Analysing Linguistic Markers on Fake News to Enhance the Explainability of Deception Detection Systems

Alba Pérez-Montero

Dept. of Software and Computing Systems, University of Alicante, Apdo. de Correos 99, E-03080, Alicante, Spain.

Abstract

The massive use of social media has increased the ease of dissemination of information. Unfortunately, every type of information can be massively disseminated (even deceptive information deliberately created to mislead). In this research we introduce a deep analysis of linguistic cues (i.e., adjectives, pronouns, complex syntax, emotion words, etc.) that can lead to distinguish which texts can be deceptive. The research focuses on extracting a combination of features: content-based, context-based, readability, virality and information richness. The objective is to test if current NLP tools for deception detection can extract in a satisfactory way these features and to examine the grade of explainability that these systems offer. The methodology starts from a multidisciplinary point of view, focusing in elaborate an integrative research. To pursue the objective of this research, we also combine both analytical and empirical methodologies. The expected impact is to enhance the ability to distinguish misinformation by improving the accuracy and transparency of deception detection systems for everyone.

Keywords

NLP, fake news, readability, explainability, virality, information richness, linguistic markers, inclusive IA, deception detection

1. Justification of the research

The use of social media has massively multiplied the last years, making very easy the communication between individuals and the spread of information. According to Gottfried and Shearer [1], nearly two-thirds of American adults retrieve information via social media. Notwithstanding, every type of information can spread quickly, even false information. The development of social media platforms has intensified the diffusion of fake news [2].

The internet not only provides a medium for publishing fake news but also offers tools to actively promote dissemination [3]. The rapid distribution of fake news is due to the widespread use of social media which offer a fertile ground for instantly sharing and circulating news with the users having no means of quality checking over the shared content [4]. This wide dissemination of information has also been studied as *virality*. This concept relates to fake news because the more viral a false information is, the more probable is to cause harm. As Esteban-Bravo et al. [2] show in their research, the potential virality of fake news can be predicted by analyzing written texts. Their proposal is to implement early stage strategies that can help to control the dissemination of false information, because once a false information is spread, debunking it is a major challenge [2].

Moreover, the increasing quantity of information online makes it every time more difficult to individually analyze it. For this reason, implementation of Natural Language Processing (NLP) techniques and tools is mandatory. As an example, Esteban-Bravo et al. [2] used machine learning models to classify fake news by their level of virality. Also Bonet-Jover et al. [5] combined machine learning and deep learning techniques to create a two-layer model architecture for automatic fake news detection.

To this day, many false information detecting tools have been developed. An example of this is the Veripol tool created by Quijano-Sánchez et al. [6] in collaboration with the Spanish National Police. It is a system that detects false reports automatically. Nevertheless, not many researches are focused on explainability of these tools. The term *explainability* refers to the ability of a machine learning model to offer a mechanism by which its decision-making can be analyzed, and possibly visualized [7]. As Kotonya et al. [7] include in their survey, it is crucial to make every NLP tool sufficiently explainable.

Doctoral Symposium on Natural Language Processing, 26 September 2024, Valladolid, Spain.

alba.perezm@ua.es (A. Pérez-Montero)

In this case, they focus on explanation functionality – that is systems providing claims to support their predictions. NLP systems need to offer explanations that are actionable, causal, coherent, context-full, interactive, unbiased, and chronological [7]. For this reason, it is important to bear in mind that every step in deception detection investigation should be transparent and easily understandable for everyone. Current technologies are mature enough to provide a sound basis for the development of components to automatically detect and remove obstacles to reading comprehension[8].

Before going into the next sections, we define some relevant concepts related to false information that we will use throughout our research:

Fake news: is the term related to fabricated information that mimics news media content in form but not in organizational process or intent [3].

Deception: is the term related to white lies, omissions, and evasions to bald-faced lies and misrepresentations [9], that is to say, messages transmitted with the objective of creating a false information different from the verifiable reality.

Misinformation: is the term related to the factually incorrect or misleading information that is not backed up with evidence [10].

Disinformation: is the term that involves misleading information knowingly being created and shared to cause harm [11].

In this PhD thesis we mainly focus on the term *deception* because its intention to confuse the receiver leaves a "linguistic impression" that can be analyzed. Thus, different research works use that terms to differentiate nuances of meaning regarding to the writer intention or the background information that is available.

The motivation of this research arises from the need to unify and compile different approaches to enhance deception detection and to revise exlpainability of deception detection systems with the aim to make these systems more accesible and inclusive.

2. Background and Related work

Previous studies have tried to delimit the linguistic markers that allow the detection of the falsehood or veracity of a message. In 2019, Gravanis et al. [4] review the most complete classifications of linguistic cues to deception. In this study, they mainly focus on analyzing three taxonomies of linguistic cues to deception: [12], [13] and [9].

As a result, they extract 27 linguistic markers that respond to dimensions such as complexity, expression of uncertainty, expressiveness, or degree of formality, among others.

However, the topic of linguistic cues to deception started some years before. From the beginning, in 2003, DePaulo et al. [14] elaborated a exhaustive experiment with participants who were instructed to write false statements of true statements. This experiment allowed the researches to analyze false and truthful texts and they extracted 158 cues to deception. This study is based on a psychological perspective, but it has laid the groundwork for subsequent researches in linguistic analysis of false statements.

In 2004, Zhou et al. [9] conducted another experiment with participants from which they extracted 9 linguistic constructs: quantity, diversity, complexity, specificity, expressiveness, informality, affect, uncertainty, and non immediacy.

On their part, Hauch et al. [15] offer a meta-analysis on the linguistic markers of deception. They review 44 previous works and extract 79 markers, examining each of them to determine whether they are really discriminatory between false and true information. Their research results show that constructs as the expression of certainty, expression of emotions, distancing from what it is being said, details and expression of cognitive processes should be taken into consideration in order to analyze deceptive texts.

More recently, in 2020, Santos et al. [16] proposed a new taxonomy of linguistic cues that includes also readability features. As they affirmed, readability features are formed by branches of features from other linguistic levels, such as morphological, syntactic and semantic, so the robustness of these features

Table 1
Review of studies on linguistic cues to deception and misinformation.

Authors	Date	Title	Criteria
B. M. DePaulo, et al.	2003	Cues to Deception	Length, Complexity, Unique Words, Sensory Information, etc.
L. Zhou, et al.	2004	Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications	Specificity, Expressivity, Uncertainty, Affect, etc.
V. Hauch, et al.	2014	Are Computers Effective Lie Detectors? A Meta-Analysis of Linguistic Cues to Deception	Mistakes, Expressivity, Emotions, etc.
R. Santos, et al.	2020	Measuring the Impact of Readability Features in Fake News Detection	Readability Index, Concreteness, Familiarity, etc.
C. Zhou, et al.	2021	Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness	Persuasive Words, Emotions, Comparative Words, etc.
M. Esteban-Bravo, et al.	2024	Predicting the virality of fake news at the early stage of dissemination	Readability, Pronouns, Informatily, Affect, etc.

could be differential in identifying different writing styles in fake news. In this regard, readability features are markers related to complexity.

Moreover, Zhou et al. [11] approach this question trying to discriminate if the quality and details of information can be useful in detecting false information. They studied persuasive, comparative, emotional and uncertainty words in the misinformation dissemination process. They also analyze if misinformation dissemination is stronger when it includes multimodal content. Their contribution is relevant because they categorized three levels of richness in online information: level 1 for text-only, level 2 for text with image, and level 3 for text with video.

In the most recent study, from 2024, Esteban-Bravo et al. [2] show that to analyze fake news it is important to consider multiple features: writing style features, readability/complexity features, and psychological features. Their contribution is a proposal of classification of levels for virality in the social network X (formerly Twitter): 50 retweets, between 50 and 1000 retweets, between 1000 and 5000 retweets, and more than 5000 retweets, which is the viral category.

These studies helps us to initiate a complete and detailed taxonomy of linguistic markers to detect deception and they provide us a valuable reference point to improve the state of the art in this topic for English. A review is made considering date, topic and the criteria they extracted of the researches, as can be seen in **Table 1**.

Previous works do not explore further in the linguistic principles or contextual variables that can infer in the interpretation of false information or analysed different languages. In this case, it is necessary to add a deeper linguistic point of view in order to elaborate a generalizing taxonomy of linguistic characteristics that could be applied to more than one language, textual genre or modality.

As Bonet-Jover et al. [5] proved, fake news combines true and false data with the intention of confusing readers. In this study they analyze digital media by using the traditional journalistic structure of news 5W1H (What, Who, Where, When, Why and How). They also demonstrate that determining the veracity of each 5W1H component using only textual information has a limited prediction performance, so adding high-level features like fact-checking information, semantic relations between components or contextual features would be beneficial.

On its part, Saquete et al. [17], elaborate a review about fake news detection from the NLP perspective. They point out that there are different subtasks within fake news detection: deception detection, stance detection, controversy and polarization, automated fact-checking, clickbait detection and credibility scores. However, in all cases they indicate that it is necessary to create both resources and standardized and balanced evaluation metrics that can be applied to every subtask.

On the other hand, conferences frequently include workshops that are competitions to evaluate NLP tasks. For fake news detection it is relevant the CheckThat! Lab, part of the 2021 Conference and Labs of the Evaluation Forum (CLEF). Nakov et al. [18] present an overview of the lab, whose main objective was to evaluate technology supporting tasks related to factuality in five different languages. This lab is

divided in different tasks: check-worthiness estimation, detecting previously fact-checked claims and fake news detection. More than 130 teams participated and created resources to test their technologies, what makes a invaluable source of resources and references that can be used to improve the NLP state of the art.

Besides conference labs, several researches create specific datasets in order to extract information in a specific domain that can be used to perform different NLP tasks. Therefore, some of the datasets are publicly available and can be used by researchers. As an example, MultiFC [19] is a corpus collected from 26 fact-checking websites in English, including metadata as well as evidence pages of reference. For languages different than English, ForceNLP [20] compile a corpus of news mainly from Mexican web sources in Spanish. Santos et al. [16] used the corpus created by [21] called Fake.Br Corpus. It was collected by crowdsourcing and has been used in several researches.

After reviewing previous researches, it becomes clear that the analysis of deception detection is approached from different disciplines that can entwine and work together to perform a complete a wide scope definition and description of the topic:

- **Linguistics:** The studies made from this discipline focus on the analysis of words, sentences and texts. It looks for words or structures that relate to truthfulness or falsehood, what we can also understand as modalization or subjectivity marks, that is to say, the linguistic elements that are present in discursive activity, indicating the attitude of the speaking subject with respect to his interlocutor and his own utterances [22]. Researches relevant in this field are [14] and [4].
- **Psychology:** The studies made from this discipline focus on techniques or metrics that analyze people's behavior (extralinguistic information) to determine whether they are telling the truth or lying. In this field, it should be highlighted the research of [23].
- **Sociology:** The studies made from this discipline focus on the sociological analysis of social media to extract cues of veracity or falsehood. It is related to the NLP task of fact-checking. Researches of [3] and [24] are relevant in this field.
- **Computer Science / Natural Language Processing:** The studies made from this discipline focus on the development of tools and techniques that enable to automate the detection and analysis of deception. Relevant researches on this field are [4], [15] and [9].

As can be seen, every discipline provides a valuable approach to deception detection. The combination of different points of view can help to improve our research and offer a more comprehensive and multi-level perspective. As Gravanis et al. [4] proved in their research, psychologists in cooperation with linguistics experts and computer scientists revealed that the potential deceivers use certain language patterns.

Nevertheless, previous researches primarily present two gaps: (1) they address the extraction of features from an particular discipline or approach (2) they focus in the impact of misinformation in the wide public and does not pay attention to the explainability. These different approaches had not been taken into consideration in a unifying research. It is necessary to fill in the gaps within disciplines and create a continuum between them, i.e., learning how intentions are embodied in discourse, examining the way NLP techniques can extract pragmatic information or how sociological theory can be applied to fake news detection.

For this reason, our expected impact is to enhance the ability to distinguish misinformation, by unifying and compelling different approaches, and to revise transparency and explainability of deception detection systems for everyone.

3. Main Hypothesis and Objectives

The main hypothesis that introduces this PhD thesis is that it is possible to delimit a generalizing taxonomy for deception markers that can be applied to more than one language, textual genre or discursive modality.

Subsequently, the main objective is to detect deception from written texts automatically extracting content-based features, contextual-based features, readability features, virality features, and linguistic richness features (task 1). The extraction of these features had been studied separately, but not as an integrative study like in this thesis proposal. In addition, our aim is to analyze existing NLP systems that include text generation to justify whether it is false or true information, focusing on its explainability to build an inclusive artificial intelligence (task 2).

As Bonet-Jover et al. [5] explain, the research community is approaching the deception detection task focusing in extracting content-based features or context-based features. In our research we try to unify and interrelate both types of features, in addition to readabality and virality features. It is necessary an integrative approach because a piece of misinformation contains physical content (such as body text, picture or video) and nonphysical content (such as emotion, opinion or feeling) [25].

Basically, the research will focus on analyzing linguistic cues to deception detection. Based on this, we will find some sub-objectives that will be part of the first task (O1, O2, O3, O4), others that will respond to the second task (O5) and others that will be common (O6, O7).

The specific scientific sub-objectives are presented below:

O1. To collect and analyze information about linguistic cues (content-based, context-based markers, virality degree and readability features) that are present in the deceptive texts. We will mainly focus in the researches of [9], [11], [14], [15], [16] and [2].

At this point, virality features can be analyzed as a complementary element. A deceptive message is deceptive not for its probability to go viral, but for its own characteristics. However, could be interesting to study at the same time if it exists any relationship between false information and a hidden intention of the sender to go viral. As Saquete et al. [17] shows, dissemination of false information can be motivated by ideological or economic interests. For this reason, virality is considered as a feature, but it is not the center of our investigation.

O2. Linking the approaches showed in previous researches, to develop a generalizing marker classification that can be applied to more than one language, textual genre and modality. We analyze the current researches to unify and extract the features that are relevant to this topic.

O3. To extract the methodology used in previous NLP tools for the deception detection purpose. It is necessary to revise researches, competitions and available datasets.

O4. To develop a methodology to assess the degree of deceptiveness/credibility of an information.

O5. To employ and test NLP tools that analyze text to detect deception. Analyze if they present a sufficient degree of explainability that provides a clear and universally accessible justification for the veracity or falsity of the information. It is necessary to implement explainability measures in NLP systems that can help every person to understand the reasons why an information is deceptive or believable in an autonomous way.

O6. To obtain results and compare them with the state of the art. Recognize weak points and implement improvements in the research, both in terms of features to detect deception and in terms of explainability degree.

O7. To carry out scientific dissemination of the processes and results obtained from the research throughout the development of the PhD thesis.

The time planning is divided into four years. As a summary, we show in which objectives the research will be focused during this project, as can be seen in **Figure 1**.

4. Methodology

The methodology used in this work starts from a multidisciplinary point of view, focusing in elaborate an integrative research. As it was said before, the study around deception converges linguistic, psychological, sociological and computational approaches. To pursue the objective of this research, we will also combine both analytical and empirical methodologies.

On the one hand, our methodology focus in carrying out an exhaustive analysis of the discourse in relation with different variables, which can provide a wider view of linguistic markers to apply them

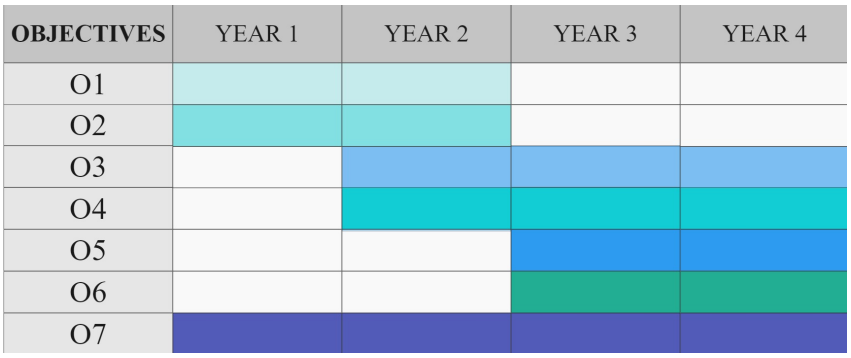

Figure 1: Gantt Chart that presents the time planning of the PhD research.

in NLP tasks. Following previous work, to extract linguistic cues it is necessary to work with written texts, preferably online resources that can be compiled easily. As happened in similar researches [2], the compilation of images is a limitation of the study. This integrative analysis extracts, compiles and test which features are relevant to take into consideration at detecting deception in written texts. The main goal is to find relevant and distinctive markers that can be generalizing in different languages, textual genres or modalities.

On the other hand, we present an empirical methodology in which we carry out a process of experimentation centered in the implementation of existing NLP systems for deception detection and test their accuracy. After that, our examination focus on their explainability, that it to say, how NLP tools display their information and outputs to build an inclusive understanding of NLP tools.

Therefore, the study is approached from a conjunction between exploration and action to create a solid theoretical foundation that can also be put into practice, and ending with a conclusion phase where the results obtained are evaluated quantitatively and qualitatively.

5. Research issues to discuss

To determine primary research issues of this PhD thesis, we used the ABC of systematic literature review [26]. In this survey, they introduce various research question development tools. These are mainly applied to health science, but we can use them to create research questions that are relevant to establish the basis of this PhD thesis.

To begin the research process of this PhD thesis establishing the following research questions:

- ***RQ1:*** *Is there a relationship between certain linguistic markers (word classes, verb tenses, pronoun usage, syntax, etc.) and the expression of truthfulness/falsehood?*
  As many studies have shown, it is possible to extract falsehood or veracity of a written text from its linguistic components [9], [14] or [4], among others. However, a compilation and improvement of a classification is an unfinished task.
- ***RQ2:*** *Is it possible to create a methodology for falsehood detection that is generalizing (different textual typologies, registers and contexts), unbiased and applicable?*
  As we introduced before, we focus in the researches [9], [11], [14], [15], [16] and [2]. Based on the information collected from this researches (displayed at Table 1), it is possible to collect all the linguistic deception cues and create a preliminary taxonomy for our research. After analyzing which cues are repeated or similar, the classification is as follows:
  - **Expressiveness:** terms referring to any type of expression of emotions, mental images or affection (positive or negative).
  - **Quantity:** referring to any type of measurement related to words or sentences. It is a content-independent variable, it is only quantifiable.

  - **Complexity/readability:** measured by unique words, complex syntax, etc. There are tools or algorithms that can calculate readability index.
  - **Cognitive processes:** referring to expressions of internal thinking or perceptual/sensory processes.
  - **Certainty:** referring to terms that show uncertainty, certainty or concreteness. This construct can contain two more concrete variables: specificity (what is more specific shows more certainty), and immediacy (when this specificity is related to time or space).
  - **Participation:** referring to expression of participation or distancing from what is being said. Mostly use of pronouns (autorreference or outer-reference).
  - **Informality:** measured by mistakes, punctuation marks, etc.
  - **Virality:** measured by the classification by Esteban-Bravo et al. [2].
  - **Information richness:** measured by the classification by Zhou et al. [11].

- ***RQ3:*** *Is it possible to generate justifications for the veracity or falsehood of a text that are understandable and accessible to everyone?*
  As Kotonya et al. [7] show, explainable machine learning shows a great deal of promise despite the particularly challenging nature of the problem. This study shows that is necessary to continue the research on the explainability and accessibility of NLP tools.
- ***RQ4:*** *Can detection of false information help people to avoid being misled, but especially people with some difficulty to investigate autonomously whether an information is truthful or not?*
  As Moreda et al. [8] affirm in the CLEAR.TEXT project, it is important to secure the ability to access written information for all people, thereby reducing the risk of exclusion for those with cognitive disability.
- ***RQ5:*** *Is there any relationship between fake news and virality?*
  As Vosoughi et al. [24] proved, falsehood diffuses significantly farther, faster, deeper, and more broadly than the truth. However, this report arises from a exclusively sociological approach, leaving behind the analysis of linguistic features that can motivate the spread of information (simpler syntax, briefer sentences, more common words, etc.).

Acknowledgments

This research work is part of the project "NL4DISMIS Natural Language Technologies for dealing with dis- and misinformation" (CIPROM/2021/21) (funded by Generalitat Valenciana (Conselleria d'Educació, Investigació, Cultura i Esport)), and of the R-D projects "CORTEX: Conscious Text Generation" (PID2021-123956OB-I00) (funded by MCIN/ AEI/10.13039/501100011033/ and by "ERDF A way of making Europe").

References

[1] J. A. Gottfried, E. Shearer, News use across social media platforms 2016, 2016. URL: https://api.semanticscholar.org/CorpusID:156553104.

[2] M. Esteban-Bravo, L. D. L. M. Jiménez-Rubido, J. M. Vidal-Sanz, Predicting the virality of fake news at the early stage of dissemination, Expert Systems with Applications 248 (2024) 123390. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417424002550. doi:10.1016/j.eswa.2024.123390.

[3] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain, The science of fake news, Science 359 (2018) 1094–1096. URL: https://www.science.org/doi/10.1126/science.aao2998. doi:10.1126/science.aao2998.

[4] G. Gravanis, A. Vakali, K. Diamantaras, P. Karadais, Behind the cues: A benchmarking study for fake news detection, Expert Systems with Applications 128 (2019) 201–213. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417419301988. doi:10.1016/j.eswa.2019.03.036.

[5] A. Bonet-Jover, A. Piad-Morffis, E. Saquete, P. Martínez-Barco, M. Ángel García-Cumbreras, Exploiting discourse structure of traditional digital media to enhance automatic fake news detection, Expert Systems with Applications 169 (2021) 114340. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417420310277. doi:10.1016/j.eswa.2020.114340.

[6] L. Quijano-Sánchez, F. Liberatore, J. Camacho-Collados, M. Camacho-Collados, Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police, Knowledge-Based Systems 149 (2018) 155–168. URL: https://linkinghub.elsevier.com/retrieve/pii/S095070511830128X. doi:10.1016/j.knosys.2018.03.010.

[7] N. Kotonya, F. Toni, Explainable automated fact-checking: A survey, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 5430–5443. URL: https://www.aclweb.org/anthology/2020.coling-main.474.

[8] P. Moreda, B. Botella, I. Espinosa-Zaragoza, E. Lloret, T. J. Martin, P. Martínez-Barco, A. Suárez Cueto, M. Palomar, et al., Clear. text enhancing the modernization public sector organizations by deploying natural language processing to make their digital content clearer to those with cognitive disabilities (2023).

[9] L. Zhou, J. K. Burgoon, J. F. Nunamaker, D. Twitchell, Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications, Group Decision and Negotiation 13 (2004) 81–106. URL: http://link.springer.com/10.1023/B:GRUP.0000011944.62889.6f. doi:10.1023/B:GRUP.0000011944.62889.6f.

[10] L. Bode, E. K. Vraga, In related news, that was wrong: The correction of misinformation through related stories functionality in social media, Journal of Communication 65 (2015) 619–638. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jcom.12166. doi:https://doi.org/10.1111/jcom.12166. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcom.12166.

[11] C. Zhou, K. Li, Y. Lu, Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness, Inf. Process. Manage. 58 (2021). URL: https://doi.org/10.1016/j.ipm.2021.102679. doi:10.1016/j.ipm.2021.102679.

[12] T. Q. J. F. N. Judee K. Burgoon, J. P. Blair, Detecting deception through linguistic analysis, in: M. R. Z. D. D. D. C. S. J. M. T. Chen, Hsinchun (Ed.), Intelligence and Security Informatics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 91–101.

[13] M. L. Newman, J. W. Pennebaker, D. S. Berry, J. M. Richards, Lying words: Predicting deception from linguistic styles, Personality and Social Psychology Bulletin 29 (2003) 665–675. URL: https://doi.org/10.1177/0146167203029005010. doi:10.1177/0146167203029005010. arXiv:https://doi.org/10.1177/0146167203029005010, pMID: 15272998.

[14] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, H. Cooper, Cues to deception., Psychological Bulletin 129 (2003) 74–118. URL: https://doi.apa.org/doi/10.1037/0033-2909.129.1.74. doi:10.1037/0033-2909.129.1.74.

[15] V. Hauch, I. Blandón-Gitlin, J. Masip, S. L. Sporer, Are Computers Effective Lie Detectors? A Meta-Analysis of Linguistic Cues to Deception, Personality and Social Psychology Review 19 (2015) 307–342. URL: http://journals.sagepub.com/doi/10.1177/1088868314556539. doi:10.1177/1088868314556539.

[16] R. Santos, G. Pedro, S. Leal, O. Vale, T. Pardo, K. Bontcheva, C. Scarton, Measuring the impact of readability features in 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. 1404–1413. URL: https://aclanthology.org/2020.lrec-1.176.

[17] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, M. Palomar, Fighting post-truth using natural language processing: A review and open challenges, Expert Systems with Applications 141 (2020) 112943. URL: https://linkinghub.elsevier.com/retrieve/pii/S095741741930661X. doi:10.1016/j.eswa.2019.112943.

[18] P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, W. Mansour, et al., Overview of the clef–2021 checkthat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 12th International Conference of the CLEF Association, CLEF 2021, Virtual Event, September 21–24, 2021, Proceedings 12, Springer, 2021, pp. 264–291.

[19] I. Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, C. Hansen, J. G. Simonsen, MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 4685–4697. URL: https://aclanthology.org/D19-1475. doi:10.18653/v1/D19-1475.

[20] J. Reyes-Magaña, L. E. A. Vega, Forcenlp at fakedes 2021: Analysis of text features applied to fake news detection in spanish, in: IberLEF@SEPLN, 2021. URL: https://api.semanticscholar.org/CorpusID:238208223.

[21] R. A. Monteiro, R. L. S. Santos, T. A. S. Pardo, T. A. d. Almeida, E. E. S. Ruiz, O. A. Vale, Contributions to the study of fake news in portuguese: new corpus and automatic detection results, Springer, 2018. doi:10.1007/978-3-319-99722-3_33.

[22] C. V. Cervantes, Modalización in diccionario de términos clave de ele, 2024. URL: https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/modalizacion.htm.

[23] Á. Almela, A Corpus-Based Study of Linguistic Deception in Spanish, Applied Sciences 11 (2021) 8817. URL: https://www.mdpi.com/2076-3417/11/19/8817. doi:10.3390/app11198817.

[24] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (2018) 1146–1151. URL: https://www.science.org/doi/abs/10.1126/science.aap9559. doi:10.1126/science.aap9559. arXiv:https://www.science.org/doi/pdf/10.1126/science.aap9559.

[25] X. Zhang, A. A. Ghorbani, An overview of online fake news: Characterization, detection, and discussion, Inf. Process. Manage. 57 (2020). URL: https://doi.org/10.1016/j.ipm.2019.03.004. doi:10.1016/j.ipm.2019.03.004.

[26] H. A. Mohamed Shaffril, S. F. Samsuddin, A. Abu Samah, The ABC of systematic literature review: the basic methodological guidance for beginners, Quality & Quantity 55 (2021) 1319–1346. URL: https://link.springer.com/10.1007/s11135-020-01059-6. doi:10.1007/s11135-020-01059-6.