LINGUISTIC ATTRIBUTES AS A VALID TOOL FOR IDENTIFYING FAKE NEWS IN SOCIAL MEDIA PLATFORMS

Gevorg Grigoryan[*]
https://orcid.org/0000-0002-1689-8375
Jiangxi Tellhow Animation Vocational College

Ning Huichun[**]
*https://orcid.org/*0000-0002-0798-2006
Jiangxi Tellhow Animation Vocational College

Jingjing Shi[***]
https://orcid.org/0000-0002-5948-3591
Taizhou Vocational and Technical College

Abstract: *The mass-spread of fake news on social media platforms has emerged as a significant challenge for many online users in recent years. This issue has prompted researchers to develop various mechanisms aimed at detecting and preventing the spread of misinformation. While numerous techniques have been proposed to identify fake content, many are hindered by inherent limitations and inaccuracies. The primary objective of this research paper is to analyze the linguistic attributes of news articles and identify key differences that distinguish fake news from authentic news. Utilizing a qualitative data analysis approach, the following study makes an attempt to examine the systematic linguistic features of both fake and legitimate news articles. The dataset comprises 20 news articles from American news websites and 20 fake news articles sourced from the Facebook social networking website. The findings clearly demonstrate that linguistic features can serve as effective indicators to help online users differentiate between unreliable and factual news content. This study contributes to the existing literature by offering a novel perspective on the role of linguistic attributes in fake news detection.*

---

[*] gevorg.grigoryan93@gmail.com
[**] ning.huichun@tellhow.com
[***] janejjshi@outlook.com

Introduction

Social media has become one of the indispensable elements of human life. Every individual is motivated towards participating in social media in exchange for rewards in the form of friendship, appreciation, knowledge, and participation (Sipahi, 2017). Individuals on social media are both producers and consumers of information (Grover et al., 2022). Facebook, Twitter (X microblog), YouTube, Snapchat, Reddit, etc., have become the key online platforms for mass communication and dissemination of information for individuals and companies. Aside from being a major communicative and informative tool for daily life, social media poses some threats to its users. One of the challenges that social media users face on daily basis is the mass spread of fake news (in this article misinformation, disinformation and fake news are used interchangeably). The easy access of social media platforms, anonymity, multiracial and multicultural audiences serve as a solid background for spreading intentional misinformation to the targeted audience. The problem of fake news intensifies especially in period of political elections, cultural and historical events, pandemics, natural disasters, etc., causing mass spread of abuse, anxiety, and hatred to targeted groups or individuals. Disinformation can easily influence the decisions and actions of social media representatives. Misinformation can easily alter the attitude of individuals towards any event or topic. Moreover, the mass spread of fake news is decisive in the results of governmental or presidential elections, pandemics, natural disasters, etc. Combatting against the spread or emergence of disinformation has been a major problem for many countries. However, the easy accessibility of creating new accounts is out of control even for social media platform owners. To solve this problem, a number of countries have employed certain measures to accelerate cyber literacy among online users. The target of cyber literacy is to literate social media users, accelerate their awareness and cognitive skills to differentiate fake news from authentic news. The analysis of linguistic attributes of news articles is an effective method in determining misinformation. The application of specific linguistic features in a text can easily reveal the authenticity of any news. Fake news creators try to manipulate the audience by generating content for specific purposes.

Thus, the objective of the current study is:

1) to analyze the content of fake and real news articles in terms of linguistic features and style;
2) to compare the distribution of linguistic attributes of news articles (Dataset 1 and Dataset 2) by employing the qualitative data analysis method;

3) to identify the key differences between fake and authentic news in terms of linguistic features (grammatical, syntactic, and stylistic).

The following research article provides fresh practical as well as theoretical insights on how the application of linguistic attributes detects fake news in social media platforms. Furthermore, the significance of this paper is highly supported by the application of qualitative data analysis in comparing and identifying linguistic attributes of news articles.

Our paper attempts to pose the following research questions:

RQ1: What are the major causes of the formation of fake news articles on social media platforms?

RQ2: What are the differences between fake and authentic news articles in terms of linguistic attributes and style?

By answering these questions, our study seeks to accelerate public awareness in detecting fake news articles. The answers to these research questions can help media specialists to better combat the spread of fake news among social media users. The results of the research will definitely serve as a blueprint for media and IT specialists, as well as linguists.

Despite the prior research on the determination of fake news articles on social media platforms, to our knowledge, there are no studies conducted before that have targeted on identifying linguistic attributes of fake and authentic news articles in political context.

Literature Review

The emergence of social media has raised the mass spread of fake news causing anxiety and hatred to a targeted group or individual. Many researchers have tried to define fake news. The term *fake news* was included in Oxford Dictionary (OxfordLearnersDictionaries.com) only in 2017 defining it as false reports of events, written and read on websites. Allcot and Gentskow (2017) were among the first researchers who defined fake news as articles that are intentionally and verifiably false and could mislead readers (in Tandoc et al., 2017). Fake news may refer to the manipulation of information that can be carried out through the production of false information, or the distortion of true information (Aïmeur et al., 2023).

Researchers Egelhofer and Lecheler have identified three recurring definitional characteristics in the literature: an item can be considered fake news "when it is low in facticity, was created with the intention to deceive, and is presented in a journalistic format" (in Boudana & Segev, 2024). For us, fake news is the intentional falsification of any news that tends to affect public opinion causing

mass spread of anxiety and hatred. The spread of misinformation has always been a part of human life. However, after the emergence of digital technologies, the detection of fake news and the suspense of the dissemination has come out of control. Fake news identification is still a complex unresolved issue (Aïmeur et al., 2023).

On social networks, the reach and effects of information spread occur at such a fast pace and are so amplified that distorted, inaccurate or false information acquires a tremendous potential to cause real-world impact, within minutes for millions of users (Figueira & Oliveira, 2017).

Nowadays, the detection of fake news on social media has become a hot research area for linguists, media specialists, journalists, etc. The multi-functionality, accessibility to a wide and diverse audience, and anonymity trigger the wide spread of fake news on social media platforms especially on Facebook (Facebook is considered to be the major communication social networking site for many English-speaking users). As Sharma, Shu, and others mention, social media has become nowadays a powerful source for fake news dissemination (Sharma et al. 2019; Shu et al. 2017).

The potential of fake news on social media was far beyond the expectations and imagination of many researchers until 2016. The impact of fake news has become critically prevalent after 2015. The 2016 U.S. Presidential election and the Brexit referendum in 2016 were both characterized by widespread disinformation and misinformation and – in the U.S. case – accusations of fake news (Broda & Strömbäck, 2024). The high proportion of fake news spread has dramatically affected the prestige and rating of the US presidential candidates making the election more severe and tough among political leaders. Moreover, fake news spread shaped the opinion of many British people in the historical period of Brexit. The negative influence of fake news has not been confined by these two major events. However, this was only the initial period of fake news emergence and dissemination.

During the outbreak of COVID-19 the mass spread of misinformation on different social media platforms has substantially increased triggering panic all around the world. The latter has led to critical health challenges in public. During the COVID-19 pandemic, the dynamic interplay between entertainment and information became more pronounced, with a significant portion of content shared for amusement or as a coping mechanism, inadvertently fueling the dissemination of misinformation (Islam et al., 2020). Studies have also revealed motivations for spreading false information, including socialization, the desire for entertainment, and a desire for social status (Islam et al., 2020).

After experiencing the devastating impact of fake news on different global events, the governments of many countries have started to take some measures, more specifically promoting media literacy among social media users, launching fact-checking programs, or strengthening new legislative regulations. However, the outcomes of these measures are not satisfactory and accurate.

Methodology

In this section, we outline the design of our proposed methodology. To begin, we collected forty news articles. Half of these articles (twenty) were sourced from prominent American news websites, including CNN (edition.cnn.com), The Wall Street Journal (wsj.com), The New York Times (www.nytimes.com), NBC News (nbcnews.com), and Fox News (www.foxnews.com). The remaining twenty articles were sourced from Facebook, a highly influential social media platform known for its widespread dissemination of information.

To analyze the linguistic peculiarities of the texts, we applied qualitative data analysis tool. Qualitative Data Analysis is a fundamental process in social and behavioural research, providing invaluable insights into the rich tapestry of human experiences (Hamed, 2024). Qualitative research is a type of research that explores and provides deeper insights into real-world problems (Moser & Korstjens, 2017).

In order to analyze the linguistic attributes of news reports, our first step is to construct data. It primarily includes two datasets. The first dataset contains 20 articles from 5 famous American news websites, whereas the second dataset includes 20 news articles from random Facebook pages. Both datasets are obtained in the form of texts. Pictures, links, and infographics have been removed. Different software tools can be employed to process qualitative data analysis. NVivo is one of them. It can easily identify and evaluate the metrics of any text, style of writing, as well as content structure based on different linguistic attributes. The collected datasets are processed by NVivo software.

Linguistic Features of News Articles

Our research paper makes an attempt to study the key differences between fake and real news by analyzing their major linguistic features. To achieve the goal of our research, we examine the frequency of different linguistic attributes in datasets 1 and 2. The list of linguistic attributes includes the following components: verbs (stative and modal verbs, infinitive and gerund constructions, phrasal verbs), active and passive voice, pronouns (subject and object personal pronouns, possessive pronouns, reflexive pronouns, etc.), adverbs, adjectives (superlative and

comparative), direct and indirect speech, long and short sentences, negations, interjections, conjunctions, quotations. The rate of frequency of the above-mentioned components has been processed computationally.

To conduct a qualitative content analysis, we have tried to code the linguistic attributes. Our research team designated 12 codes and 11 sub-codes to both datasets. The matrix of designated codes and sub-codes is displayed in Figure 1.

Figure 1: The matrix of codes and sub-codes

Codes	**Sub-codes**
Verbs	Stative verbs
Active and passive voice	Modal verbs
Pronouns	Verb phrases
Adverbs	Superlative degree
Adjectives	Comparative degree
Direct and indirect speech	Subject personal pronoun
Long and short sentences	Object personal pronoun
Negations	Reflexive pronouns
Interjections	Possessive pronouns
Conjunctions	Infinitive
Negation	Gerund

Linguistic attributes in both fake and authentic news texts were examined separately. After studying the news articles and their relevant linguistic features, we employed qualitative data analysis software to conduct calculations of the frequencies of linguistic attributes.

To reveal the differences our research team compared the outcomes of the linguistic features of the first dataset to the outcomes of the second dataset. After

that, we applied the results of the calculations to determine the linguistic features of fake news. The model of our research methodology is illustrated in Figure 2.

Figure 2: Model of Research Methodology
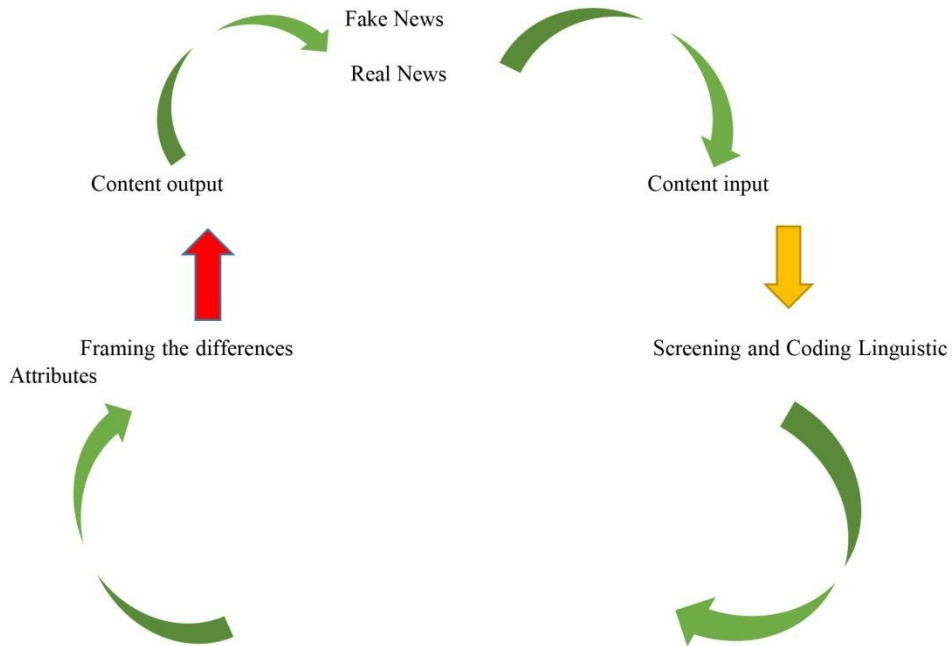
As mentioned, our data contains 40 news reports collected from the most famous American news websites and random Facebook pages. This section includes many graphs that display the application of linguistic attributes in both datasets. The illustration of these tables enables us to detect the deceptive characteristics of fake news. The interconnection of linguistic features has a substantial role in identifying the reasoning mode. The outcomes depict the detailed list of linguistic features in both datasets. A profile is created to illustrate the frequency rate of the features that are employed in the framework of our data. The below-illustrated graphs depict the analysis of these linguistic attributes.

Table 1 illustrates the detailed list of linguistic attributes that determine the authenticity of news sources. More specifically, the linguistic attribute of direct and indirect speech is the most frequently used feature in real news rating with a score of 21.6%. All the authentic news sources were marked with the presence of direct and indirect speech. This grammatical feature is followed by active and passive voice and negations. The majority of the tested material contained elements of

voice and negation. Genuine texts include a high frequency of comparative adjectives, common nouns, and quotations as well. Comparative adjectives show a higher score than superlative adjectives. Long sentences are common to real news as well.

Table 1: List of Linguistic Attributes (Factual News)

1.	Codes and Subcodes	Word count	Codes by Percentage	Recorded cases by percentage	Non-binary words
2.	Direct and Indirect Speech	41	21.6%	98%	1465
3.	Active and Passive Voice	34	17.8%	97.8%	1558
4.	Common nouns	29	15.1%	88.89%	978
5.	Adverbs	25	12.4%	87.1%	817
6.	Comparative Adjectives	17	5.8%	57.3%	746
7.	Quotations	14	5.5%	43.5%	645
8.	Subject personal Pronouns	10	4.9%	35.6%	515
9.	Stative verbs	9	2.2%	33.41%	201
10.	Infinitive to	5	2.0%	21.2%	98
11.	Modal Verbs	3	1.6%	19.8%	174
12.	Long sentences	2	1.1%	11.4%	183

Table 2 contains the detailed information of all the tested linguistic attributes relevant to the second dataset. As one can see from the table below, proper nouns rank as the most frequently employed linguistic feature. Active and passive voice, direct and indirect speech are excessively applied as well. The third most commonly used lexical attribute in fake news is adjectives, covering 92.3 % of the tested material. Dataset 2 was also marked with the application of interjections and negations. Compared to Dataset 1, Dataset 2 does not have any shortage of pronouns, especially personal pronouns. And finally, Dataset 2 has plenty of long sentences.

Table 2: List of Linguistic Attributes (Fake News)

1)	Codes and Subcodes	Word count	Codes by Percentage	Recorded cases by percentage	Non-binary words
2)	Proper Nouns	54	13.2%	75.2%	1157
3)	Active and Passive Voice	49	11.5%	84.2%	935
4)	Direct and indirect Speech	42	10.3%	90%	284

5)	Superlative Adjectives	39	9.8%	92.3%	362
6)	Metaphors	36	8.3%	69.5%	184
7)	Negations	29	6.5%	47.8%	119
8)	Personal pronouns	23	4.7%	61.8%	736
9)	Short sentences	19	3.4%	33.9%	837
10)	Superlative adjectives	15	2.2%	65.4%	2314
11)	Infinitive to	4	1.7%	15.4%	204
12)	Modal Verbs	2	0.8%	9.8%	936

The application of proper nouns, personal pronouns, comparative adjectives, and verbs in the present tense is leading compared to other linguistic features in real news articles. Both fake and authentic news articles are abundant in reported speech and active and passive voice structures. Top linguistic features also contain adverbs, negations, interjections, and quotations. The coverage of linguistic features of fake news articles is demonstrated in Table 2.

Results and Discussions

The mass production and dissemination of fake news is a tremendous threat to online users, companies, political parties, etc. Fake news can easily convince and mislead the readers. The outcomes of our research paper can help readers to identify the content of fake news from those of authentic news. The linguistic analysis reveals that fake news contains many proper noun elements, while genuine news does not. This finding complied with the previous research conducted by Horne and Abele, who state that fake titles are longer, use few stop words, and fewer nouns but more proper nouns (Horne & Abele, 2017). Undoubtedly, the excessive application of proper nouns tends to attract the attention of social media users. The usage of proper nouns has a repetitive characteristic.

At the same time, fake news is marked with the abundance of pronouns. Aside from pronouns, deceptive news contains many verbal constructions mainly in the Present Tense. These results accord with the outcomes stated by a group of British researchers who are trying to create tools for detecting fake news and stopping its dissemination. More specifically they state that fake texts contain more frequent use of verbs, especially in the present tense. Also, more use of pronouns and adjectives (Linguistic Cues Could Be Key to Exposing Fake News, 2022).

The findings of our research illustrate that creators of fake news make use of plenty of exaggerated linguistic attributes, more specifically adjectives in superlative degree. Generally, the creators of fake news try to employ some

emotive and evaluative adjectives to attract and influence the readers. The application of adjectives in comparative degree is at a low rate. These results are absolutely in accord with the work of Ott et al., who mentioned the overuse of superlative adjectives in fake news content (Ott et al., 2011). The application of adjectives in a superlative degree has a rhetorical effect. The application of superlative adjectives mainly tends to convince readers without providing any room for suspicion.

The application of emotional words (metaphors) is an indirect way of luring readers. Metaphors and emphatic words mainly seek for persuasion. Fake news is identified as a sinister form of mass persuasion (Nyilasy, 2019). Metaphors create some sort of uncertainty and anxiety among social media users. This idea was stated in many psychological research papers.

Our study also displays that fake news articles employ more quotations and negations to make their content look more authentic.

Identifying fake news articles is really challenging. Using linguistic attributes to determine fake news is an effective approach.

Conclusion

Our study makes an attempt to determine the linguistic differences peculiar to fake news and factual news. The outcomes of our research show that proper nouns, superlative adjectives, subject personal pronouns, and long and repetitive sentences dominate in fake news content, while factual news mainly contains common nouns, adverbs, comparative adjectives, and short sentences. Fake and authentic news articles also share some common linguistic features. Both of them employ a high proportion of reported speech and passive and active constructions. Fake news is also marked with the abundance of emphatic words, metaphors, interjections, repetitive phrases and sentences, etc., while the content of genuine news mainly does not include these linguistic elements. Fake news creators use emotive words to persuade and mislead the audience, so the style of fake news is predominantly characterized by its persuasive and emotive overtones. Meanwhile, authentic news incorporates a neutral tone in its style trying to provide an accurate and fact-based content to its readers. Overall, linguistic features are good indicators of exposing fake news articles.

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining, 13(1).* https://doi.org/10.1007/s13278-023-01028-5

Boudana, S., & Segev, E. (2024). Fake News Makes the News: Definitions and Framing of Fake News in Mainstream Media. *Journalism Practice, 1-20.* https://doi.org/10.1080/17512786.2024.2379898

Broda, E., & Strömbäck, J. (2024). Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review. *Annals of the International Communication Association*, 1-28. https://doi.org/10.1080/23808985.2024.2323736

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, *121*, 817-825. https://doi.org/10.1016/j.procs.2017.11.106

fake news noun – Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com. (n.d.). https://www.oxfordlearnersdictionaries.com/definition/english/fake-news?q=fake+news

Grover, P., Kar, A. K., & Dwivedi, Y. (2022). The evolution of social media influence - A literature review and research agenda. *International Journal of Information Management Data Insights*, *2*(2), 100116. https://doi.org/10.1016/-j.jjimei.2022.100116

Hamed U. (2024). Qualitative Data Analysis. https://doi.org/10.13140/RG.2.2.18417.25440

Horne Abele, B. D. S. (2017). *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News.*

Islam, A. N., Laato, S., Talukder, S., & Sutinen, E. (2020). Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. *Technological Forecasting and Social Change*, *159*, 120201. https://doi.org/10.1016/j.techfore.2020.120201

Islam, T., Mahmood, K., Sadiq, M., Usman, B., & Yousaf, S. U. (2020). Understanding Knowledgeable Workers' Behavior Toward COVID-19 Information Sharing Through WhatsApp in Pakistan. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.572526

Linguistic cues could be key to exposing fake news. (2022). Department of Literature, Area Studies and European Languages.

Moser, A., & Korstjens, I. (2017). Series: Practical guidance to qualitative research. Part 1: Introduction. *European Journal of General Practice*, *23*(1), 271-273. https://doi.org/10.1080/13814788.2017.1375093

Nyilasy, G. (2019). Fake news: When the dark side of persuasion takes over. *International Journal of Advertising*, *38*(2), 336-342. https://doi.org/10.1080/-02650487.2019.1586210

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. ArXIV preprint arXiv:1107.4577

Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining "Fake News." *Digital Journalism*, *6*(2), 137-153. https://doi.org/10.1080/21670811.2017.1360143

Sharma K., Seo S., Meng C., Rambhatla S., Liu Y. (2020). Covid-19 on social media: analyzing misinformation in Twitter conversations. arXiv preprint arXiv:2003.12309

Shu K., Wang S., Lee D., Liu H. (2020c) Mining disinformation and fake news: concepts, methods, and recent advancements. In *Disinformation, misinformation, and fake news in social media,* 1-19. Berlin, Springer. https://doi.org/10.1007/ 978-3-030-42699-6_1

SiPahi, E. (2017). SOSYAL MEDYA İLE KURUMSAL İTİBAR OLUŞTURMA. *The Journal of Social Sciences, 4*(10), 598-617. https://doi.org/10.16990/ sobider.3346

Zafarani, R., Zhou, X., Shu, K., & Liu, H. (2019, July). Fake news research: Theories, detection strategies, and open problems (pp. 3207-3208). In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* https://doi.org/10.1145/3292500.3332287

Գ. Գրիգորյան, Ն. Հուխչուն, Զ. Շի – *Լեզվաբանական հատկանիշները որպես սոցիալական մեդիա հարթակներում կեղծ լուրերի բացահայտման գործիք.* – Սոցիալական մեդիա հարթակներում կեղծ լուրերի զանգվածային տարածումը վերջին տարիներին դարձել է առցանց օգտատերերի ամենամեծ մարտահրավեր-ներից մեկը։ Կեղծ լուրերի ազդեցությունը գրավել է բազմաթիվ հետազոտողների ուշադրությունը, որի շնորհիվ նրանք մշակում են հանրության շրջանում ապա-տեղեկատվության տարածումը հայտնաբերելու և արգելելու մեխանիզմներ։ Կեղծ տեղեկատվությունը հայտնաբերելու համար մշակվել են բազմաթիվ տեխ-նոլոգիաներ, որոնք, այնուամենայնիվ, սահմանափակ են։ Այս հետազոտական հոդվածի նպատակն է վերլուծել լրատվական հոդվածների բովանդակությունը և բացահայտել լեզվական հատկանիշների հիմնական տարբերությունները, որոնք կարող են մատնանշել կեղծ և ճշգրիտ լուրերի հիմնական անհամապատասխա-նությունները։ Կիրառվել է որակական տվյալների վերլուծություն՝ կեղծ և վավե-րական լուրերի կանոնավոր և համակարգված լեզվական առանձնահատկու-թյունները հայտնաբերելու և որոշելու համար։ Հոդվածի տվյալների շտեմարանը

ներառում է 20 լրատվական հոդված ամերիկյան լրատվական կայքերից և 20 կեղծ լուր՝ Ֆեյսբուք սոցիալական մեդիա հարթակից: Հետազոտության արդյունքները ցույց են տալիս, որ լրատվական հոդվածների լեզվական հատկանիշները կարող են էապես օգնել առցանց օգտատերերին տարբերակել ոչ հավաստի լուրերը փաստացի լուրերից:

Բանալի բառեր. ճշգրիտ լուրեր, կեղծ լուրեր, տեղեկատվություն, լեզվական հատկանիշներ, ապատեղեկատվություն, սոցիալական մեդիա հարթակներ

Г. Григорян, Н. Хуэйчунь, Ц. Ши – *Лингвистические атрибуты как действенный инструмент для определения фейковых новостей на платформах социальных сетей.* – Массовое распространение фейковых новостей на платформах социальных сетей стало одной из самых больших проблем для пользователей сети в последние годы. Влияние фейковых новостей привлекло внимание многих исследователей, что привело к разработке механизмов обнаружения и запрета распространения дезинформации в общественных местах. Разработан широкий спектр методов для ежедневного обнаружения фейкового контента, однако все эти методы отмечены определенными ограничениями и неточностью. Целью данной работы является анализ содержания новостных статей и выявление основных различий с точки зрения лингвистических атрибутов, которые могут выявлять основные несоответствия между фейковыми и подлинными новостными статьями. В настоящем исследовании использовался качественный анализ данных для обнаружения и определения регулярных и систематических лингвистических особенностей фейковых и подлинных новостей. Материал исследования включает 20 новостных статей с американских новостных сайтов и 20 фейковых новостных статей, взятых с платформы социальной сети Facebook. Результаты исследования показывают, что лингвистические атрибуты новостных статей могут значительно помочь пользователям сети отличать недостоверные новостные статьи от фактических новостей.

Ключевые слова: подлинные новости, фейковые новости, информация, лингвистические атрибуты, дезинформация, платформы социальных сетей