Gender-based violence in fake news: A corpus analysis of semantic collocations

Patrizia Sorianello[1], Giovanni Vinciguerra[1]

[1]University of Bari 'Aldo Moro', Piazza Umberto I, 70121 , Bari, Italy

Abstract

This study, conducted within the SERICS's project, investigates the semantic collocations associated with gender-based violence by comparing a corpus of fake news headlines with a corpus of real news headlines. Using a corpus-based approach, the analysis identifies recurrent collocational patterns involving the words woman and violence that contribute to the discursive construction and manipulation of gender-based violence narratives. These patterns often serve to distort, trivialise, or ideologically frame acts of violence, often reinforcing stereotypes and disinformation. By highlighting the role of collocational choices in shaping meaning, this study contributes to research on linguistic manipulation in fake news and offers insights into the representation of gender-based violence in misleading digital discourse, offering potential linguistic indicators of online social risks.

Keywords

Gender-based violence, Fake news, Semantic collocations, Corpus linguistics

1. Introduction

Falsehoods and lies, in all their forms, undermine the veracity of information and constitute linguistic manipulation. They are prevalent in politics, advertising, ideological and religious propaganda, and journalism. Owing to the numerous socio-cultural, emotional, and psychological repercussions of this pervasive phenomenon, urgent identification and prevention measures are required. For this reason, misinformation represents a significant risk factor within the broader field of cyber social security.

Manipulation is a multidisciplinary process that affects a variety of domains, including language. However, most studies of linguistic disinformation tend to fall under the remit of the sociology of communication and social psychology rather than linguistics. Notable exceptions include the research of [1] and [2] on political propaganda in Italy.

One of the main forms of linguistic manipulation is fake news. They can be understood as a form of partial falsification of information that reverses the traditional news hierarchy by altering the prominence and salience of conveyed details. This process ultimately aims to distort the formation of objective personal judgement. Scholars claim that there is no simple definition of fake news; the term is generally used to refer to "any news that is suspected to be inaccurate, biased, misleading, or fabricated" [3, 1]. Research into fake news has grown significantly in recent years, driven by mounting concerns over its detrimental impact on society. This is particularly due to the ability of fake news to generate real fears, with direct and tangible repercussions on daily life [4]. Academic research on fake news is, therefore, pivotal. For some time now, the linguistic sciences have increasingly focused on the systematic analysis of its linguistic properties. Indeed, linguistics and related disciplines – such as philosophy of language and rhetoric, have long shown sensitivity toward defining and understanding truth [5]. The two branches of linguistics that have particularly shaped research perspectives on mis- and disinformation are corpus linguistics and computational linguistics. More specifically, the vast majority of research on the language of fake news has been conducted within the field of NLP and has focused on the development of tools for the automatic detection of fake news [6, 7, 8]. Linguistic

1st Workshop on Symbiotic AI for Cyber Social Security: Human–AI Collaboration for Social Resilience and Trustworthy Digital Ecosystems (SAI4CCS 2026), co-located with the 2nd Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09–13, 2026, Cagliari, Italy

patrizia.sorianello@uniba.it (P. Sorianello); giovanni.vinciguerra@uniba.it (G. Vinciguerra)

0000-0002-6632-0555 (P. Sorianello); 0009-0000-9473-5566 (G. Vinciguerra)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-14

CEUR-WS.org/Vol-4329/paper17.pdf

research in this area often clashes with instrumental and methodological limitations. First, scholars report a lack of a standard definition of "fake news," which does not always allow for a clear binary classification (true vs. false). In addition, there is a shortage of manually annotated news corpora that are well-balanced in terms of size, type, authorship, and sources; furthermore, much of the existing research is conducted exclusively in English and replies on the same datasets. Finally, a major limitation concerns the lack of an underlying theory explaining why real and fake news should differ linguistically. Indeed, although numerous NLP studies have achieved high levels of accuracy in distinguishing news types [9], such results should be interpreted with caution, as this line of research often prioritises maximising accuracy over explaining patterns of language usage [3].

Generally, analyses conducted to date have tended to focus on three levels of linguistic analysis: the lexical level, which examines words and their syntagmatic organisation (e.g., collocations); the morpho-syntactic level, which addresses the grammatical structure of utterances; and the discursive level, which investigates a wide range of phenomena, from pragmatics to broader discourse features (e.g., texts and sentences length).

According to [3], who analysed a series of truthful and false articles written by Jayson Blair for The New York Times, difference between real and fake news can be traced to a stylistic variation related to the authorial intent. In particular, real news is characterised by higher information density, the presence of epistemic markers, and longer, complex noun phrases and lexical items, which distinguish it from manipulated content.

Fake news headlines have also attracted the attention of linguists. For example, [10] investigated the role of implicit pragmatic phenomena in a corpus of Hungarian-language fake news. It is well established that newspaper headlines function almost as autonomous textual genres within the overall structure of a news article; their primary function is not only informative, but also to establish a dialogue with the reader [11]. The authors observe that fake news headlines often exhibit the implicit omission of nuclear arguments, such as subject and object. This strategy can be interpreted as a manipulative device aimed at generating clickbait. Second, some headlines perform speech acts with incomplete propositional content, while others convey an implicit illocutionary force. The resulting combination of explicit and implicit information creates information gaps that trigger psychological effects, such as fear, hope, or curiosity, designed to attract clicks.

Studies have shown that fake news differ significantly from real news in terms of content and semantic organisation. As a result, the information they convey varies both in presentation and lexical choice, shaping different perceptions of the news as a whole. To investigate this, it is necessary not only to quantify the distribution of linguistic processes, but also the examine the textual and semantic structures of the corpora.

2. The semantic collocations

In order to quantitatively assess the importance of semantic choices in news transmission, we focused on semantic collocations. Semantic collocations have a composite nature and are of particular interest in several fields of research, including second language (L2) acquisition purposes, foreign language translation, lexicology, and corpus linguistics, as well as, more recently, computational linguistics.

First identified by [12], the process refers to a meaningful combination of multiple words that forms a recognisable, but not fixed, phraseological unit in natural languages [13, 14]. For example, the Italian phrase *domare un incendio* (to put out a fire) is a typical semantic collocation, as this combination of words tends to appear frequently. This means that the meaning of a word depends not only on the word itself, but also on how it is coupled with the surrounding words, forming what are known as linguistic chunks [15]. Sinclair's approach is frequency-based; nevertheless the statistical frequency of the multi-word unit is not the only salient parameter to consider. In a collocation, a lexical base (or node, in Sinclair's terminology), such as *domare*, combines with one or more collocates, such as *incendio*, contributing to the construction of specific semantic properties and potentially modifying the base word's primary meaning.

Semantic collocations cannot be predicted based on their semantic or syntactic properties. For this reason, they need to be listed in specific dictionaries (see, for example, [16] for Italian or [17] for L2 learners). However, the phenomenon is complex. On the one hand there is no universal agreement on its definition, so it is often necessary to distinguish the semantic collocations from semantic preferences (which allow for greater freedom in combination), idiomatic expressions (which have a fixed degree of combinability), and compounds. On the other hand, collocations are characterised by an arbitrary combination of words (they are considered semi compositional) and a high frequency of the combination itself. They have a varying degree of semantic transparency. To this end, [13] identified the following properties: i. non-compositionality, ii. the impossibility of changing a word with similar lexemes (for example, *tirare un appello* is not an accepted synonym of *lanciare un appello*), iii. a wide range of possible configurations, iv the ability to insert something within the construction, differently to idiomatic expressions (for example, *domare velocemente un incendio*, 'to quickly put out a fire'). Overall, each collocation is characterised by a certain degree of semantic cohesion, which is conveyed by the dual conditioning conveying of the words that form the construction (principle of proximity). Collocations may have diverse syntactic structures; the most frequent are verbal, (V + N , such as *esprimere un desiderio*, 'to make a wish'), Nominal, (N+N, such as *birra alla spina*, 'draft beer'), Adiectival (N+ Adj, pioggia torrenziale, ' torrential rain) [18].

3. The Research

Previous studies have shown that manipulation relies on specific linguistic structures and content. Accordingly, one of the SERICS's primary objectives was to identify the most common linguistic processes involved in the production of disinformation, as linguistic manipulation represents a significant risk in social communication and can impact cyber social security. A computational lexical extraction procedure was applied to the data to identify recurring semantic structures and to measure indices of lexical representativeness. In line with the objectives of the SERICS project, the analysis addressed gender-based violence by examining falsehood and disinformation through lexical content. The study analyses a sample of fake news headlines (FN-h) alongside a control corpus of real news headlines (RN-h). To enhance analytical accuracy, the analysis focused on the most frequent semantic collocations in both corpora associated with the sensitive lexical items *donna* (woman) and *violenza* (violence), which served as thematic nodes for the gender-based violence filter. The study evaluate the strength of these semantic associations in order to identify potential semantic risk indicators and contribute to the development of preventive measures.

It was hypothesised that fake and authentic news differ in their representation of gender-based violence. In particular, we expected the lexical items *donna* and *violenza* to exhibit distinct collocational patterns in the FN-h dataset compared to the RN-h dataset.

The research was conducted using the online software *SketchEngine* [19] on selected keywords. The software is a well-established tool for computational lexical extraction, enabling the analysis of large text corpora to identify grammatical relations, categorised collocations, and lemma sketch in context [20]. The results show for each keyword the concordances, as well as the distribution and frequency of its collocations.

3.1. Material and methods

To achieve this goal, the research analysed two balanced datasets of real and fake headlines drawn from online newspapers filtered for content related to gender-based violence. The analysis was conducted with the methodological awareness that the newspaper headline constitute a distinct textual genre that is inherently prone to manipulation [21].

The first sample comprised 1,200 fake news headlines (FN-h) identified by Vox (https://voxnews.info), while the second sample comprised 1,200 real news headlines (RN-h) drawn from the online edition

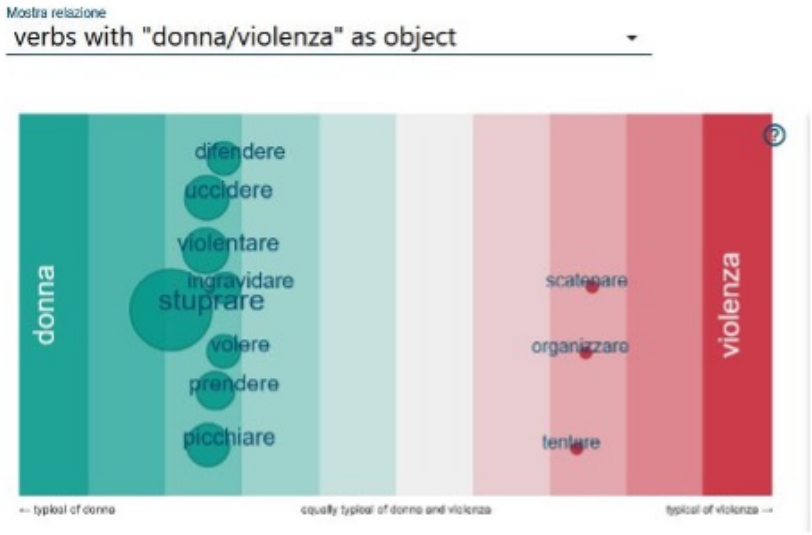


Figure 1: FN-h: Sketch Difference of *donna/violenza*

of *Corriere della Sera* (https://www.corriere.it) published between 2023 and 2024 [1] . The two resulting datasets each consist of 32.000 tokens. Given the limited size of the dataset, currently under construction, the research should be regarded as preliminary and exploratory.

For each sensitive lexical items, *Word Sketches* were extracted for donna (woman) and violenza (violence), including frequency and score values. The *Word Sketch* (WS) provides an automatic one-page summary of a lemma's collocational behaviour, based on the *logDice* score, which quantifies grammatical relations and the semantic strength of collocates. In addition, the *Word Sketch Difference* (WD) function allows for the comparison of two lemmas, displayed in different colours, highlighting differences in usage and contextual distribution.

3.2. Semantic collocations of sensitive nodes

The analysis shows that FN-h are characterised by altered textual structures and weakened structural cohesion. Meaning units tend to be saturated with sensationalist, unverifiable, and often irrelevant or incomplete information. One recurring pattern involves the deliberate manipulation of morphosyntactic structures at both the words and sentence level to influence interpretation. The results further indicate that several linguistic processes, such as the alteration of punctuation, direct speech and quotation forms, occur more frequently in FN-h than in RN-h. In addition, FN-h display a higher frequency of morphological and lexical intensifiers, performative speech act verbs, and negative connotated lexemes.

We consider the collocations of the nodes *donna/violenza* (woman and violence) in the two corpora. Figures 1 and 2 present intensity graphs, where the colour gradient (blue for *donna*, red for *violenza*) and the size of each bubble indicate the strength of the semantic collocation. These visualisations allow for a clear comparison of the semantic relationships in fake versus real news headlines, highlighting the different ways in which sensitive lexical items are framed.

The analysis of verbs for which *donna/violenza* (woman/violence) serve as syntactic object reveals distinct patterns. In the FN-h sample, the strongest association is with the verb *stuprare* (to rape) reflecting a negative and sensationalist framing (Figure 1). In contrast, in the RN-h sample (Figure 2), the lemma donna is most strongly associated with verbs such as *aiutare* (to help), *salvare* (to save) and *difendere* (to defend), while the collocational pattern of violenza are less extreme and display different contextual patterns. These results suggest that fake news tends to focus on violent and sensational

[1]The choice is not random: Vox is an anonymously Italian website that regularly disseminates false information in sup-port of an anti-immigration agenda, whereas Corriere della Sera is one of Italy's oldest daily newspapers, published in Milan and founded in 1876.

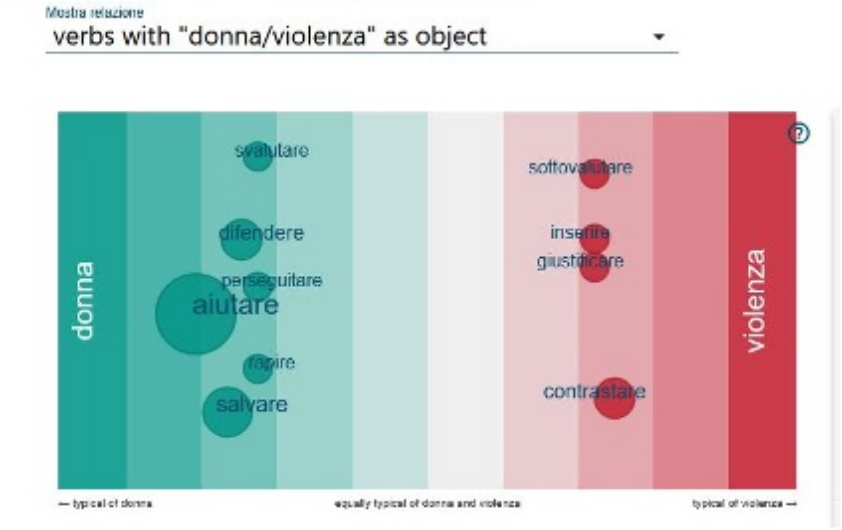


Figure 2: RN-h: Sketch Difference of *donna/violenza*

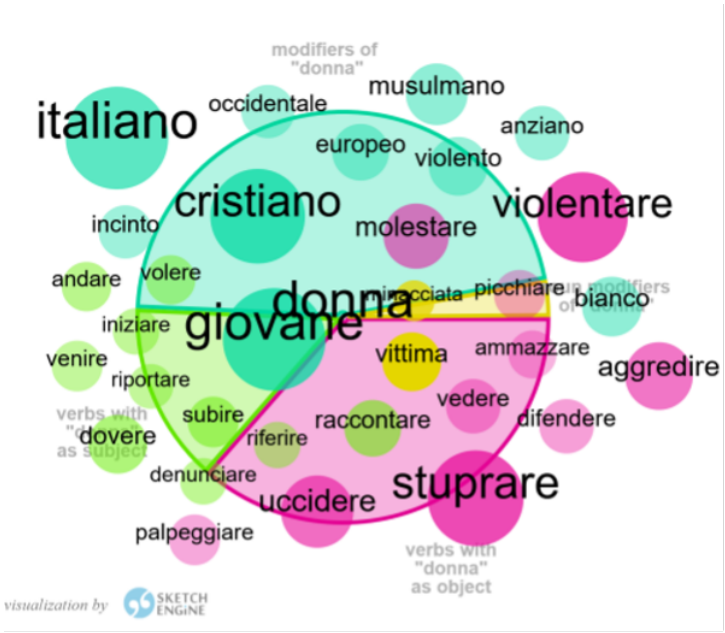


Figure 3: FN-h: WS of the term *donna*

content portraying women as targets of violence, whereas real news headlines more often associate women with protective or positive actions, portraying women as the victims to save.

A similar pattern emerges when analysing the same words from a grammatical perspective. The graphs shown in Figures 3 and 4 highlight the grammatical subjects, object complements and modifiers most frequently associated with word *donna*, with the target word placed at the centre of each graph. Each grammatical function is represented by a different colour, and the size of each lemma's circle ad the font size reflects the strength of its collocation based on frequency. For each lemma, the associated words and their respective frequencies are displayed.

This analysis shows that in RN-h (Figure 4), the most prominent collocation is between donna 'woman' and the modifier *vittima* 'victim'. In parallel, donna is the strongest grammatical object of *aiutare* (to help), followed by *difendere* (to defend) and *salvare* (to save), consistent with the most frequent modifiers *vittima* (victim). In contrast, in FN-h (Figure 3), the strongest object-verb relation is between donna and the verbs *stuprare* and *violentare* (to rape), followed by *uccidere* (to kill) and



Figure 4: RN-h: WS of the term *donna*

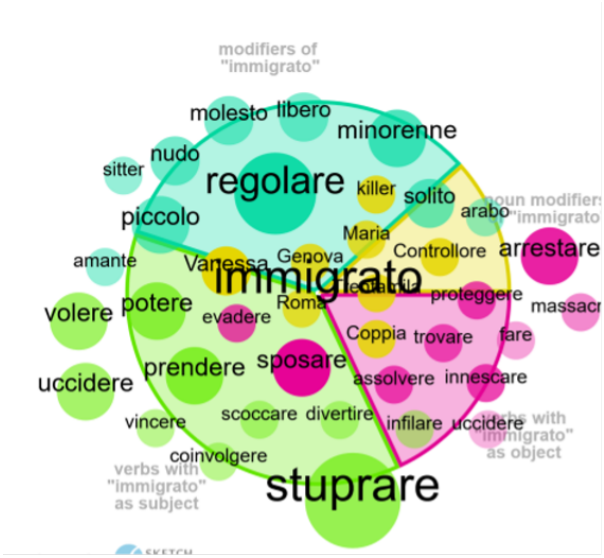


Figure 5: FN-h: WS of the term *immigrato*

picchiare (to beat), while the more frequent modifier of donna is *giovane* (young).

The search for the terms *immigrato* (immigrant) and *africano* (African) reveals a striking contrast between fake ad real news (Figures 5 and 6). These words do not appear in RN-h, but in the FN-h sample they show strong statistical associations. Specifically, *immigrato* frequently functions as the grammatical subject of the verb *stuprare* (to rape), while *africano* (African) is the most frequent modifier of the noun *stupratore* (rapist) (Figure 5). This pattern illustrates how fake news headlines manipulate language to create highly charged and sensationalist narratives. By associating immigrant and African identities with violent actions, the headlines construct a biased, emotionally provocative portrayal that reinforces stereotypes and amplifies fear. Such strategies demonstrate the deliberate use of lexical and grammatical structures to influence readers' perceptions, highlighting the manipulative potential of language in online misinformation.

Consider the examples (1) and (2) extracted from the FN-h corpus:

(1) Violentata nel parcheggio di un supermercato: è stato il solito AFRICANO (Raped in a supermarket car park by the usual AFRICAN man) [https://voxnews.info/2023/11/27/]

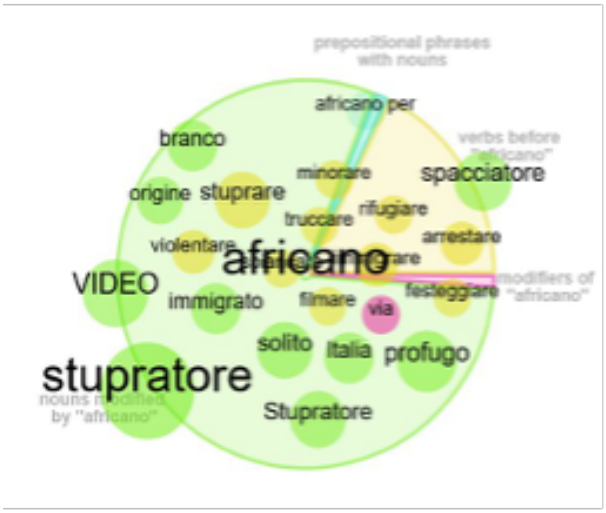


Figure 6: FN-h: WS of the term *africano*

(2) Tombolo, altra donna uccisa da un africano: quando li sbatteremo fuori? (Tombolo, another woman has been killed by an African: when will we kick them out?) [https://voxnews.info/2023/09/21/]

The use of the phrase *il solito africano* (the usual African man) implies that African men rape Italian women. This phrase conveys a personal and stereotyped opinion influenced by racial prejudice, and is presented to readers without any opportunity for rebuttal. The use of a capital letter for *AFRICANO* further emphasises the racial stigma conveyed by the headline. We did not observe this type of framing in the RN-h sample used as a control. Colons, italics. and quotation marks are also frequently employed in FN-h (see Example 2). Presenting the victim's exact words can create an illusion of reliability, a strategy commonly used to increase the persuasive effect of manipulative information. Similarly, there is widespread use of speech acts intended to incite action and violence), which amplifies the negative semantic weight of terms, like the word *egiziano* (Egyptian) in (see Example 3).

(3) *Stupro Catania. «Tremavo come una bambina», egiziani urlavano: «Io sono pazzo, posso ammazzarti»* (Rape in Catania. 'I was trembling like a little girl', Egyptians shouted. 'I am crazy; I can kill you') [https://voxnews.info/2024/02/05]

On the same vein, the lexical extraction for the term *violenza* (violence) provides different findings. First of all, we noticed that the frequency of this lemma differs between the two corpora. In FN-h, the word *violenza* exhibits a higher distribution score (5574,86 token per million, 0.56%), whereas in RN-h, it occurs less frequently (2865,06 token per million, 0.29%). This suggests that *violenza* is more commonly used in FN-h to highlight the aggressive content of the news. This semantic emphasis is further amplified by the frequent selection of the noun *stupro* (rape). In contrast, in RN-h the word *violenza* is not only less frequent, but it also does not typically collocate with terms such as *stupro* or *aggressione* (assault), thereby conveying a more neural meaning. Additionally, it was observed that in RN-h, common collocates of *violenza* include specific terms such as *femminicidio* (feminicide), *genere* (gender), *donne* (women), *contro* (against). These terms do not appear in the collocates of *violenza* in the FN-h dataset.

4. Discussion

The research is still ongoing, but the results are promising. Analysis of the semantic collocations automatically extracted using *SketchEngine* reveals that the two datasets - authentic (RN-h) and fake



Figure 7: FN-h: collocation network of the word *violenza*

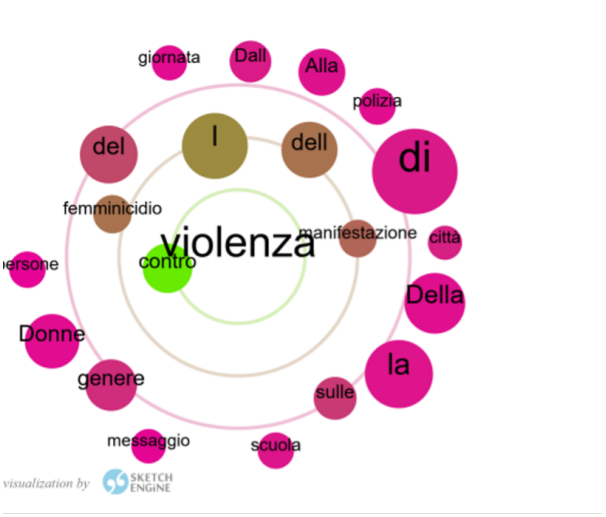


Figure 8: RN-h: collocation network of the word *violenza*

(FN-h) - differ not only in form but also in content. The latter often subtly distorts and influences the audience's understanding of the news. One of the most reliably parameter for differentiating fake news (FN-h) from real news (RN-h) is the presence of prejudicial content, which was found exclusively in the FN-h sample.

It is important to note that FN-h also contain many performative speech acts, more specifically commissives, since the speaker commits to a future course of action. This means that the words used in these headlines do not merely describe events, but also direct the reader's action, thus having practical effects. The content does not simply inform, but actively seeks to persuade the reader to adopt the opinion presented in the news. As a result, fake news headlines are designed to capture the reader's attention and encourage them to share the information, making this the primary objective.

Such linguistic strategies play a central role in the rapid spread of falsehood, shaping perceptions and behaviours with potentially social consequences. Therefore, recognising these rhetorical mechanisms is essential for fostering a more critical awareness of how disinformation circulates and how readers can resist manipulation. This highlights the importance of analysing the linguistic construction of fake messages, as understanding their persuasive techniques is crucial for counteracting their impact.

The lexical choices also differs between the two datasets. First of all, lexical extraction of the semantic collocates for *donna* and *violenza* reveals the presence of different words associated with these target

terms. Both the *Word Sketch* and the *Word Sketch Difference* show how the range of collocates around *donna* varies. In FN-h donna is the grammatical object of verbs with a strong negative connotation, such as *stuprare* (to rape), *violentare* (to assault), *aggredire* (to assault), *ammazzare* (to murder), *palpeggiare* (to grope). In contrast, in RN-h, the same lemma donna select verbs that have an opposite semantic polarity, such as *aiutare* (to help), *salvare* (to save), *difendere* (to defend), *proteggere* (to protect)[2]. When the news is manipulated and biased, *donna* is very often portrayed as a young, targeted individual subjected to some form of sexual violence by men, who are typically depicted as immigrants, particularly Africans. By contrast, RN-h tends to present a more neutral account of gender-based violence, in which women are depicted as victims of negative events who require help, rescue and protection. Terms such as *stuprare* (to rape) instead of *assalire* (to assault), *ammazzare* (to murder) or *massacrare* (to massacre) instead of *uccidere* (to kill), *branco* (gang) instead of *gruppo* (group) fall into the category of negative vocabulary. These words are not interchangeable, as they function differently at the cognitive and emotional processing level, potentially fostering enduring social biases over time.

Furthermore, in RN-h, there is no performative incitement to action or violence; however this does not mean that performative speech acts are entirely absent. The only forms identified are we are calls to protest and to raise awareness against feminicides or other forms of gender-based violence. The interconnection of linguistic structures and processes further reinforces this effect, amplifying it in a manner consistent with confirmation bias. The linguistic coding of macro-categories and processes for rendering implicit meaning, as well as the computational filtering of corpora, can provide a more precise characterisation of the alterations and distortions present in misleading information, enabling early detection and mitigation of social and cybersecurity risk factors. Future research should aim to expand the corpus in terms of size and diversity, and explore additional lexical queries, not only in headlines but also within the full texts of the corresponding news articles, in order to gain a more comprehensive understanding of manipulative strategies across different textual levels.

Acknowledgments

This Publication was produced with the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 - Partnerships extended to universities, research centres, companies and research D.D. MUR n. 341 del 15.03.2022 – Next Generation EU (PE0000014 - "SEcurity and Rights In the CyberSpace - SERICS" - CUP: H93C22000620001).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] E. Lombardi Vallauri, Pesare gli impliciti, in: A. Ferrari, L. Lala, R. Stojmenova (Eds.), Testualità. Fondamenti, unità, relazioni, Franco Cesati, Milano, 2015, pp. 61–81.

[2] E. Lombardi Vallauri, V. Masia, Implicitness impact: Measuring texts, Journal of Pragmatics 61 (2014) 161–184.

[3] J. Grieve, H. Woodfield, The Language of Fake News, Cambridge University Press, Cambridge, 2023.

[4] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection, in: R. Barzilay, M.-Y. Kan (Eds.), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics,

[2]We have no data with which to compare our findings. However, some recent research on the language used in the Italian media's coverage of feminicides, based on corpora of newspaper articles about crimes against women, shows some similarities [22, 23]. One of the most relevant findings is the frequent association of donna with he role oft vic-tim. while men are depicted as violent perpetrators to be arrested.

Vancouver, Canada, 2017, pp. 422–426. URL: https://aclanthology.org/P17-2067/. doi:10.18653/v1/P17-2067.

[5] M. Taboada, Linguistic approaches to fake news research are growing and maturing: commentary on a special issue, Linguistics Vanguard (2025). URL: https://doi.org/10.1515/lingvan-2025-0160. doi:doi:10.1515/lingvan-2025-0160.

[6] N. K. Conroy, V. L. Rubin, Y. Chen, Automatic deception detection: Methods for finding fake news, Proceedings of the Association for Information Science and Technology 52 (2015) 1–4. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pra2.2015.145052010082. doi:https://doi.org/10.1002/pra2.2015.145052010082. arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.2015.145052010082.

[7] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media, Big Data 8 (2020) 171–188. URL: https://journals.sagepub.com/doi/abs/10.1089/big.2020.0062. doi:10.1089/big.2020.0062. arXiv:https://journals.sagepub.com/doi/pdf/10.1089/big.2020.0062.

[8] X. Zhou, R. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities, ACM Comput. Surv. 53 (2020). URL: https://doi.org/10.1145/3395046. doi:10.1145/3395046.

[9] R. Oshikawa, J. Qian, W. Y. Wang, A survey on natural language processing for fake news detection, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Twelfth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 6086–6093. URL: https://aclanthology.org/2020.lrec-1.747/.

[10] E. N. T., Z. Németh, K. N. C., Implicit pragmatic phenomena in headlines of hungarian health-related fake news, Linguistics Vanguard (2025). URL: https://doi.org/10.1515/lingvan-2024-0088. doi:doi:10.1515/lingvan-2024-0088.

[11] T. A. van Dijk, Racism and the Press, Routledge, London, 1991.

[12] J. R. Firth, A synopsis of linguistic theory, 1930–1955, in: Studies in Linguistic Analysis, Longman, London, 1957, pp. 1–31.

[13] S. Evert, The Statistics of Word Cooccurrences: Word Pairs and Collocations, Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Stuttgart, 2005.

[14] M. Lewis (Ed.), Teaching Collocation: Further Developments in the Lexical Approach, Hove, England, 2000.

[15] J. Sinclair, Corpus, Concordance, Collocation, Oxford University Press, Oxford, 1991.

[16] P. Tiberii, Il dizionario delle collocazioni, Zanichelli, Bologna, 2012.

[17] S. Spina, The dictionary of italian collocations: Design and integration in an online learning environment, in: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Malta, 2010, pp. 3202–3208.

[18] M. Benson, Collocations and idioms, in: R. Ilson (Ed.), Dictionaries, Lexicography and Language Learning, Pergamon Press, Oxford, 1985, pp. 61–68.

[19] A. Kilgarriff, P. Rychlý, P. Smrž, D. Tugwell, The sketch engine, in: Proceedings of the Eleventh EURALEX International Congress, Lorient, France, 2004, pp. 105–116.

[20] R. Xiao, T. McEnery, Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective, Applied Linguistics 27 (2006) 103–129. URL: https://doi.org/10.1093/applin/ami045. doi:10.1093/applin/ami045. arXiv:https://academic.oup.com/applij/article-pdf/27/1/103/447211/ami045.pdf.

[21] S. Spina, News as a conversation: A comparative analysis of the language of online and printed newspapers in italy, Recherches en Communication 39 (2013).

[22] L. Busso, C. R. Combei, O. Tordini, La rappresentazione lessicale della violenza di genere: 'donne come vittime' nei media italiani, in: B. Aldinucci, V. Carbonara, et al. (Eds.), PAROLA. Una nozione unica per una ricerca multidisciplinare, Edizioni Università per Stranieri di Siena, Siena, 2019, pp. 261–279.

[23] L. Busso, C. R. Combei, O. Tordini, The mediatization of feminicide: A corpus-based study on the representation of gendered violence in italian media, L'Analisi Linguistica e Letteraria 28 (2020) 29–48.