undermining trust in institutions and media (Lewandosky et al., 2023).

Paslawska and Papielska-Borys (2018) define fake news as "any incorrect information that has been created intentionally under the guise of a fact". The dissemination of fake news involves cognitively biased reasoning, including automatic and subconscious processing. For example, Lewandowski et al. (2012) argue that the behavior of spreading untruth involves the engaging in two pre-digital processes: the embracement of urban legends and the engagement in individual automatic, subconscious presumptions (prejudices). Studies in automaticity show that implicit associations are created when information is automatically processed, and then we expect associations

to occur unless we consciously flag information as false. In other words, when the reader temporarily accepts the processed information as “probably true” a cognitive link is created automatically. On the other hand, the overthrow of a piece of information as “false” requires the person to mentally articulate and consciously question it beyond human presumptions (Ward, 2019). In other words, automatic cognitive connections constantly reinforce false information, as seeing headlines in our news feed repeatedly, over time, increases the confidence that they are credible.

Social media has become a powerful vehicle for the fast and effective spread of fake news, especially at times of crisis and uncertainty (Spence et al., 2016, Jang & Kim, 2018; Jang et al. 2017). A recent study conducted in Italy showed that links containing fake news represented 23.1% share of all articles reviewed (Moscadelli et al., 2020).

A growing bulk of work in the last decade, suggests that the automatic and emotional nature of news processing posits a number of implications on the nature of “spreadable” information. That is, fabricated news—intentionally created to be appealing to the reader and readily replicated—are expected to be emotionally charged and easier to process. Along these lines, psycholinguists have looked at how the dynamics of information transmission may favor some language traits over others.

In an era of explosive growth of information, computational linguistics has become a crucial tool to make sense of patterns of information distortion and patterns of news spreading (Ruffo et al., 2023). For example, recent computational work has shown that language traits that are easier to process tend to proliferate in large communities of speakers, while complex traits tend to wind down. This suggests that highly shared pieces of information, and possibly other aspects of culture, may become simpler at the structural level as our world becomes increasingly interconnected (Realí et al., 2018).

Accordingly, the main goal of this paper was to identify measurable differences between true and false statements in the Spanish online news. We focus on the aspects of complexity, familiarity and valence. As we see in the next section, a number of studies have shown that fake statements tend to have higher emotional valence and to be less complex (shorter and simpler sentences and words). Our research extends these ideas to the case of Spanish. To do that, we present a quantitative analysis of some psycholinguistic features present in fake news (in comparison with truth-checked ones) using a Spanish corpus. Analysis of fake news’ lexical features have been done using databases mainly in English but not in Spanish. Then, the current analysis was designed to assess whether the regularities that have been observed in fake news in other languages reproduce in a database of Spanish database.

The remaining of this paper is organized as follows. In the next section, we review studies that investigated the kind of linguistic features that make news spread faster. Then, in the following section, we review what psycholinguistic aspects have been documented to differ in true vs false statements. Then, a quantitative analysis is presented, using a publicly accessible database of fake and real news in Spanish to assess whether psycholinguistic features previously documented in other languages are sufficient to differentiate fake from true information. In the final section, results and its implication are discussed.

Spreadability: What Makes News Spread Faster

Users choose the information that consolidates their views, cementing previous beliefs. Differential diffusion of true and false stories has been investigated. Vosoughi, Roy and Aral (2018) investigated the patterns of diffusion of over 126000 verified true and false stories tweeted by 3 million people. They found that false stories diffused significantly farther, faster, deeper, and more broadly than true ones in all categories of information. In addition, false stories inspired fear, disgust, and surprise in their audiences, while true stories inspired anticipation, sadness, joy, and trust.

Some research has looked into the kind of linguistic cues that make news more spreadable. Brady et al. (2017) used a large sample of social media communications about polarizing moral/political issues to show that the presence of moral-emotional words in messages increased their diffusion significantly (by a factor of 20%). In another study, Namkoong et al. (2019) studied diffusion in social media under conditions of uncertainty. They found that, when causal uncertainty is high—that is, when causes of anxiety-producing events are not clear—, people react more positively towards abstract messages and replicate them more. They interpreted this finding as consistent with previous research that shows that abstract thinking helps reduce causal uncertainty.

Many studies have shown that emotional information is key to determine social transmission. In general, emotional content increases the probability and frequency of a message being reposted (Xu et al., 2022). Some have documented that messages conveying extreme negative emotions have been found to be more spreadable (Anbin & Wei-yi, 2018). In another study, Pivecka et al. (2022) examined the role of valence and arousal in the social transmission of over 3000 politicians' messages in English on Twitter. They found that tweets eliciting a negative valence decreased retweet probability, whereas tweets eliciting a high (vs. low) arousal increased retweet probability. The prevalence of high arousal in retweets aligns with previous results. However, they found a bias towards positive emotions transmission, which is at odds with previously reported bias towards negativity.

Linguistic Features in False and True Statements

Recent studies have looked at whether it is possible to differentiate false from true statements looking at linguistic cues in texts. Research looking at this question comes from two different yet overlapping fields of study: social sciences to digital technologies studies. The former aimed at exploring the nature of deceptive behavior and veracity judgement, and the latter aimed at exploring the nature of fake news and its spreading dynamics.

Deception Produced under Experimental Conditions

Linguistic cues to deception have been explored in the aim of understanding psychological aspects of deceptive behavior. Cognitive constraints, particularly cognitive loads, may play a role in the production of false statements under conditions of prompted lying—that is, condition under which participants do not have much time to prepare an elaborated answer. For example, Newman et

al. (2003) suggest that due to cognitive load, “liars” may produce more simple statements rather than complex evaluations and judgments because the former are “easier” to combine into a false narrative. Studies in social science fields typically explore linguistic traits of false and true statements produced by experimental participants (Masip et al., 2012; Sarzynska-Wawer et al. 2022) or by specific groups of people in more natural contexts. For instance, in one study, Larker and Zakolyukina (2012) used a set of criteria to identify severity of accounting problems to label a set of financial restatements in English as “truthful” or “deceptive”. They found that deceptive statements used more references to general knowledge, included less mildly positive emotions but more extreme positive emotions. In a comprehensive meta-analysis, Hauch et al. (2015) analyzed written and transcribed statements from different studies designed to identify and quantify linguistic cues to deception in English. They concluded that deceptive statements include more negative emotion words, distance themselves from events by using fewer self-references (first-person pronouns) and more other-references (second and third-person pronouns). Moreover, and consistently with the idea that false statements are associated with a reduced cognitive load (Newman et al., 2003), they found that liars construct shorter statements (fewer words and fewer sentences), which are more concise, less elaborated and less complex.

A few studies have analyzed the language of false and true statements in languages different from English. For instance, Schelleman-Offermans and Mercklerbach (2010) analyzed statements in Dutch, failing to confirm differences in complexity between true and false statements as measured by the use of motion verbs. They also found no differences in the use of emotionally negative terms. However, false statements differed from true ones only in their less frequent use of exclusive words. In another study, Masip et al. (2012) analyzed Spanish handwritten studies and judged other participants’ stories veracity. They used the Spanish translation of the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Francis, & Booth, 2001; Ramírez-Esparza, Pennebaker, García & Suria, 2007) to identify and quantify linguistic cues to deception. LIWC software that analyzes transcripts with a dictionary-based approach in which each word is compared against a file of words divided into linguistic dimensions. They found that true statements contained more preposition than false ones, while false statements contained more positive emotions and affective processes.

In a recent study, Sarzynska-Wawer et al. (2022) studied the extent to which true and false statements in Polish differed linguistically in terms of complexity and sentiment. Specifically, they investigated whether linguistic features that differentiate true and false utterances in English—namely utterance length, concreteness, and particular parts-of-speech—are also present in the Polish language. After analyzing 1,500 true and false statements, they concluded that false statements are less complex in terms of vocabulary—more concise and concrete—and include words that are more positive.

Taken together, it can be concluded that false statements created under conditions of prompted lying (typically under experimental conditions) are less complex, and presumably easier to process. It should be noted, though, that these are statements often created under time-constrained conditions and where the interlocutor's knowledge must be taken into account. Under these conditions, higher cognitive loads associated with fabrication pressure may strongly influence the outcome (Newman, 2003). A different question, however, is whether fake news fabricated with less or no time restrictions and with an intention to be appealing to an audience share similar psycholinguistic features.

Psycholinguistic Characteristics of Fake News in the Media

The second source of literature comes from the field of digital technologies aimed at developing automatic fake news detectors. These studies typically use banks of fake vs real news to train classifiers based on linguistic cues. Fake news detection has gained increasing importance among the research community in recent years due to the widespread diffusion of false information through media platforms. Annotated datasets have been released in the last few years, aiming to assess the performance of fake news detection automatic methods (see D'Ulizia et al., 2021 for a recent review).

What are the linguistic features that differentiate fake from real news? Intentional factors should not be underestimated. Deception present in fake news constitutes intentional manipulation aimed at social media consumers, sometimes targeted at shaping political views. Producers of fake news may prepare content to favor fast spreading. Then, we should expect fake messages to be simpler and richer in emotional content to boost appeal. Support for this idea comes from many studies. For example, fake news detection studies looking at readability—a measure of the extent to which a text is easy or enjoyable to read—, confirmed that fake news tend to be less complex (Pérez-Rosas et al., 2018; Santos et al., 2020).

More generally, a range of different stylometric features associated to sentence complexity (e.g., sentence length, word frequencies, word length) and features associated to semantic content (imaginability, concreteness, arousal) characterize statements in fake news (see survey on Kumar & Shah, 2018). False statements include shorter sentences (Mukherjee, 2013), more words with higher imaginability (Ott et al., 2013), more exaggerations (Ott et al., 2013). In terms of sentiment, false information was found to include more negative high emotions (Ott et al., 2013) and to be more polarized and sentimental (Li et al., 2014). In a recent study, Galletta Horner et al. (2021), built on economics of emotion theories to explore the idea that fake news headlines are created to evoke emotional responses in readers that will cause them to interact with the article, a hypothesis empirically supported by their behavioral study.

Fake news are typically fabricated to manipulate the reader's emotions. Ghanem et al (2021) showed the importance of flow of affective information in fake news, using neural network models trained to detect false news. In another study, Ghanem et al. (2020) compared false and real news, showing that false information has different emotional patterns in propaganda, hoax, clickbait and satire articles, concluding that

emotions play a role in deceiving readers. Others have studied whether emotional information affects credibility. For example, Using LSTM models, Giachanou et al (2019) showed that emotional signals are useful cues to differentiate credible from non-credible news, using real world datasets, showing that emotional signals are cues are important in credibility assessment.

Taken together, studies looking at stylistic features of fake news and false statements, suggest that false information tends to be easier to process and likely to boost emotional reactions. False information contains shorter sentences (Mukherjee, 2013), shorter words with higher imaginability scores (Ott et al., 2013), and more concise vocabulary (Sarzynska-Wawer et al, 2022). Another indication that false statements may be “simpler” or easier to process is the lower number of prepositions in false statements (Masip et al., 2012), as a higher number of prepositions in an utterance suggests syntactic complexity (Shaffer, 2022). The second main pattern that emerges from different studies is that false information tends to be more polarized and sentimental (Kumar & Shah, 2018). False statements tend to present more extreme emotional content, both negatively (Ott et al, 2013) and positively polarized (Masip et al., 2012; Larker & Zakolyukina, 2012).

Most of the work on fake news, however, examines online media written in English. In order to fully understand the nature of misinformation, it is important to look at fake news in different languages. Previous studies have explored psycholinguistic features of deceptive statements in Spanish created under controlled conditions in the context of deception detection (e.g., Masip et al. 2012). However, these features have not been explored in fake news naturally occurring in Spanish media.

Accordingly, the main goal of this paper was to identify differences between true and fake news using a corpus of annotated media news in Spanish, a resource has been publicly available to analyze deceptive information (Aragón et al., 2021; Gomez-Adorno et al., 2021; Posadas-Duran et al., 2019). The focus will be on lexical aspects of sentiment and complexity. Consistently with previous results, we expect to find that fake news has easier-to-process content reflected in a number of cues, such as reduced sentence length, reduced number of prepositions, and higher familiarity and imaginability scores at the lexical level. In addition, we expect fake news to be more sentimental and emotionally polarized.

Method

Data

To test whether false and true statements in Spanish differ in terms of linguistic cues, we used a database of fake and true news. Data for analysis were taken from The Spanish Fake News Corpus Version 1.0 and Version 2.0, provided by Posadas-Durán et al. (2020), a free resource available at <https://github.com/jpposadas/FakeNewsCorpusSpanish>. The corpus consists in a series of news from various websites annotated with two labels (real and fake), which enables the automatic analysis of linguistic cues using various dictionaries of lexical norms. Version

1.0 was used for the Fake News Detection Task in the MEX-A3T competition at the IberLEF 2020 congress (see Posadas-Durán et al., 2020 for details) and it consisted on 971 news taken from different news websites in Spanish. From these, 480 were manually annotated as fake and 491 were annotated as real. Websites were taken from 21 news Spanish sources from Spain, Mexico, Argentina, US and England.

The Version 2.0 was used for the Fake News Detection in Spanish (FakeDeS) shared task at the IberLEF 2021 congress, and contains pairs of fake and true publications collected from November 2020 to March 2021 from newspapers and media companies websites and fact-checking websites. Version 2.0 comprised 572 news, and was balanced with respect to false and real news.

For the purpose of this study, Version 1.0 and 2.0 were merged, and sentences were taken as the unit for analysis, in order to consider complexity at this level. Specifically, we measure sentence-length and number of prepositions per sentence as a proxy for syntactic complexity (Shaffer, 2022). Tabs in the news database were taken as end of sentence markers. After segmenting the corpus in utterances, news labeled as fake comprised 6381 sentences and news labeled as true contained 9229 sentences for analysis. In average, fake news contained 8.33 sentences per story, while real news contained 11.87 sentences per story.

Analysis of Data

We conducted an automatic analysis aimed at identifying linguistic differences between true and false news in the Spanish corpus described. The analysis contrasted the relative prevalence of linguistic cues in fake vs real news using lexical norms values taken from available dictionaries (see next subsection). Codes for analysis were implemented in Perl programming using the corpora texts as input. Sentences in news were assigned values for linguistic features of interest. We focused on lexical aspects of complexity and sentiment; specifically stylometric features of words, proxies of syntactic complexity and semantic cues. After coding sentences for linguistic cues, we assessed whether significant differences exist between sentences in fake and real news in terms of the prevalence and nature of the linguistic features considered.

Linguistic Measurements

The following linguistic traits were measured in order to compare their prevalence in fake and real news: 1. stylometric features were measured, including log frequency of content words and letters per word; 2. sentences' complexity was measured using length of utterances and number of prepositions per sentence; 3. Words semantic features were obtained from rating norms (see below) and averaged per sentence. Semantic features included concreteness, familiarity and imaginability, and emotional valence and arousal.

Stylometric Features

Mean word length per sentence was calculated using Perl script in a UNIX environment. Log-frequency norms of content words were taken from the publicly available Spanish annotated dictionary described in Duchon et al., (2013) and available at <http://www.bcbl.eu/databases/espal/>. Sentences in false and true stories were assigned mean logarithmic frequency scores averaging over their components words.

Syntactic Complexity

Sentence length and number of prepositions per sentence were taken as a proxy of complexity (Shaffer, 2022) and calculated using Perl programming in a UNIX environment. We use utterance length as a proxy of complexity based on psycholinguistic studies that have argued that length of utterances is one of the measurement indices sensitive to changes of complexity through development. In accordance with this, authors such as Johansson (2009) and Parker and Brorson (2005) affirm that this is one of the classic and reliable indicators that has been used to capture the complexity of language. The number of prepositions is a relevant measure since previous research has found that false statements created by participants contain less prepositions than true ones for the case of Spanish (Masip et al. 2012). The number of prepositions was normalized by sentence length.

Subjective Norms of Concreteness, Imaginability, Familiarity and Emotionality

Sentences in fake and true stories received scores of concreteness, familiarity and imaginability based on the average word norms per sentence. We used a public database that gathers subjective scores of concreteness, imaginability and familiarity for 6400 words in Spanish (Duchon et al., 2013), which is freely available at <http://www.bcbl.eu/databases/espal/>. Words were evaluated by Spanish speaking participants on a scale of 1–7 for values of concreteness, familiarity, and imageability. Subjective ratings were obtained from the questionnaires answered by undergraduates from 12 different Spanish universities. The data present in EsPal is the average ratings for over 6,400 words from at least 30 participants and from at least two universities (see Duchon et al., 2013 for details of EsPal subjective ratings collection).

Emotional valence and arousal scores were achieved using a dictionary of 14032 manually annotated words for sentiment labels created by Stadthagen-Gonzalez et al. (2017) and available at <https://link.springer.com/article/10.3758/s13428-015-0700-2>. Valence and arousal are subjective characteristics that reflect experiences and associations with particular objects, events, and words. In their study, Stadthagen-Gonzalez et al., (2017) collected values for these variables via subjective ratings by participants exposed to

lists of words and asked to state how happy or sad, how active or quiet, each stimulus makes them feel (please see details in the cited work by Stadthagen-Gonzalez et al. 2017). For the purpose of this study we defined the variable valence residual, consisting of the absolute value of the difference between the valence rating of a word and the mean rating for all studied words, in order to capture the extremeness of the valence either negative or positive in nature.

Arousal scores were calculated by averaging the arousal norms for words in an utterance. Previous research had shown that, in English and other languages, emotional valence in false statements tends to be extreme, particularly in the negative direction (Ott et al, 2013; Li et al., 2014). To capture polarized values of emotional valence we calculated valence ratings residuals for each word—that is the absolute value of the distance from the word's valence norm and the dictionary mean. In turn, sentences were assigned valence residual scores based on word averages.

Results

Sentences in fake and true news were analyzed and compared. Student's t-tests were used to examine differences between stylometric, complexity and lexical norms in fake and real stories. Second, a logistic regression analysis was run to assess the extent to which veracity was predicted by the linguistic variables considered.

When testing differences between sentences in true and fake news, significant differences ($p < 0.01$) were found in almost all measures with the exception of arousal and concreteness ($ps > .1$). Fake news presented significantly lower number of words per statements ($t = -7.04$; $p < .001$), significantly lower number of prepositions per statement ($t = -11.98$; $p < .001$), lower number of letters per word ($t = -5.07$; $p < .001$). Content words in fake news were significantly more frequent ($t = 8.25$; $p < .001$) and rated more familiar ($t = 7.32$; $p < 0.001$) and imaginable ($t = 2.6$; $p = 0.009$). Finally words in fake news presented more extreme values of positive and negative valence, as showed by marginally significant higher valence residual values ($t = 2.21$; $p = .02$). Figure 1 illustrates the false/real statements mean differences across measures.

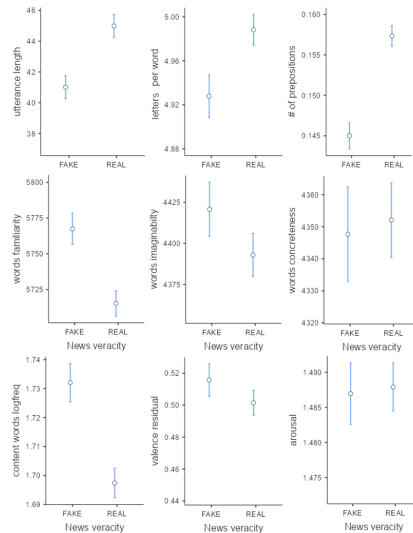
Variables were entered simultaneously as predictors in a logistic regression analysis, entering News veracity as the dependent variable. Due to the high collinearity value associated to concreteness scores ($VIF > 2$), this variable was removed from the model, after which collinearity assumptions were met satisfactorily (all $VIF < 2$), resulting in predictive model significantly different from chance ($\chi^2(8) = 253$; $p < 0.001$). Table 1 presents the predictive variable's coefficients and its significance for individual feature on News Veracity, when controlling for all other variables. Familiarity, imaginability, content word frequency and valence residuals, predicted fake news directly,

Table 1. Model Coefficients - News Veracity

Predictor	<i>Estimate</i>	<i>SE</i>	<i>Z</i>	<i>p</i>
Intercept	1.27725	0.4697	2.72	0.007
words per state- ments	0.00414	6.87e-4	6.02	<.001
number preposi- tions	3.01706	0.3639	8.29	<.001
letters per word	0.17293	0.0392	4.41	<.001
word familiarity	-2.17e-4	5.35e-5	-4.05	<.001
word imaginability	-1.21e-4	3.51e-5	-3.44	<.001
word log frequency	-0.19480	0.0966	-2.02	0.044
valence residual	-0.20216	0.0595	-3.40	<.001
arousal	0.25610	0.1363	1.88	0.060

Note. The regression takes real news as baseline; and estimates represent the log odds of "News veracity = REAL" vs. "News veracity = FAKE"

Figure 1. Comparison of Variable Means (95% CI) in Fake and Real News



while number of prepositions, letters per word and words per statement were inversely associated with false statements. The results are consistent with previous research in that sentence length, number of prepositions, word length were higher in veridical news, while lexical simplicity and easiness (frequency, imaginability and familiarity) were higher in fake news sentences. An exception was the case of concreteness which did not differ between fake/real stories. Finally, emotional valence was more extreme in false stories, but no difference was found between arousal scores.

Discussion

Several voices have raised concern about the growing infusion of misinformation in popular media in recent years, particularly about the proliferation of fake news in public political discourse (Lewandosky et al, 2023). Fake news are intentionally created to spread false information readily, effectively and easily. Then, we would expect false information to contain material that is easy-to-process and shaped to catch the reader's attention. Our data contribute to add evidence in support of this idea, showing that fake news contain less complex, more familiar and imaginable language.

Most fake news studies have focused on Anglo-American and European media. However, in order to fully understand the nature of misinformation in discourse in a broader sense, it is important to study the construal of fake news transculturally and across different languages. Here we compared fake and real news in Spanish, using available lexical norms, finding regularities similar to those previously reported in online media in other languages. This is consistent with previous studies looking at psycholinguistic features of deceptive statements in Spanish produced under prompted and controlled conditions (e.g., Masip et al. 2012). Here we extend this to the context of fake and real news naturally occurring in online media, something that has not been explored before for the case of Spanish.

In sum, fake news contain shorter utterances, more concise words, fewer prepositions per sentence and more familiar and imaginable words. In addition, the emotional valence is more extreme in fake news. Being easier and more familiar to the reader, fake news are prone to be more readily transmitted, boosting diffusion to a larger extent. Extreme values of valence (highly positive or negative words) were found to be more common in fake news—another trait that has also been linked to increased replication in the literature.

Overall, the results highlight two aspects of fake news that have implications at higher levels of discourse and information transmission. First, linguistic cues prevalent in fake news coincide with traits associated with increased replication, such as simplicity and emotionality (Paslawska & Papielska-Borys, 2018). That is, the pieces of information that are easy-to-process and emotionally polarized tend to be repeated and amplified in social media.

On a more theoretical plane, computational models of diachronic cultural transmission have suggested that language traits that are easier to process tend to proliferate in large communities of speakers, while complex ones tend to die out, particularly in communities of highly interconnected speakers (Reali et al., 2018). In turn, this suggests that cognitive language processing constraints may cause certain pieces of information to become progressively simpler at the structural level as the community of speakers becomes increasingly interconnected. If this were the case, the ease of processing false information may have broader implications on the expected distribution of news, particularly when social media—characterized by high levels of interconnection between speakers—becomes the principal means for diffusion. Namely, we should expect that easy-to-process news (a robust trait of fake ones) progressively overcome complex pieces of information.

The emerging picture is one in which media consumers spread simple and emotional news (much of them false) in an uncritical and visceral manner, reproducing a modern version of what Freud, a hundred years ago, called an “automaton” to refer to a crowd full of enthusiasm, lacking a critical spirit and highly susceptible to brightly colored images and the repetition of simple ideas. In the era of post-truth, the results presented here add to the accumulating evidence that points towards the need of critically evaluating veracity and sources when consuming information (Harari, 2019).

Conflict of Interest Disclosure

The Author reports no conflicts of interest.

Funding

No external funding.

Research Ethics Statement

We followed all ethical guidelines as described by the American Psychological Association with regard to designing, conducting, and reporting on the study. This study did not involve human participants, and therefore did not require ethical approval.

Data Availability Statement

Data associated with the current study will be made available upon reasonable request to the corresponding author.

References

- Aragón, M. E., Jarquín, H., Gómez, M. M. Y., Escalante, H. J., Villaseñor-Pineda, L., Gómez-Adorno, H., ... & Posadas-Durán, J. P. (2020, September). Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in Mexican Spanish. In Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain.
- Brady W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. doi:10.1073/pnas.1618923114
- Burkhardt, J. M. (2017). Combating fake news in the digital age. *American Library Association*.
- De Francisci Morales, G., Galeazzi, A., Quattrocioni, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the USA*, 118(9), e2023301118. doi:10.1073/pnas.2023301118
- A., Caschera M.C., Ferri F., & Grifoni P. (2021). Fake news detection: A survey of evaluation datasets. *Peer Journal of Computational Science*, 18 (7):e518. doi: 10.7717/peerj-cs.518.
- Duchon A, Perea M, Sebastián-Gallés, N, Martí, A. & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavioral and Research Methods*, 45(4):1246-58. doi: 10.3758/s13428-013-0326-1.
- Ecker, U., Roozenbeek, J., van der Linden, S., Tay, L.Q., Cook, J, Oreskes N., & Lewandowsky, S. (2024). Misinformation poses a bigger threat to democracy than you might think. *Nature*, 630(8015), 29-32. <https://www.nature.com/articles/d41586-024-01587-3>
- Freud, S. (1921[1955]). *Group psychology and the analysis of the ego*. In *The standard edition of the complete psychological works of Sigmund Freud* (J Strachey, editor). Macmillan.
- Galletta Horner, C., Galletta, D., Crawford, J, & Shirsat, A. (2021). Emotions: The unexplored fuel of fake news on social media. *Journal of Management Information Systems*, 38(4), 1039-1066. DOI: 10.1080/07421222.2021.1990610
- Ghanem B., Rosso P., Rangel F. (2020) An emotional analysis of false information in social media and news articles. In: *ACM Transactions on Internet Technology (TOIT)*, 20(2), 1-18 <https://doi.org/10.1145/3381750>
- Ghanem B., Ponzetto S., Rosso P., Rangel F. (2021) FakeFlow: Fake news detection by modeling the flow of affective information. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, 679–689. <https://doi.org/10.18653/v1/2021.eacl-main.56>
- Giachanou A., Rosso P., & Crestani F. (2021). The impact of emotional signals on credibility assessment. *Journal of the Association for Information Science and Technology*, 72(9), 1117-1132. <https://doi.org/10.1002/asi.24480>

- Gómez-Adorno, H., Posadas-Durán, J. P., Enguix, G. B., & Capetillo, C. P. (2021). Overview of FakeDeS at IberLEF 2021: Fake news detection in Spanish shared task. *Procesamiento del Lenguaje Natural*, 67, 223-231. <https://doi.org/10.26342/2021-67-19>
- Graham, M., Zook, M., & Boulton, A. (2013). Augmented reality in the urban environment: Contested content and the duplicity of code, *Transactions of the Institute of British Geographers*, 38(3), 464. <https://doi.org/10.1111/j.1475-5661.2012.00539.x>
- Jang, S.M., & Kim, J.K. (2017). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, 80, 295-302, <https://doi.org/10.1016/j.chb.2017.11.034>.
- Johansson, V. (2009). *Developmental aspects of text production in writing and speech*. PhD disertación. Department of Linguistics and Phonetics. Lund University.
- Kumar, S., & Shah, N. (2018). False Information on web and social media: A survey. ArXiv, abs/1804.08559.
- Harari, Y. (2019). *21 lessons for the 21st century*. Vintage.
- Hauch V., Blando n-Gitlin, I., Masip, J. & Sporer, S.L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and social psychology review*, 19(4), 307–342. <https://doi.org/10.1177/1088868314556539>
- Larcker, D.F., & Zakolyukina, A.A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2), 495-540. <http://dx.doi.org/10.2139/ssrn.1572705>
- Lewandowsky, S., Ecker, U. K. J., Cook, J., van der Linden, S., Roozenbeek, J., Oreskes, N. (2023) Misinformation and the epistemic integrity of democracy, *Current Opinion in Psychology*, 54, 101711, <https://doi.org/10.1016/j.copsyc.2023.101711>.
- Lewandowsky, S., & Van Der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, 32(2), 348-384. Advance online publication. <https://doi.org/10.1080/10463283.2021.1876983>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science and Public Interest*, 13(3), 106-31. <https://doi.org/10.1177/1529100612451018>.
- Li, J., Ott, M., Cardie, C., & Hovy, E. (2014) Towards a general rule for identifying deceptive opinion spam. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014) (pp. 1566–1576). Baltimore, MD, USA.
- Masip J., Bethencourt M., Lucas G., Segundo, M.S.S., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology*, 53(2), 103–111. <https://doi.org/10.1111/j.1467-9450.2011.00931.x>

- McIntyre, L. (2018). *Post-truth*. The MIT Press.
- Moscadelli, A., Albora, G., Biamonte, M. A., Giorgetti, D., Innocenzio, M., Paoli, S., Lorini, C., Bonanni, P., & Bonaccorsi, G. (2020). Fake news and covid-19 in Italy: Results of a quantitative observational study. *International Journal of Environmental Research in Public Health*, 17(16), 5850. <https://doi.org/10.3390/ijerph17165850>.
- Namkoong, J.E., & Henderson, M. D. (2019). Responding to causal uncertainty through abstract thinking. *Current Directions in Psychological Science*, 28(6), 547-551. <https://doi.org/10.1177/0963721419859346>
- Newman, M.L., Pennebaker, J. W., Berry, D.S., Richards, J.M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675. <https://doi.org/10.1177/0146167203029005010>
- Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N.S. (2013). What yelp fake review filter might be doing? In *Proceedings of the 7th International AAAI Conference on Web and Social Media*.
- Ott, M., Cardie, C., & Hancock, J.T. (2013). Negative deceptive opinion spam. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, 25(3), 365-376. <https://doi.org/10.1177/0142723705059114>
- Paslawska, P., & Popielska-Borys, A. (2018) Phenomenon of fake news. *Social Communication Special Issue*, 4(1), 36–140.
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)* (pp. 3391–3401). Santa Fe, NM, USA.
- Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). Linguistic inquiry and word count: LIWC. Lawrence Erlbaum Associates,
- Posadas-Durán et al. (2019). Posadas-Durán, J.P., Gómez-Adorno H., Sidorov, G., & Escobar, J.J.M. (2019). Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent & Fuzzy Systems*, 36(5):4869–4876. <https://doi.org/10.3233/JIFS-179034>.
- Ramírez-Esparza, N., Pennebaker, J.W., García, F. A., & Suriá, R. (2007). La psicología del uso de las palabras: Un programa de computadora que analiza textos en español [The psychology of word usage: A computer program that analyzes Spanish texts]. *Revista mexicana de psicología*, 24(1), 85-99.
- Reali, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B*, 285(1871), 20172586.. <https://doi.org/10.1098/rspb.2017.2586>
- Ruffo G., Semeraro A, Giachanou A., & Rosso P. (2023). Studying fake news

- spreading, polarisation dynamics, and manipulation by bots: a tale of networks and language. *Computer Science Review*, 47, 100531, <https://doi.org/10.1016/j.cosrev.2022.100531>
- Santos, R., Pedro, G., Leal, S., Vale, O., Pardo, T., Bontcheva, K., & Scarton, C. (2020). Measuring the impact of readability features in fake news detection. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)* (pp. 1404–1413). Marseille, France: ELRA.
- Sarzynska-Wawer, J., Pawlak, A., Szymanowska, J., Hanusz, K., Wawer, A. (2023). Truth or lie: Exploring the language of deception. *PloS ONE* 18(2): e0281179. <https://doi.org/10.1371/journal.pone.0281179>
- Shaffer, J.D. (2022). An improved measure of grammatical complexity: A first step in more accurately assessing the readability of a text. *Journal of Shizuoka University Education*, 18, 1-10. <https://doi.org/10.14945/00028800>
- Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7(3), 247–260. <https://doi.org/10.1002/jip.121>
- An-bin, S., & Wei-yi, Q. (2018). Governmental Crisis Communication and Emotion Guidance in Age of Social Media: A Case of 2015 Shenzhen Landslide. *Modern Communication (Journal of Communication University of China)*, 4, 34-41.
- Spence, C., Okajima, K., Cheok, A.D., Petit, O., Michel, C. (2016). Eating with our eyes: From visual hunger to digital satiation. *Brain Cognition*, 110, 53-63. <https://doi.org/10.1016/j.bandc.2015.08.006>.
- Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M.A., Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavioral Research Methods*, 49(1), 111-123. <https://doi.org/10.3758/s13428-015-0700-2>.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>.
- Ward, A. (2019). *Fake news: read all about it*. New York: New York Times Educational Publishing.
- Wu, Y., Ngai, E., Wu, P., & Wu, C. (2022). Fake news on the internet: a literature review, synthesis and directions for future research. *Internet Research*, 32(5), 1662-1699. <https://doi.org/10.1108/INTR-05-2021-0294>
- Xu, M., Wei, Z., & Wu, J. (2022) How emotional communication happens in social media: Predicting “Arousal-Homophily-Echo” emotional communication with multi-dimensional features. *Telematics and Informatics Reports*, 8, 100019, <https://doi.org/10.1016/j.teler.2022.100019>.