Discerning Fake News

An Automated Analysis Using the *ReaderBench* Framework

SIMINA-MARIA TERIAN, LAURENTIU-MARIAN NEAGU, ANCA-SIMINA MARTIN, STEFAN RUSETI, MIHAI DASCALU

1. Introduction

FAKE NEWS is not only one of the largest threats levelled against democratic societies, but also an unprecedented challenge for contemporary science. Due to its elusiveness, fake news is notoriously difficult to define, classify, and describe, while constant mutations call for uninterrupted effort to rethink the phenomenon and update its toolbox. It is for this reason that fake news analysis oftentimes rests on interdisciplinary cooperation, which helps ensure that the findings of a particular approach are the product of rigorous objective research, and not a set of fallacies arising from methodological bias.

This study sets out to conduct an automated analysis of two interconnected hypotheses. The first one posits that, despite relatively limited related research, a series of fake news subcategories can nonetheless be distinguished (in our opinion, two). The second one is that these two subtypes relate differently to the truth (and, implicitly, to its counterfeiting) and exhibit other dissimilar structural properties, the nature of which we set out to explore in what follows.

Several contextual clarifications regarding the emergence of the two hypotheses are in order. Most attempts to delineate (sub)types of fake news have focused strictly on discussions as to whether it is appropriate to expand the newly formulated concept to include certain traditional categories of the journalistic discourse. A telling example in this regard is Edson C. Tandoc Jr., Zheng Wei Lim, and Richard Ling's study, whereby the authors argue for a six-fold classification of fake news as "news satire," "news parody," "news fabrication," "(photo) manipulation," "advertising," and "propaganda."[1] Maria D. Molina, S. Shyam Sundar, Thai Le, and Dongwon Lee's taxonomy distinguishes, on similar grounds, seven microgenres of fake news: "false news, polarized content, sat-

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS–UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III. It was also financed by Lucian Blaga University of Sibiu & Hasso Plattner Foundation research grants LBUS-IRG-2019-05.

ire, misreporting, commentary, persuasive information, and citizen journalism."[2] Romy Jaster and David Lanius have made a decisive step away from this trend when they suggested that "a news report is true" when the truth is found "both in its *literal content* ('what is said') and in its *communicative content* ('what it pragmatically conveys')."[3] Put differently, fake news subsumes reports that falsify either the literal or the communicative content, or even both. As shown elsewhere, such categorizations are inexact, as both types of fake news are found time and again to have counterfeited the literal content, admittedly in different ways.[4] Instead, we differentiate between "propaganda news" reports, which falsify content via omission, and "fabricated news," whose content features counterfeit textual additions or substitutions.[5] These are the two (sub)categories of fake news we investigate in what follows.

As for the second hypothesis, we can only guess the "structural properties" which describe best these two types of fake news. If the "propaganda news" vs. "fabricated news" binary opposition was juxtaposed with well-established truth value clines, we could say with a high degree of certainty which one of them exhibits the *largest number* of "structural properties." For instance, PolitiFact (www.politifact.com), Poynter Institute's famed fact-checking website, operates with six truth ratings, as follows: "*pants-fire, false, barely-true, half-true, mostly-true,* and *true.*"[6] The rating system behind Factual (www.factual.ro), its Romanian counterpart, is a faithful replica in many respects, with the amendment that it rests on an eight instead of six-fold truth value scale, which ranges from *la mişto* (in jest), *la derută* (misleading), *impostură* (imposture), and *100% inventat* (100% invented) to *legătură falsă* (false connection), *context fals* (false context), *manipulare* (manipulation), and *propagandă* (propaganda).[7] If we were to superimpose these clines onto our classification of fake news, it would not be too farfetched to assume that "fabricated news" articles lie closer to the pole of falsehood (i.e., in the vicinity of categories such as *pants-fire, false, misleading, imposture, 100% invented*), whereas "propaganda news" reports lie at the other end of the spectrum, somewhere near the pole of truthfulness (i.e., in the proximity of categories such as *barely-true, half-true, mostly-true, manipulation, propaganda*), which, in turn, makes the latter less amenable to clear-cut distinctions from truth-content proper. In other words, since "fabricated news" articles are subjected to a larger number of truth-distorting operations than "propaganda news," the former are more likely to exhibit a more varied range of specific "structural properties," whereas the latter is expected to have a more "neutral" tone. It remains to observe whether this supposition holds true for the findings of our automated analysis.

2. Corpus and Method

2.1. Corpus

THIS STUDY considered an initial dataset of $N = 300$ manually extracted fake news reports from 70 Romanian-language websites,[8] which were compiled to form a corpus. The dataset is available open access in FAKEROM, a project which aims to define various Romanian fake news datasets and apply various deep learning models for automated classification. Published between March 2020 and March 2021, the articles submitted for analysis focus on the evolution of the COVID-19 pandemic and were selected from a wide variety of websites, which include news outlets and weblogs. The texts were divided by 2 experts into 4 categories according to the following labels: real news, authentic news, fabricated news, and propaganda news. Of these 4 categories, we will focus on a side-by-side comparison of the two fake news subcategories, namely fabricated and propaganda news; the total number of articles found to reflect these subtypes amounts to $N = 248$ Romanian news (see Table 1).

TABLE 1. LANGUAGE LEVELS CORPUS STATISTICS

Class	# Documents	# Sentences	# Words
Fabricated news	106	3,963	84,961
Propaganda	142	3,700	85,567

2.2. The *ReaderBench* Textual Complexity Indices

ALL THE texts were processed using the *ReaderBench* framework,[9] which provides a multilevel analysis of text characteristics grouped by scope (see the dedicated Wiki page[10]). The available indices for Romanian are computed at 4 granularity levels (i.e., Document—D, Paragraph—P, Sentence—S, and Word—W) and take into account 3 aggregation functions (i.e., mean—M, standard deviation—SD, and maximum—Max). To ensure ease of interpretability, we considered only the mean as an aggregation function; also, all indices at document level, except for word entropy, were disregarded to reveal specific writing styles at a more fine-grained level instead of the overarching document level.

It should be noted that the *ReaderBench* surface indices generally consider simple counts of words, punctuation marks, sentences, and entropy[11] to reflect vocabulary diversity. In terms of morphology, *ReaderBench* returns statistics for each part of speech (i.e., nouns, verbs, adjective, adverbs) using spaCy's PoS tagger.[12] From a syntactical standpoint, *ReaderBench* provides statistics for various dependency types available in the spaCy Romanian parser, while also considering the depth of the parsing tree, which, in turn, reveals the sentence structure complexity. In point of semantics, *ReaderBench* builds the Cohesion Network Analysis graph[13] using ROBERT,[14] which provides suitable contextualized embeddings. To increase interpretability, we limited the scope of the study to an analysis of the level of cohesion between adjacent sentences and paragraphs, thus emphasizing the importance of cohesion flow throughout each news article. Specific discourse connectors—e.g., coordination, conjuncts—are also counted at paragraph and sentence levels. Word level tags derived from spaCy's Named Entity Recognizer, which include person, product, and location, were also considered and computed. Due to their liability to subjective interpretation, the indices derived from Wordnet—e.g., depths in hypernym tree, sense count—were ignored, while searching for peculiarities of the fake news writing style.

To distinguish between the 2 categories of fake news, 3 new custom features were considered in addition to the indices offered by *ReaderBench* (see table 2).

TABLE 2. CUSTOM FEATURES SUBMITTED FOR ANALYSIS

Feature	Description
Quotes	The number of quotes available in the text, which explicitly reference texts from external sources.
References	The count of external links/URLS in the article, which were presumably included to increase believability.
Modality adverbs	Counts of a specific category of adverbs such as "desigur," "neapărat," "negreşit," "poate," "probabil," "pesemne," and "fireşte," which reflect the author's level of confidence in the content of the report.

2.3. Statistical Analyses

A STATISTICAL approach was used to identify which features exhibited significant differences between the 2 categories of fake news. All previous variables—the *ReaderBench* textual complexity indices and the custom features—were checked for linguistic coverage, to ensure that an index has non-zero values for at least 20% of the documents, and normality, which was verified by considering skewness and kurtosis values[15] smaller than 2. All *ReaderBench* complexity indices that were not representative in terms of coverage and which displayed non-normality were removed. The indices were also checked in terms of homogeneity of variances using Levene's test.[16] Independent t-tests were afterwards carried out on all the remaining indices to assess the extent to which writing style indices differed between the 2 categories of fake news. Multicollinearity was assessed as pair-wise Pearson correlations ($r > .90$) and only the indices with the strongest effect sizes were preserved.

Since all 3 custom features were representative in terms of linguistic coverage, albeit with non-normal patterns of distribution, we decided to perform non-parametric Mann-Whitney Z tests to compare the mean ranks between the 2 categories of fake news. All analyses were carried out using SPSS v28.[17]

3. Results

TABLE 3 depicts the results for the statistical analyses performed on the remaining *ReaderBench* textual complexity indices. The acronyms stand for aggregation function (M—mean for all our features), index abbreviation, and the granularity at which the aggregation occurs (after "/"). Significant features ($p < .05$) are highlighted in italics. The most predictive feature, word entropy ("WdEntr"), suggests the use of a more varied vocabulary in fabricated news reports. *However, propaganda news articles exhibit a more elaborated sentence structure, with more unique words ("UnqWd") and verbs ("POS_verb"), additional specific syntactic dependencies ("obl"oblique nominal, "cc"—coordinating conjunction, "nsubj"—nominal subject), and a larger parse tree depth. As expected, there were no significant differences in terms of cohesion, various other types of connectors (i.e., order, contrast, logic), syntactic dependencies, and the use of pronouns (overall at POS level or for specific subcategories—e.g., indefinite, third person).

TABLE 3. INDEPENDENT T-TESTS ON *READERBENCH* TEXTUAL COMPLEXITY INDICES

ReaderBench feature	M (SD)* Propaganda news	M (SD) Fabricated news	*t*(246)	*p*
M(WdEntr/Doc)	4.775 (0.402)	4.964 (0.397)	-3.688	<.001
M(Dep_obl/Sent)	1.334 (0.533)	1.123 (0.463)	3.262	.001
M(UnqWd/Sent)	22.370 (5.192)	20.570 (4.922)	2.761	.006
M(POS_verb/Sent)	3.981 (1.141)	3.610 (1.038)	2.636	.009
M(ParseDepth/Sent)	5.812 (0.924)	5.533 (0.980)	2.296	.023
M(Connector_coord/Sent)	2.457 (0.891)	2.229 (0.750)	2.131	.034
M(Dep_cc/Sent)	0.648 (0.299)	0.572 (0.256)	2.114	.036
M(Dep_nsubj/Sent)	1.356 (0.458)	1.243 (0.386)	2.048	.042
M(Connector_order/Par)	0.108 (0.102)	0.087 (0.090)	1.71	.089
M(Dep_obl/Par)	1.336 (0.666)	1.212 (0.641)	1.473	.142
M(NmdEnt_person/Sent)	1.104 (0.675)	0.981 (0.620)	1.463	.145
M(Connector_conj/Par)	0.400 (0.327)	0.358 (0.277)	1.073	.284
M(POS_pron/Sent)	0.971 (0.475)	0.910 (0.452)	1.036	.301
M(Pron_indef/Sent)	0.839 (0.316)	0.876 (0.35)	-0.862	.39
M(Connector_addition/Sent)	0.602 (0.344)	0.572 (0.288)	0.733	.465
M(Dep_aux/Par)	0.820 (0.467)	0.781 (0.475)	0.639	.523
M(Connector_contrast/Par)	0.215 (0.181)	0.203 (0.168)	0.505	.614
M(Dep_det/Sent)	1.290 (0.499)	1.318 (0.531)	-0.428	.669
M(Dep_iobj/Par)	0.138 (0.145)	0.145 (0.141)	-0.363	.717
M(SentAdjCoh/Par)	0.222 (0.096)	0.227 (0.096)	-0.329	.742
M(Dep_amod/Sent)	1.385 (0.599)	1.360 (0.583)	0.328	.743
M(Dep_cop/Sent)	0.338 (0.186)	0.342 (0.173)	-0.19	.849
M(Pron_thrd/Sent)	0.569 (0.262)	0.575 (0.284)	-0.189	.850
M(Dep_advmod/Sent)	1.347 (0.635)	1.353 (0.517)	-0.071	.943
M(Connector_logic/Sent)	0.799 (0.370)	0.796 (0.337)	0.069	.945

Table 4 introduces the Mann-Whitney Z scores for the custom features. Only one of them was significant ($p < .05$), namely the number of references, which is lower for propaganda news relative to fabricated news.

TABLE 4. MANN-WHITNEY Z SCORES FOR THE CUSTOM FEATURES

Feature	M (SD) Propaganda news	M (SD) Fabricated news	Z	p
Quotes	5.61 (15.75)	5.64 (8.61)	-0.320	.750
References	0.15 (0.84)	0.38 (1.57)	-2.285	.026
Modality adverbs	1.52 (3.00)	1.75 (4.25)	-0.483	.630

4. Discussion and Conclusions

THE FINDINGS above highlight several major differences between propaganda and fabricated news across 1/3 of the analyzed textual traits, which, in turn, allows for a more applied discussion. It remains, however, to be seen whether they describe a system or, at the very least, a converging direction. At first glance, propaganda news (PN) reports put forward a *more sophisticated* discourse, whereas fabricated news (FN) articles propose *more rudimentary* narratives. This observation is reinforced by 7 of the 8 textual complexity indices for which the t-test generated significant differences: table 3 shows that PNs exhibit more complex lexical and grammatical structures, as is evident from their multi-level parse trees, the additional specific dependencies (nominal subject and oblique nominal) they feature, and the larger number of unique words, particularly verbs. It should be also noted that these properties manifest at sentence level (/Sent) rather than across paragraphs or an entire document.

In fact, the only relevant textual complexity index applied at document level (/Doc) seems to contradict the interpretation above. This is the case of word entropy (WdEntr), which outlines not only a higher lexical diversity for FNs, but also the largest difference in value illustrated in table 3. However, this observation does not necessarily point to a contradiction with the other seven indices discussed above. Not only do they operate at different levels (/Sent and /Doc, respectively), but their effects are also different. Specifically, the FNs' higher lexical diversity serves to *improve* the report's *vividness* in the eyes of the readership, whereas the more complex discursive structure of the PNs is intended to *make a more convincing case for* the events therein discussed. The explanation is that both FN and PN *strive to compensate* for the discursive elements they have lost while interfering with the truth-content: when certain circumstances of the reported event are suppressed, PNs are more likely to experience a decrease in the level of discursive coherence, which, in turn, calls for efforts on the part of the authors to scaffold the argumentative dimension. Conversely, since they falsify the truth-content through addition or substitution, FNs are more exposed to coming across as artificial against the real backdrop of the reported event, so that authors feel compelled to add nuance in order for the invented facts to resemble the rest of the information. A similar pattern emerges from the only relevant feature in table 4, i.e., the reference count. PNs do not make extensive use of references, as they are quite well grounded in reality; rather, they are more concerned with discovering the most seamless modalities of removing uncomfortable circumstances and delivering a simplified propagandistic message. Conversely, in the case of FNs, the illusion that fabricated facts are legitimized by external sources is a vital component; hence, the higher number of references, sometimes unrelated to the reported event.

The observations above point to several conclusions. First, the automated analysis has demonstrated there is a substantial number of textual indices that reinforce the relevance of our distinction between propaganda news and fabricated news, which was initially grounded in linguistic and philosophical research. Second, the latter hypothesis of the study was not only refuted, but also systematically disproven: despite our arguments in favor of its validity, PNs have a larger number of "specific structural properties" than FNs, which appear to exhibit a more "neutral" tone. What accounts for this asymmetry is that both PN and FN strive to compensate for the elements which have been affected by the process of distorting the truth: argumentative coherence and accuracy of detail in the case of PNs and FNs, respectively. Yet, this mechanism gives way to a third, cautionary conclusion: since both types of fake news aim to obscure and counteract precisely those elements that would give them away as fake, it would be a grave mistake to judge a report based on superficial properties such as syntactic complexity or the abundance of references, which fake news overemphasize to conceal their shortcomings. Paradoxically, an ostentatious display of elements otherwise associated with professional journalism should, in this case, serve not as a reassuring factor, but as a warning trigger.

❑

Notes

1. Edson C. Tandoc Jr., Zheng Wei Lim, and Richard Ling, "Defining 'Fake News:' A Typology of Scholarly Definition," *Digital Journalism* 6, 2 (2018): 141–147.
2. Maria D. Molina, S. Shyam Sundar, Thai Le, and Dongwon Lee, "'Fake News' Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content," *American Behavioral Scientist* 65, 2 (2021): 186.
3. Romy Jaster and David Lanius, "Speaking of Fake News: Definitions and Dimensions," in *The Epistemology of Fake News*, edited by Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann (Oxford–New York: Oxford University Press, 2021), 21.
4. Simina-Maria Terian, "What Is Fake News: A New Definition," *Transilvania*, 11–12 (2021): 115.
5. Simina-Maria Terian, "What Kind of News Is Fake News: A New Taxonomy," *Transylvanian Review* 31, Suppl. 1 (2022): 70–83.
6. William Yang Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," in *The 55th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference, Vol. 2 (Short Papers): July 30–August 4, 2017, Van-*

couver, Canada, edited by Regina Barzilay and Min-Yen Kan (Stroudsburg, PA: Association for Computational Linguistics, 2017), 424.

7. See the following analysis based on the Factual corpus: Costin Busioc, Vlad Dumitru, Stefan Ruseti, Simina Terian-Dan, Mihai Dascalu, and Traian Rebedea, "What Are the Latest Fake News in Romanian Politics? An Automated Analysis Based on BERT Language Models," in *Ludic, Co-design and Tools Supporting Smart Learning Ecosystems and Smart Education: Proceedings of the 6th International Conference on Smart Learning Ecosystems and Regional Development*, edited by Óscar Mealha, Mihai Dascalu, and Tania Di Mascio (Singapore: Springer Nature Singapore, 2022), 201–212.
8. What follows is a complete list of the source websites: https://www.activenews.ro, https://ortodoxinfo.ro, https://josbotnita.ro, https://dantanasescu.ro, https://omuldinnordforever.wordpress.com, https://daniel-roxin.ro, https://www.qmagazine.ro, https://iosefinapascal.com, https://www.stiripesurse.ro, https://oficialmedia.com, https://rapcea.ro, http://fymaaa.blogspot.com, https://karensmithdotblog.wordpress.com, https://rtnews.ro, https://octavpelin.wordpress.com, http://romanii-liberi.ro, http://todayactual.site, http://ro.portal-news-24h.xyz, https://sputnik.md, http://romanortodox.info, https://www.justitiarul.ro, https://lupuldacicblogg.wordpress.com, https://vp-news.com, https://yogaesoteric.net, https://www.national.ro, https://burebista2012.blogspot.com, https://www.romaniatv.net, https://evz.ro, https://saccsiv.wordpress.com, https://vremuritulburi.com, https://bucovinaprofunda.com, https://www.nationalisti.ro, https://www.incorectpolitic.com, https://glasul.info, https://jurnalul-patriot.ro, https://bpnews.ro, https://60m.ro, https://constiintacolectiva.net, https://4media.info, https://www.aparatorul.md, https://raintv.eu, https://flux.md, https://www.marturisireaortodoxa.ro, https://jurnalul-patriot.ro, http://moldova24.info, https://www.hranaspirituala.ro, https://invictuswebmedia.com, https://www.infocrestin.com, https://www.andreearaicu.ro, https://noi.md, https://altarulcredintei.md, https://ecology.md, https://autovindecarea.blogspot.com, https://romania-unita.ro, https://www.dir.org.ro, https://www.doctorulzilei.ro, http://www.napocanews.ro, https://ziare.com, https://www.bzi.ro, https://tribuna.us, https://www.cotidianul.ro, https://www.recentnews.ro, https://www.lovendal.ro, https://www.marturisireaortodoxa.ro, https://www.diacaf.com, https://spirituletern.blogspot.com, https://razboiulinformational.ro, https://spynews.ro, https://magnanews.ro, https://fabricatinromania.info.
9. Mihai Dascalu, *Analyzing Discourse and Text Complexity for Learning and Collaborating: A Cognitive Approach Based on Natural Language Processing* (Cham: Springer, 2014); Mihai Dascalu, Philippe Dessus, Maryse Bianco, Stefan Trausan-Matu, and Aurélie Nardy, "Mining Texts, Learner Productions and Strategies with ReaderBench," in *Educational Data Mining: Applications and Trends*, ed. Alejandro Peña-Ayala (Cham etc.: Springer, 2014), 345–377.
10. https://github.com/readerbench/ReaderBench/wiki/Textual-Complexity-Indices.
11. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, 3 (1948): 379–423.
12. https://spacy.io/models/ro.
13. Mihai Dascalu, Danielle S. McNamara, Stefan Trausan-Matu, and Laura K. Allen, "Cohesion Network Analysis of CSCL Participation," *Behavior Research Methods* 50, 2 (2018): 604–619.

14. Mihai Masala, Stefan Ruseti, and Mihai Dascalu, "RoBERT—A Romanian BERT Model," in *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8–13, 2020*, edited by Donia Scott, Núria Bel, Chengqing Zong (Barcelona: International Committee on Computational Linguistics, 2020): 6626–6637.
15. Kenneth D. Hopkins and Douglas L. Weeks, "Tests for Normality and Measures of Skewness and Kurtosis: Their Place in Research Reporting," *Educational and Psychological Measurement* 50, 4 (1990): 717–729.
16. Howard Levene, "Robust Tests for Equality of Variances," *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by Ingram Olkin (Palo Alto, CA: Stanford University Press, 1960), 278–292.
17. https://www.ibm.com/products/spss-statistics.

Abstract

Discerning Fake News:
An Automated Analysis Using the *ReaderBench* Framework

Fake news is a global phenomenon and one of the largest threats against democratic societies. This study sets out to analyze whether it is relevant to distinguish between two different subcategories of fake news, namely propaganda and fabricated news. To this end, a selection of 300 fake news from the FAKEROM corpus was subjected to an automated analysis using the *ReaderBench* framework. The results point not only to the usefulness of differentiating between the two subtypes of fake news, but also to a particular modality in which they operate, whereby both fake news subcategories strive to compensate rhetorically for the discursive aspects which were affected by the process of distorting reality.

Keywords

fake news, propaganda versus fabricated news, automated text analysis, *ReaderBench* framework