



Fighting the Fake: A Forensic Linguistic Analysis to Fake News Detection

Rui Sousa-Silva^{1,2}

Accepted: 6 April 2022 / Published online: 28 April 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Fake news has been the focus of debate, especially since the election of Donald Trump (2016), and remains a topic of concern in democratic countries worldwide, given (a) their threat to democratic systems and (b) the difficulty in detecting them. Despite the deployment of sophisticated computational systems to identify fake news, as well as the streamlining of fact-checking methods, appropriate fake news detection mechanisms have not yet been found. In fact, technological approaches are likely to be inefficient, given that fake news are based mostly on partisanship and identity politics, and not necessarily on outright deception. However, as disinformation is inherently expressed linguistically, this is a privileged room for forensic linguistic analysis. This article builds upon a forensic linguistic analysis of fake news pieces published in English and in Portuguese, which were collected since 2019 from acknowledged fake news outlets. The preliminary empirical analysis reveals that fake news pieces employ particular linguistic features, e.g. at the levels of typography, orthography and spelling, and morphosyntax. The systematic identification of these features, which will allow mapping linguistic resources and patterns used in those contexts, contributes to scholarship, not only by enabling a streamlined development of computational detection systems, but more importantly by permitting the forensic linguistics expert to assist criminal investigations and give evidence in court.

Keywords Forensic linguistics · Cybercrime · Language crimes · Fake news · Fact-checking · Disinformation

✉ Rui Sousa-Silva
rssilva@letras.up.pt

¹ Faculty of Arts and Humanities, CLUP - Centre for Linguistics of the University of Porto, University of Porto, Porto, Portugal

² Faculdade de Letras, Universidade do Porto, Via Panorâmica, s/n, 4150-564 Porto, Portugal

1 Introduction

Over the last two years, the world has faced a serious COVID-19 pandemic, which resulted in millions of deaths, long-lasting global lockdowns, and limitations on free movement due, e.g., to travel bans. At the time of writing, the world is also witnessing a war triggered by Russia's invasion of Ukraine, which has originated polarized views and positions. Alongside this literal, health pandemic and this war, the world has also been stricken by two other inter-related, technologically motivated, metaphorical wars and pandemics: cybercrime and disinformation. Although neither of these phenomena emerged during the COVID-19 pandemic, they are known to have skyrocketed during the pandemic. In the UK, for instance, not only has the number of cybercriminal activities increased during the pandemic, but also this number was especially high during the period in which the lockdown policies and measures were the strictest [7]. Among the most common categories of cybercrime that were subject to this increase, the authors found frauds related to online shopping and auctions, and hacking of social media and email, targeting, in particular, individual victims. At the same time, the uncertainty underlying COVID-19 treatments and public policies, as well as the polarized views of the Russia–Ukraine war, enabled the exploitation of platforms such as those offered by online and social media to spread disinformation and conspiracy theories, which are among the main threats to public health [19] and, ultimately, to democracy [2].

However, neither cybercrime, nor disinformation are new phenomena that emerged as a result of the COVID-19 pandemic, or the Russia–Ukraine war. Indeed, the technological developments of the last decades have brought along new communication possibilities across the world, and, with them, new cybercriminal activities and structured disinformation campaigns. These activities, despite taking place online, are neither merely virtual, nor unrelated to offline crime; cybercrime has been found to mimic and adapt reality [40], by using technology-enabled (online) possibilities. Cybercriminals thus tend to use sophisticated technological methods and techniques to perform their criminal activities anonymously, including: cyber-trespass, e.g. unauthorized access to passwords, identity theft, or destruction of sensitive information; cyberporn, including e.g. illegal use of pornographic contents, unauthorized use of nudity, sexual exploitation, extortion and 'revenge porn' (in which contents such as nudes are disseminated publicly without permission); cyber-violence, e.g. defamation, cyberthreats, dissemination of dangerous/harmful contents, online harassment, cyber-bullying/ cyber-stalking and hate speech, often leading to physical/emotional trauma or death; and cyber-deception/theft, such as illegal access to information/materials online, theft of intellectual property online/digital piracy [74]. The latter group, in addition, includes new cybercriminal activities, such as 'doxing' (in which someone else's email address and real name is revealed online against their will) and fake news, disinformation and misinformation activities, which have been considered to pose threats to democracy worldwide [75].

The recent technological developments did not have an impact on more evident cybercriminal activities only. Journalism, too, has changed dramatically lately largely because the new information and communication technologies, and in

particular the sharing possibilities offered by the social media, now allow information to be disseminated immediately and widely without much effort. Alongside these developments, participatory journalism principles and policies, which have been promoted in schools of communication throughout the world over the last decades, gave ordinary citizens the power to collect, report, analyse and—in particular—disseminate news and information, actively and timely [5]. Of course, participatory (or citizen) journalism is not nefarious in itself, rather the opposite; its advantages are obvious. For the society in general, the active involvement of ordinary citizens in news-gathering results in an easier and faster access to news-worthy information; for media conglomerates, involving citizens in media communication meant that they could release part of their human (and consequently financial) resources, and thus significantly improve their financial sustainability prospects—and, in turn, eventually guarantee the independence of media outlets; for citizens, the main advantage of participatory journalism is that it confers them an opportunity to have a say on what information they value most.

Importantly, however, participatory journalism does not come without disadvantages, and its two main challenges cannot be backgrounded. The first, of a professional nature, is that ordinary citizens do not receive education and training in the field, so the facts that they witness will always be observed through the lens of a citizen and not of a professional, mainstream journalist. Professional journalists are academically trained to preserve the three core principles of journalism: factuality, objectivity, and neutrality. Ordinary citizens, as amateur journalists, on the contrary, do not receive specific training on how to guarantee these principles, regardless of how much civic education they may have received. The second challenge is of a deontological and even legal nature: unlike professional journalists, who are bound by ethical and deontological codes and principles, and who must follow the established codes of practice strictly, citizen journalists do not have to abide by any codes, and neither are they subject to disciplinary action taken by professional unions and associations (they can still, of course, be subject to legal action, although this will be rather unlikely in most cases). Consequently, this increases the possibility of ethical and legal issues occurring, potentially including propagation of mis- and disinformation.

Disinformation has long been identified as a serious problem, and different approaches have been adopted since to fight against it, including (semi-)automatic systems to identify users and networks that are known to spread fake news, machine-learning tools that screen fake news by analysing metadata, and especially fact-checking systems.

This article builds upon the circumstances underlying cybercriminal activities, in general, and disinformation, specifically, to argue that most of the systems proposed so far, including fact-checking systems, to a lesser or greater extent fail to effectively detect disinformation. This is because, as will be argued, disinformation does not necessarily offer false facts, but rather produces a slanted, biased, and manipulated version of often verified facts. Conversely, as will be argued, linguistic analyses of disinformation like the ones used in forensic contexts have the potential to establish whether a piece of news is truthful, or, conversely, whether it is fake.

The article is structured as follows: the next section discusses the disinformation phenomenon in the context of the ethical and legal principles of journalism. The article then discusses and defines the concept of disinformation and ‘fake news’, by exploring the anatomy of fake news. Subsequently, fake news detection methodologies are discussed, by contrasting the traditional fact-checking methods and the advantages of a linguistic approach. The subsequent sections present the data and methodology used in this study, followed by the results of the analysis, the discussion of the findings and the final remarks.

2 The Ethical and Legal Breaches of Journalism

The ‘fake news’ phenomenon cannot be discussed independently of media and communication studies, where ethical and legal issues have long been considered. Over the last decades, for instance, communication sciences furthered their self-reflection as several renowned mainstream journalists were involved in unethical and even illegal activities, including plagiarism scandals. In 2011, Johann Hari, former journalist of *The Independent*, was accused of plagiarizing quotes that his purported interviewees had previously given to other journalists, by passing off those quotes as a product of his own interviews. As a consequence, he was asked to return the Orwell Prize that he had been awarded in 2008, was suspended as a columnist of *The Independent*, and eventually resigned.

Hari’s case was neither the first, nor the last one involving renowned journalists. In 2003, *The New York Times*, too, was involved in a scandal after one of its most distinguished journalists, Jayson Blair, was accused of plagiarism and fabrication in his news stories. The investigation started after similarities were found between a story that Blair had written, and an article authored by reporter Macarena Hernandez, published one week earlier by *San Antonio Express-News*. Blair eventually resigned and his case (which is recounted in the documentary *A Fragile Trust: Plagiarism, Power, and Jayson Blair at The New York Times*) was widely used to discuss the boundaries of ethics, deontology, and malpractice in journalism.¹

In 2007, across the Atlantic, a reporter of the Portuguese newspaper *Público*, too, was found to have plagiarized, albeit in a form that is different from usual plagiarism: she was accused by the newspaper of having improperly reused texts from the *Wikipedia* and *NewScientist*, which she then translated into Portuguese to produce an article that was later published in the newspaper’s Sunday supplement. Her misdeeds included inadvertently leaving part of the text that she copied from the *Wikipedia* in English, although her article was written in Portuguese. She initially denied the accusations, by arguing that a news report is not an academic work, but eventually admitted to plagiarism.²

¹ See, e.g., <https://www.nytimes.com/2003/05/11/us/correcting-the-record-times-reporter-who-resigned-leaves-long-trail-of-deception.html>, last accessed 12.02.2022.

² See <https://www.publico.pt/2007/01/07/jornal/uma-forma-de-plagio-115478>, last accessed 12.02.2022.

These cases show that ethical and legal breaches, including but not limited to plagiarism and fabrication of data, can and do have serious consequences, in the first instance on journalists' lives, and then on society at large. A different case is that of participatory journalists: which disciplinary action is taken against them when they breach anonymity and confidentiality, or when they plagiarize, fabricate data, produce and spread mis-, disinformation and fake news, or are granted illegal access to—and share—sensitive information is unknown. Most likely, none is taken, since participatory journalists are in essence ordinary citizens who negligently and unwittingly attempt to emulate the work of mainstream journalists.

Altogether, these circumstances may appear to result in a greater propensity for spreading fake news. However, they also add another layer of difficulty to fake news detection because not all news pieces that are inherently fake are equally damaging [52]. Therefore, a clear definition of fake news is needed before appropriate detection methods and procedures can be proposed.

3 Defining Fake News, Mis- and Disinformation

'Fake news' has been a vigorous research topic for some time, although it has gained particular attention after the 2016 US Presidential primaries and general election campaign and the subsequent election of Donald Trump, in the USA, and Jair Bolsonaro, in Brazil (in 2018), and remains a topic of concern in democratic countries worldwide [48] due to its stealthy power to interfere with democratic systems [2, 75]. More recently, the COVID-19 pandemic has been fruitful ground for fake news spreading [11], often in association with conspiratorial narratives [19]. COVID-related disinformation has claimed, for instance, that ice-cream prevented COVID-19, that hydroxychloroquine was an effective medicine against the infection, that massive vaccination was simply a plan of a higher order to rule the world, or even that the infection was related to 5G—either because the technology helped spread the infection, or because anti-COVID vaccines were just a pretext to implant 5G chips in the human body. Some of these views, largely conspiratorial, were appropriated by extremist groups as an attempt to destabilize democratic systems by disrupting public policies.

Contrary to what is commonly believed, however, fake news is not a recent phenomenon. As reported [48], in the 1800 election involving Thomas Jefferson and John Adams, Jefferson paid a pamphleteer to spread rumours about his opponent, while Adams, in turn, made false accusations about Jefferson. Therefore, as the author writes, “[i]n the twenty-first century, the only surprising thing about noxious political propaganda is that anyone still finds it surprising” (p. 58).

Notwithstanding the fact that political propaganda, like deception, is as old as humankind, what one may find surprising is the new forms used to produce and propagate false, mis- and disinforming, slanted and biased information, which not only are an approximation to what is usually considered cybercrime, but more importantly an appropriation of many of the methods and tools used by cybercriminals. Methods used to produce and propagate fake news, much like those used for cybercriminal activities, are subject to mutations, as they change and adapt to

circumstances over time, and this raises the single biggest challenge to fight the fake news phenomenon: the difficulty in detecting them. Additionally, as was previously argued [75], this difficulty is increased by the broad, legal protection granted to false (political) speech, which constrains the viability of legal solutions to the problem. This issue is furthered by the ease with which disinformation propagates all over the world, across jurisdictions, which often find themselves impotent to act.

Fake news pieces have always existed also in the form of satirical reports, of which *The New Yorker's Borowitz Report*, the British *Private Eye* or the French *Charlie Hebdo* are all excellent examples. It is true that because satire uses well-defined resources, including humour, irony and exaggeration, these media outlets are easily identifiable by the readers as being satirical (not to mention the fact that a clear statement is usually made that those news stories are satirical); yet, the news that they report are not any less false. A distinct, more serious, insidious, and noxious form of fake news is where facts are distorted, and data are manipulated and presented with a slant, typically to the financial or political benefit of one party. The latter, rather than the former, is what is commonly referred to as 'fake news'.

Notwithstanding, the concept of fake news is highly pervasive, which explains why a consensus has not yet been reached on a clear definition. In its most general sense, fake news can be described as 'news that is not real,' which however can have diverse meanings in different contexts. The first, obvious defining feature of fake news is its failure to observe and guarantee the integrity of the facts, which explains why the topic has immediately attracted the attention of computer scientists and led to the development of fact-checking systems. However, as discourse analysts well know—and so do journalists and communication scientists, consciously or unconsciously—'fact' and 'reality' are not absolute values and can indeed be duplicitous phenomena. Therefore, although a strong argument has been made that 'fake news' should only be so labelled when they reproduce objectively false content [75], the crux of the matter is not simply whether an event is factually true or false, in absolute terms, but whether the facts are reported falsely, are slanted or biased.

As counsellors will know from experience, eyewitnesses often have conflicting views of the same event, and this does not always necessarily mean that one is lying, while the other is being truthful, but rather that they have different views of the witnessed event. The same holds true when language is used to report facts: even if the facts reported may be the same, *how* those facts are reported will produce different meanings for the reader. This is one of the features underlying hyperpartisan news—with a potential as well for 'fake news' detection [54], as can be empirically observed: in one case of Portuguese disinformation, while the mainstream media reported that a bank that had been subject to bailout would request another 1.25 million Euros, a fake news outlet reported that the Portuguese taxpayers would *pay* another 1.25 million Euros towards the bank. The two news reports do not differ with respect to the fact itself; the bank was *factually* going to be paid another 1.25 million Euros, so that fact is undisputable. Where those news pieces differ strikingly is in how that fact is reported: the agency attributed to the Portuguese taxpayers in the fake news piece, in comparison to the omission of agency in the mainstream media, is not innocent; rather the contrary, it aims to blame the government for 'outrageously' spending taxpayers' money on a bailout bank (and hence fuel the

citizens' mistrust in the democratic system, which could spend that money e.g. on treating cancer patients, instead). This is a clear example of a news piece whose truthfulness to the journalistic principles remains questionable, despite passing the fact-checking test.

3.1 The Anatomy of Fake News

An attempt has been made to define fake news by proposing a typology of scholarly definitions. Some authors [68] argue that fake news “appropriates the look and feel of real news” (p. 147), by imitating the look of mainstream media websites, the form of the articles, or even including attribution to photos—all of which can be indicative of truthful news pieces. Often, fake news reports include reference to external sources, as mainstream news reports do, and even the titles are an imitation of mainstream media's: “ABCnews.com.co”, “Denver Guardian”, “News Examiner”, “SubjectPolitics.com”, “YourNewsWire.com”—or, in Portugal, “Magazine Lusa”, which appropriates part of the name of the Portuguese news agency, “Agência Lusa”. Such an appropriation, as the authors claim, is an attempt of fake news outlets to gain some form of legitimacy and credibility from how the news looks. Previous research concurred [53], by reasserting that fake news mimics the real, by being packaged like news, although they are not subject to news norms.

Despite their look and feel, fake news pieces offer some degree of fabrication and especially manipulation, while partly or completely flouting the typical core principles of journalism: factuality, objectivity and neutrality. These principles have long been debated in the realm of communication sciences, as the portrait of the journalist as objective and neutral, in particular, clashes with the intrinsically subjective nature of human beings. Importantly, fake news stories typically misrepresent facts (even if those facts may be intrinsically true), by offering a manipulated report of those facts. Thus, fake news pieces presuppose not only an interference with the level of facticity, but also an intent to deceive, usually for financial or political reasons [53] (or, one may add, both).

The typology of fake news proposed by [68], based on a review of previous literature which used the term ‘fake news’, consists of six categories: news satire, news parody, fabrication, manipulation, advertising, and propaganda. It appears uncontroversial to assume, however, that some of these categories partially overlap, other categories (e.g., advertising or propaganda) are permissible text genres in specific contexts—as is the case of commercial or political campaigns, respectively, while other categories are not considered.

In this article, five categories of fake news are suggested, based on previous literature on the topic: (1) disinformation (including its next of kin misinformation); (2) bias (and its close relative hyperpartisanship); (3) clickbait; (4) sensationalism; and (5) satire (including parody). Although all these categories of fake news meet a crucial criterion to be considered fake—that of lack of facticity—this criterion needs to be balanced against another important criterion: that of intent to deceive [68]. Although satire (as well as parody) can be scored high in how they flout the principle of facticity, it has a low intent to deceive. The main communicative function of

satire is not to deceive the audience, but instead to ridicule someone (usually a high-profile individual or a specific, stereotyped group of people) for the entertainment of the audience. Sensationalism, in turn, doesn't primarily aim to ridicule a person or a group of people to amuse the audience; rather, it aims to appeal to the emotions of the audience, typically by backgrounding the core facts and foregrounding emotional or dramatic elements [43] that stir the feelings of the audience, customarily to gain popularity. The ultimate intention of sensationalism is not typically to deceive, but to be popular.

Clickbait, bias and disinformation, on the contrary, all show a degree of manipulation of facticity and cumulatively an intention to deceive. Clickbait consists of fabricating news stories and giving them catchy headlines to lure the audience, especially in the social media, to gain more website visits and, as a result, increase advertising revenue. Clickbait is used both by disinformation outlets and mainstream media. However, since it allows media outlets to obtain higher ranks in search engines, they are frequently used by fake news outlets to move up to a point where the news is believed by a general audience to be true. Bias, in turn, is based on a manipulation of facticity by producing slanted news with the intent of supporting (and benefiting) one point of view over the other [53]. Bias is typically associated with hyperpartisan pro-right movements, partly owing to the association of the fake news phenomenon to Donald Trump, in the USA, and Bolsonaro, in Brazil. However, pro-left hyperpartisan news outlets, as well as both anti-left and anti-right fake news outlets also exist.

Disinformation consists of news whose facticity is manipulated with a high intent to deceive the audience and thus obtain a financial and/or political gain. Disinformation and misinformation (both of which are commonly defined as information that is presented as truthful, but shown to be false later on [44]) are frequently used interchangeably. Such use is, however, inaccurate and a clear distinction should be made between the two: whereas misinformation is low on intent, as it can be false information that is produced inadvertently or recklessly and then propagated, disinformation is false information that is produced with an intent to deceive [4].

Therefore, given the complexity underlying fake news, it is less accurate to approach fake news as a binary system, in which a piece of news is judged to be true or false, than to place fake news in a continuum, where there are lies, half-truths and misleading (though mostly true) news.

3.2 Detecting Fake News

On the technological side, sophisticated computational systems have been developed to detect fake news. Two approaches to fake news detection have been discussed [25, 42]: the first is 'human intervention' and consists of recruiting human fact-checkers to verify information integrity and veracity; the other alternative is 'using algorithms' to identify fake content and validate the information sources. The former has several limitations, such as the time required to check the facts manually, possible inaccuracies resulting from human fatigue and, more importantly, depending on the

target subject and on the salience of the claim being made, the effectiveness of the fact-checking method may vary [27].

The computational alternative has some obvious advantages, in particular its ability to process a high volume of data quickly and homogeneously, without the interference of human fatigue, and the supposed resistance to bias. Such algorithms operate based on the content of the text (by focusing on the accuracy of the content-checking, more than on the reputation of the sources), on the dynamics behind the propagation of the message (i.e., they use the dynamics of the propagation to detect whether the information is credible or not), or on a ‘hybrid algorithm’ (i.e., an algorithm that builds on a group of features to feed a learning algorithm).

A different approach has been proposed [42] that consists of identifying the origins and patterns of evolution of false information. In this case, the authors produced an evolution tree based on an examination of the root content, the producers of the original source, and evolution patterns of false information spread on Twitter. Although this approach provides an insight into the origin and nature of false tweets, and in particular contributes to identifying information produced by non-credible sources or related bot accounts, the effectiveness of this approach to detect fake news is limited, not the least because credibility does not necessarily imply truthfulness. Previous doubts about the role and effectiveness of fact-checking and crowdsourcing approaches to fake news detection [27] challenge the relevance of considering the reputation of the sources in the detection procedure, and as has been aptly stated [25], some approaches to fake news detection are still naïve and require, at least, cross-checking, in addition to the need to consider the reputation of the sources, both historically and dynamically.

In the same vein, several authors have argued that fact-checking methods need to be streamlined [45, 58, 63]. Indeed, there has been a shift [38] in the process of determining the facts from the traditional “newsrooms, away from the public eye, as journalists considered conflicting reports, weighed up incoming information and made decisions on what to publish” (p. 665), to the public sphere, on social media platforms, which are overflowed with contradictory information and reports, as well as rumours and speculation, and where “confirmation and verification circulate via social interaction in a compressed news cycle” (p. 665). In this post-modern world of ‘alternative facts’, information has often been discussed in terms of post-truth [6], where “[f]acts are messy, difficult to determine and they are often dependent on interpretation” (p. 301). This, as the author states, not only erodes the traditional idea of the function of watchdog of the press over society—regardless of how crucial that may be [58]—but also blurs the lines separating truth, half-truths and blatant lies. Therefore, the messier the facts, and the more difficult it is to establish them, the harder it is to verify and check them, and hence to tell truths from lies. This is especially the case in a networked world, where ‘information’ is propagated instantly and often mechanically.

Complementarily, as found by a study conducted to assess the perceptions of fake news of journalists from Australia and from the UK [58], at least the journalists interviewed showed concern about a decrease in public trust in the media, while they sought a better understanding of how disinformation is propagated online, as well as of the process of automation of that spread by using bots.

3.3 The Case for Forensic Linguistics

Despite all effort made so far, appropriate fake news detection mechanisms have not yet been found, for some obvious—and other not so obvious—reasons. In addition to the difficulties underlying the distinction between facts and alternative facts, the technological approaches developed so far are likely to be inefficient [53], given that some fake news are based mostly on partisanship and identity politics, and not necessarily on outright misinformation and deception. Hence, since it is not the facts that are false, but rather how these facts are constructed, even the most sophisticated fact-checking and verification methods are bound to fail. Conversely, as partisan and hyperpartisan information is inherently expressed linguistically, this is a privileged room for linguistic analysis, in general, and forensic linguistic analysis, in particular.

However, no research has been conducted specifically in the field of forensic linguistics on the topic, and scarce research has been conducted using linguistic approaches in general—and the one conducted has focused more on demonstrating the existence of fake news pieces, than on detecting and analysing them. The latter is exemplified by a corpus linguistics analysis of Trump's discourse on Twitter, which highlighted his accusations that mainstream journalism was fake news, and investigated how he operated as a serial spreader of mis- and disinformation [56]. The method used by the authors consisted of building corpora of Trump's tweets, on the one hand, and of other politician's tweets, on the other. Subsequently, the authors conducted a corpus linguistic analysis, to extract word frequencies and thus determine the words and word clusters most frequently used by Trump in his tweets, as compared to word frequency and word clusters in the typical tweets of other politicians. These results were then classified according to a typology of four discrete strategies to help understand Trump's behaviour on Twitter: 'pre-emptive framing,' 'diversion,' 'deflection,' and 'trial balloon.' The study focused on the analysis of the lexical items, while discarding function words, to identify features of the rhetorical language used by Trump in his tweets. (Interestingly, by removing grammatical words, the authors discard words such as 'very,' 'the,' and 'and', all of which are highly frequent in Trump's discourse.) The authors concluded that the strategy of deflection is the strategy used predominantly in Trump's tweets to attack the messenger and change direction; in other words, the authors argue that Trump uses these accusations both to demonstrate his commitment and as a cover for his own propagation of mis- and disinformation, while framing it as truth.

This corpus linguistic analysis is a significant contribution to help understand the rhetorical discourse of Donald Trump on Twitter and shows the relevance of addressing fake news as more than a simple representation of false facts. Nevertheless, it is of limited application in forensic contexts, given its lack of robustness to serve as evidence in court cases involving dissemination of disinformation. Given the nature of fake news, mis- and disinformation and their potential embodiment of cybercriminal activities, a forensic linguistic analysis is able, not only to assist its investigation and detection, but also to provide evidence in legal cases.

Forensic linguistics, which consists of using applied linguistics theories, methods and approaches to study and respond to real-life, forensic problems, can be defined both in a broad and in a narrow sense [13, 16]. In a broad sense, forensic linguistics

subsumes three sub-areas: (a) the Written Language of the Law, (b) Spoken Interaction in Legal Contexts, and (c) Language as Evidence; the narrow definition of forensic linguistics, conversely, restricts the discipline to Language as Evidence. Therefore, Forensic Linguistics has been used over the last decades in several different contexts, from the provision of linguistic evidence in cases of plagiarism [65, 72], authorship attribution [35], and disputed meanings [10], to the analysis of contracts and statutes [51, 64], courtroom discourse [21] and police interaction [55], among others.

The interaction between language and the law is undeniable, and although it is particularly appealing to applied linguists, whose main concern is addressing language issues in real-life, it is also of interest to legal scholars and practitioners. Law is language [32] and although, as a social institution, it is manifested in non-linguistic, multimodal ways [50], it is largely a linguistic institution [32]: “[L]aws are coded in language, and the processes of the law are mediated through language” (p. 156). This view that law is a linguistic institution has motivated critical analyses both from linguists and legal practitioners. One excellent example of such critical analysis is the view of the relevant role of linguists, if given the appropriate tools, in preventing miscarriages of justice resulting from unreliable confessions, for instance by contributing their expertise to the analysis of the discursive structure and linguistic content of interrogations [1].

Arguably linguists operating in forensic contexts are qualified to use any of the tools and resources in their ‘forensic linguist’s toolbox’ [18], such as appraisal analysis to investigate the use of stance markers during cross-examination [31], phonetics to assess the reliability of voice recognition by ‘ear witnesses’ [28], narrative analysis to increase jurors’ understanding of expert witness testimony and to increase their confidence in the expert [61], an analysis of discourse processes and topic management in false confession contamination by police investigators [30], corpus linguistics to analyse attitude and emphasis in legal writing [26], entextualisation to assist the collection of oral evidence [55], or genre theory to help interpret diversionary justice [49], among many others.

Complex issues such as establishing whether a suspect knowingly and intelligently waived his rights, or whether such waive appears to have been coerced, if a confession is credible evidence of guilt, if an admission of involvement is due to the police deceptively promising leniency, or whether the reliability of a purported confession is questionable, all lend themselves well to the kind of analysis done by linguists [1]. Linguists have also acted as experts in court [71], including in legal cases of terrorism and murder (see e.g. [13, 17, 18, 35, 62]), plagiarism [15, 66, 72], trademark disputes [8, 9], or composition, identification, and assessment of adequacy of consumer product warnings [20, 70].

More recently, linguistic research has also focused on addressing cybercrime, i.e. crimes perpetrated online, connected devices [14]. Cybercriminal activities subsume hate crimes, threats, slander, libel and defamation, as well as fraud, identity theft (in particular via the creation of fake profiles in social media outlets), electronic vandalism and violation of intellectual property rights, when committed online, as part of both organized and unorganized crime. Research in this area contributed to assisting the investigation of cases of cybercrime, for instance to study identity assumption

and deceptive identity performances by undercover police officers in online investigations against child sex abusers [33, 34, 46], or moves and strategies of online grooming [12], so that police officers involved in the investigation of cases of paedophilia can pose as the children being groomed to eventually identify the paedophiles behind the computer screen.

In addition to the linguistics research conducted for purposes of identifying cybercriminal activities, linguistic evidence has been given in legal cases of cybercrime. One case in which a linguistic expert report was provided was a case of digital piracy [67], where contents such as books, films and music were illegally shared on a website, while the announcements were made on the Facebook page. The Prosecutor's Office in charge of the case had access to the identity of the website administrators but was unable to identify the administrators of the Facebook page, as the company would not release the information. The solution found by the Prosecutor's Office was to request an authorship analysis of the website and the Facebook page to ascertain the likelihood that both sets of documents were written by the same authors. Two other cases involved defamation. In one case, three emails were circulated among the staff of a department store allegedly denouncing illegal behaviour of the managers, including physical aggression. A third case consisted of analysing a set of SMS messages sent from unregistered phone numbers denouncing the alleged infidelity of a man. In all three cases, a forensic authorship analysis was conducted to identify patterns in the writing of the texts, and subsequently establish the consistency of those patterns across the texts in the same group, and their distinctiveness when compared to the text of other groups, as previously proposed [35].

The third case, in addition to authorship analysis, also involved sociolinguistic profiling. Sociolinguistic profiling is usually requested when the investigation does not have strong hypotheses about the identity of the author(s) and asks the linguist to analyse the texts to find clues to the age, gender, social and regional background of the writer. In the case in point, the linguistic analysis concluded that the writer used linguistic choices typical of African Portuguese, which allowed the investigation to narrow down the pool of suspects.

Altogether, these studies demonstrate the relevance of applying forensic linguistic approaches to disinformation detection and analysis.

The next section discusses the concept of fake news and argues that it is a form of disinformation, and hence potentially a language crime. Henceforth, the terms 'fake news' and 'disinformation' are used interchangeably to mean 'disinformation'.

4 Fake News as Language Crimes

Language is often used to commit crimes. Most crimes require from the defendant an *actus reus* and a *means rea*, and so usually involve some sort of physical violence, although it is not uncommon for crimes to be committed by speech or writing [71], rather the opposite; language is regularly used as the sole tool for committing unlawful acts [60], including threats, bribery, extortion, defamation, or solicitation of things such as murder, illicit sex or paedophile activities, but also perjury and blackmail. Such activities can be labelled 'language crimes' [60, 71] because they

are primarily accomplished through language, rather than through physical acts, and can lead to litigation. All these unlawful activities involve in one way or another an illegal speech act [3, 59] to be performed, which means that language is used, not simply to communicate information, but rather to do things—in this case, illegal. Language crimes such as solicitation, conspiracy, bribery, threatening and perjury all differ in terms of the elements of each crime, with respect to the extent to which the speech acts affect the recipient, and whether the speaker must be sincere, or only appear sincere [71]. Conversely, except for perjury, all language crimes share the following criteria: (a) they can be commissioned both directly and indirectly; (b) they require some kind of intent; and (c) they are committed (primarily) utilizing speech acts.

Language crimes, unsurprisingly, are boosted by online interaction, which is a fruitful ground for committing cybercriminal activities. On the one hand, the perceived anonymity of internet users has led to a general, widespread illegal behaviour online [41]. Therefore, to commit language crimes online, users don't necessarily need to be actually anonymous on the internet, but rather perceive themselves as being anonymous. On the other hand, the online, networked environment offers its users relative anonymity, enhanced by using public access computers (such as those available in cybercafés, public libraries, or other institutions) or IP address hide software and cloaking technology to erase their digital fingerprint. This is a problem to policing because it prevents the successful identification of the criminals involved, and consequently a successful prosecution of their crimes. In effect, the type of technology used for cybercriminal activities has allowed such activities to become more professional, stealthier, automated, much larger, more complex, and much different from normal routine activities. Sexting and other social media-originated crimes are illustrative of such complexity. Chans, internet fora and, ultimately, the dark web—where the users' anonymity is virtually impossible to be breached unless the users make a mistake, e.g. by sharing material containing identifying metadata—are all a fruitful ground for cybercriminal activities, in general.

In parallel, jurisdictions worldwide are faced with a serious challenge: communication is global, but regulation is local. While computer forensics has been given priority in fighting against cybercrime, this article argues that forensic linguistic analysis remains one of the most powerful tools, if not the most powerful tool, in the fight against cybercrime: whereas identifying elements resulting from technology use online can in extreme cases be virtually wiped, language can hardly be manipulated; as research on theories of idiolect and idiolectal style [17, 35, 73] has demonstrated, each speaker of a language has his/her own way of speaking or writing, whose forensic linguistic analysis enables the sociolinguistic profiling or even a positive identification of the individual speaker.

Disinformation meets all the criteria [71] of language crimes: (a) it can be commissioned both directly and indirectly; (b) it requires some kind of intent; and (c) it is committed (primarily) using speech acts. This makes forensic linguistic analysis particularly apt for the detection of fake news, in general, and disinformation, in particular, although the challenges for the linguistic analysis differ significantly from those faced e.g. in cases of authorship analysis, sociolinguistic profiling, disputed meanings or trademark disputes; what is of interest to linguists analysing

disinformation is not so much identifying the author of the information being propagated online, or the profiling of the author of such information to narrow it down to a group of suspects, but rather to establish whether the information is likely to be truthful or intentionally falsified, based on *the language used* and on *how* it is used.

The following section briefly illustrates how a preliminary forensic linguistic analysis can be useful in establishing whether a news piece is truthful or, on the contrary, a piece of disinformation.

5 Data and Methodology

For the purposes of detecting fake news, mis- and disinformation, the narrow definition of forensic linguistics will be adopted, since it focuses on analysing language as evidence (which includes the ability to assist the investigation). Consequently, this research concentrates on the work of the forensic linguist as expert witness. It applies linguistic theories and methods that have typically been used in forensic cases in general, because such theories and methods, together with a critical discourse analysis of fake and hyperpartisan news, have a significant potential to be developed and applied to the detection of disinformation and the development of disinformation preventive systems.

The data analysed in this article consist of information identified as being fake and/or hyperpartisan news published in English and in Portuguese. The data were collected between February 2019 and May 2020, to allow for a recent corpus of texts, and are part of the CONSPIRACY corpus (my own, DIY [47] open CORpus of News Scams and Partisan Information for Research and Analysis of CYbercrime). Despite the open nature of the corpus, the analysis presented in this article focuses on a corpus of small size, in order to allow for a fine-grained analysis of the patterns identified.

The CONSPIRACY corpus has been set up as a bilingual comparable corpus. Hence, it includes texts of the same type and genre, and with the same communicative function, in English and in Portuguese (but to our knowledge not translations). It is a synchronic, open corpus, which processes contemporary data and which will furthermore be expanded regularly over time, as more texts are processed and added to it. Although ultimately the corpus should allow some quantitative data to be extracted, particularly concordances (to establish which word combinations are more common in the corpus) and word frequency lists (i.e., which words are more frequently used in the texts included in the corpus), a corpus size was not pre-established, neither with respect to the number of words, nor regarding the number of texts.

The articles collected for the corpus were selected randomly from acknowledged fake news outlets, which are identified as being sources of fake news, mis- and disinformation, and were not individually fact-checked. This approach is known in computational linguistics as ‘silver standard’ [36]. The sources included in the corpus were: *ABCnews.com.co* (at the time of writing *keyc.tv*), *Conservative Daily Post*,

News Examiner,³ and *YourNewsWire.com* (at the time of writing *News Punch*), for English; *Bombeiros24*, *Semanário Extra*, *Jornal Diário Online*, *Magazine Lusa* and *Notícias do Viriato*, for Portuguese. The fake news outlets in English are mostly pro-right/ anti-left, although one of them is identified as both anti-left and anti-right (*Conservative Daily Post*) and another one is identified as inflammatory for both sides as well as conspiratorial (*YourNewsWire.com*, at the time of writing *News Punch*).

The criterion of individual authorship was not considered when collecting the texts, since the research focuses on the fake news genre, and not on the authorship analysis of the individual texts. Thus, any text published in the fake news, mis- and disinformation outlets was eligible. Once collected, the texts were saved as text files, as well as pre-processed and included in the corpus management tool *Corpógrafo* [57]. This tool, which is available at <https://www.linguateca.pt/corpografo/>, was favoured to the detriment of other corpora management tools, including commercially available software, because it offers several advantages: (a) it is a free, open-source tool; (b) it works online, although access is protected by username and password; and (c) it can process Portuguese, contrary to other tools, which have problems handling diacritics, among other language-specific features.

The methodology adopted in this research focuses, firstly, on conducting a corpus linguistics analysis of the texts in the two languages to identify peculiar words and word sequences, as well as unusual collocations (i.e., words that co-occur with other words). These can give an important insight into disinformation detection, especially as they may be indicative of flouting of the principle of neutrality, which is so cherished by mainstream journalists.

Subsequently, a detailed, fine-grained linguistic analysis of the texts—identical to the analyses done in forensic authorship scenarios—will be conducted, to identify idiosyncratic features of fake and hyperpartisan news. The systematic identification of these features aims to help map linguistic resources and patterns used in disinformation pieces. These features include, from a quantitative perspective, average paragraph, sentence and word length, use of punctuation, and idiosyncratic markers at the levels of typography, orthography and spelling, and morphosyntax. These are based on a preliminary study, which revealed that fake news, mis- and disinformation texts showed regular patterns in this respect.

Therefore, the analysis of the two sets of texts (English and Portuguese) included in the CONSPIRACY corpus consists of: (a) computing simple text statistics; (b) identifying structural linguistic patterns that are idiosyncratic, and which may be typical of fake news texts; and (c) analysing some discourse features, by applying principles of critical discourse analysis [23], to identify aspects like agency and theme and rheme [37]. These aspects are not exclusive of fake news texts; theme and

³ Access to News Examiner was later prevented due to restrictions imposed on users in Europe. At the time of writing (4 February 2022), the website returns the following error message: “451: Unavailable due to legal reasons. We recognize you are attempting to access this website from a country belonging to the European Economic Area (EEA) including the EU which enforces the General Data Protection Regulation (GDPR) and therefore access cannot be granted at this time. For any issues, contact web-support@newsexaminer.com or call 765–825-0581.”

rheme, for example, have long been studied as part of Systemic Functional Linguistics, SFL [37], to analyse written and spoken text.

Text statistics, structure and discourse have been demonstrated to be relevant in analysing online communications [39]. An identification of agency (i.e., *who* is responsible for *what*) and theme can be crucial, especially in cases of bias and hyperpartisan news because negative aspects associated with the opposing party are usually given agency and thematized, so as to be foregrounded, whereas the action of the interested party is only given agency and thematized if it is positive (otherwise, the information is backgrounded).

The most prominent features of fake news texts are then compared against contemporary mainstream news pieces, which were collected from quality papers in the same period. The results of the analysis are presented in the following section.

6 Analysis

6.1 Analysis of Text Statistics

When observing a fake news piece, in comparison with a news piece from mainstream media, one is intuitively led into assuming that mainstream media tend to use, if not longer words, at least longer sentences and paragraphs. The statistical analysis of the texts in the CONSPIRACY corpus, normalized over 1000 words, does not confirm this assumption entirely.

The analysis of the length of the paragraphs in words reveals that the mainstream media includes the shortest paragraphs (8 words only), with the shortest paragraph in the fake news sub-corpus including 11 words. Where the two corpora differ considerably is in the upper length of the paragraphs: whereas the longest paragraph in the English fake news sub-corpus is 40 words long, the English mainstream news sub-corpus includes three paragraphs containing 41, 44 and 46 words each, respectively. These findings are not fully consistent with the analysis of the Portuguese corpus, where the fake news sub-corpus includes the two longest (50 and 38 words), as well as the shortest paragraphs (4 and 5 words). In the range between 27 and 35 words, the mainstream news texts exceed the disinformation texts. The English and Portuguese corpora, however, coincide in that most of the paragraphs of the mainstream media fall in the middle, whereas the fake news texts tend to fall in the extreme ends.

The analysis of sentence length, in words, shows that disinformation texts reproduce both the shortest and the longest sentences, with the two longest sentences including 88 and 72 words, respectively. Interestingly, the fake news sub-corpora show a higher number of sentences in the range between 10 and 25 words, whereas the mainstream media texts are more pronounced in the range between 39 and 47 words. This pattern is also observed in the Portuguese corpus, where the length of most sentences falls in the middle (31 to 42, 45 to 51 and 55 to 61 words), whereas the disinformation texts occupy the extreme ends: the shortest sentence is 1-word long, while the longest sentence is 100-words long.

The observation of the word length, in characters, shows that the two sub-corpora are identical, although the disinformation texts tend to use the longest words: the longest word in this sub-corpus is 21 characters, whereas the longest word in the mainstream media sub-corpus is 18 characters long. An identical pattern is not observed in the Portuguese corpus: although one 30-character word is used once in each sub-corpus, the second and third longest words are to be found in the mainstream media sub-corpus. Notwithstanding, this sub-corpus has the highest volume of short words, whereas the fake news corpus occupies the middle area.

As far as punctuation is concerned, in the English corpus the fake news sub-corpus shows an overall higher volume of punctuation marks. The percentage of periods, commas and dashes is higher in the mainstream media sub-corpus (though not significantly), whereas the disinformation texts show more colons, semicolons, brackets and question marks than the mainstream media sub-corpus (the latter is seven times more frequent in the fake news sub-corpus). The Portuguese corpus shows striking differences: whereas the texts in the mainstream media sub-corpora use mostly periods, commas, semicolons and dashes, the fake news sub-corpus shows a higher volume of exclamation and question marks, as well as colons. The two sub-corpora do not differ significantly in the volume of brackets used. Therefore, an identical pattern cannot be observed in the English and the Portuguese texts, although there are similar patterns in some respects.

6.2 Structure

Structural elements are one of the aspects that have been considered to be of relevance in computer-mediated communication [39] and indeed a preliminary study conducted in 2019 confirmed that typography, orthography, syntax, and 'low level' features, in particular, have a high discriminatory potential to detect fake news. Interestingly, in this respect the English and the Portuguese corpora of fake news texts show identical features. The analysis of the texts in the CONSPIRACY corpus shows several instances of typographical idiosyncrasies, such as the lack of trailing spaces after punctuation. The text statistics presented above showed some insights into how punctuation is used differently in the disinformation texts, when compared to the mainstream media texts. In addition to the more frequent use of exclamation and question marks, the disinformation texts also show a high volume of ellipsis, unlike the mainstream media texts, where this feature is rare, if at all observed. Punctuation errors are also observed in both the English and the Portuguese corpora, although they are more noticeable in the Portuguese sub-corpus, where ungrammatical use of commas is frequent. The texts are also problematic concerning orthography, with frequent spelling errors. These errors include frequent replacement of letters with numbers that are visually similar. For example, in one of the Portuguese disinformation texts, 'crime' is spelled 'cr1me'.

Grammatical mistakes, too, are frequently observed in the disinformation texts, including problems with sentence structure (e.g., with subordinate clauses), lack of agreement in number (e.g., by confusing singular and plural forms)—and, in the case of Portuguese, problems with gender agreement (e.g., by resorting to inflection

in the masculine form where it should be feminine, and vice versa), and lack of prepositions. Grammatical inconsistencies are also found in verb tenses, as verb forms are mistakenly used.

Interestingly, special formatting, such as text in italics and bold, are frequently observed in disinformation texts, although this is not common in mainstream media texts. This is a relevant semiotic resource that is worth being further explored.

6.3 Discourse

Some of the most striking differences between the fake and mainstream media news texts can be found at the level of discourse. This is a relevant element when investigating disinformation because, as critical discourse analysis approaches have demonstrated [22–24, 69], analysing *how* something is said, in addition to *what* is said, gives a relevant insight into how power is exerted and how ideologies are constructed. The issue of ideology construction in fake news in general, and in hyper-partisan news, in particular, is imperative.

The study of word frequencies provides some important information in this respect, as it allows an overview of the vocabulary used, and consequently an identification of which semantic field is foregrounded. In this respect, the noun ‘truth’, as well as other words from the same semantic field (such as ‘real’ or ‘fact’), are very frequent, and are often employed in the form of an adverb, ‘truly’, which, in this case, works as an intensifier. In these cases, fake news portray themselves as the real media, contrary to the mainstream media, which are viewed as ‘establishment skills’ [48].

Adverbs are also frequently used in disinformation texts, however, to make evaluative statements. In one fake news story, the news text starts with the following sentence: “Disturbingly, Partners In Health is funded by George Soros Open Foundations and the Bill and Melinda Gates Foundation, and has Chelsea Clinton on its board.” Sentences starting with evaluative adverbs are not commonly found in mainstream media.

The headline of the same, recent publication in the fake news outlet *News Punch* read:

‘Contact Tracing’ Group Funded By Soros and Gates, With Chelsea Clinton on Board, Wins Gov’t Contract

Every sentence provides information that is given, the *theme* (hence known to the interlocutor), as well as information that is new, the *rheme*. In this sentence, “‘Contact Tracing’ Group Funded By Soros and Gates, With Chelsea Clinton on Board” is the theme, and “Wins Gov’t Contract” is the rheme. This structure, by providing the new information towards the end of the sentence, foregrounds, in the theme, the involvement of the people to which reference is made. A similar procedure can be observed in another headline from the same outlet:

President of Pro-Migrant Group, Who Opened His Home To Muslim Migrants, Found Beaten To Death

In this case, “President of Pro-Migrant Group, Who Opened His Home To Muslim Migrants” is the theme (the given information), whereas “Found Beaten To Death” is the rheme (the new information). In both cases, the headline includes a subordinate clause in the theme, “With Chelsea Clinton on Board” and “Who Opened His Home To Muslim Migrants”, respectively, which, not being relevant for the headline, foreground the sensationalist part of the headline. Interestingly, the rheme operates differently in the two cases: in the first example, the agency of those reported in the theme is emphasized, implying their direct involvement in the outcome, whereas the sentence in the second example is agentless, which moves to suggest some implicature between the Muslim migrants to whom he opened his house and the fact that he was found dead.

Agency is a powerful device when constructing ideologies because it allows positive actions to be directly attributed to us, while negative actions are attributed to our opponents, the *other*. Similarly, agency is usually omitted by an interested party both when they are responsible for some negative action, and when the opponent is responsible for a positive action. For instance, one of the fake news texts that is part of the Portuguese corpus reports that an old man who had sexually abused his granddaughter was sentenced to prison but later had his sentence reduced ‘by the court’, whereas the news story in the mainstream media simply reported that the man had his sentence reduced (by whom it was not stated; the assumption is that the reader will know who has the power to reduce sentences). By foregrounding the agency (of the court), the fake news story portrays the leniency of the court—which is part of the establishment—with such a hideous crime, thus gaining the sympathy of the popular audience; the quality newspaper, on the contrary, by not clearly specifying the agency, produces a more neutral news piece in that the focus of the story is not on who reduced the sentence, but rather the event (the fact that the man had his sentence reduced).

7 Discussion of the Findings

The analysis of the fake news, mis- and disinformation texts that are part of the small, DIY CONSPIRACY corpus, which was briefly presented in the previous section, shows that a study based on a forensic linguistic investigation of text statistics, text structure and discourse has the potential to help detect disinformation. This analysis builds on the assumption that not all news is factually false; some news reports are factual and will pass the ‘fact-checking’ exam.

The analysis of the text statistics, which has considered specifically paragraph, sentence and word length, as well as use of punctuation, revealed some patterns typical of fake news stories. The statistics regarding paragraph length revealed that, although some differences can be observed in disinformation texts in English and Portuguese when compared to their mainstream media counterparts, and the length of the paragraphs in fake news stories varies between very short paragraphs and very long paragraphs, the paragraph length of mainstream media texts tends to fall in the middle—i.e., they are neither too short, nor too long. It should be highlighted that the texts included in this corpus were all collected online, a

medium which influences the length of paragraphs; it is a known fact that, even when analysing the same news report, printed media usually reproduce longer paragraphs when compared to the same media online. An identical behaviour can be observed in sentence length, where fake and mainstream news texts differ, too: the disinformation texts occupy the extreme ends, with both the shortest and the longest sentences, whereas the length of sentences in the mainstream media texts falls in the middle. The English and Portuguese sub-corpora differ, however, in word length: in the English corpus, the longest words are used by the fake news texts, whereas in Portuguese, the longest words are used by the mainstream media.

The results of the analysis of punctuation across all the groups of texts show that English fake news texts use more punctuation than mainstream media texts, whereas in the Portuguese the amount of punctuation is identical. It should be highlighted that Portuguese and English use punctuation differently, and whereas in English, a moderate use of punctuation in news reports is equated with neutrality, in Portuguese, given in particular the more complex syntax, more punctuation is expected from well-written texts. A pattern that has been observed in both English and Portuguese is the use of punctuation that is typically associated more with opinionative and emotional texts, than with informative texts: question and exclamation marks. This is consistent with a higher level of subjectivity, which is proportional to a lower level of objectivity and facticity. It is important to note that not all media outlets, disinformation or mainstream, adopt a writing style that can be consistently considered ‘fake’ or ‘mainstream’, and the fact that different news outlets from the same group have different styles may have an impact on the overall statistics. More in-depth research, with larger corpora, will be needed to clarify this point.

The analysis of the structural elements reveals that disinformation texts are prone to grammar and spelling mistakes at different levels, as well as to an erratic use of punctuation and idiosyncratic typography. It is noteworthy that one of the spelling mistakes found in the Portuguese corpus, ‘cr1me’, where the ‘i’ was replaced with a numeral, ‘1’, is typical of the use of bots, which is consistent with previous studies which reported that bots are often used for fake news production and propagation. Contrary to mainstream news reports, disinformation texts frequently use evaluative adverbs to emphasize the hideousness of the actions described in their stories, as the sentence from *News Punch* reproduced in the previous section demonstrates. This type of structure is not usually found in mainstream media.

The discourse analysis of the texts shows that fake news reports systematically attribute the agency of what is popularly portrayed as negative to the opposing party (the *alien*, the *other*), commonly referred to as *othering*, while agency is foregrounded when associated with positive actions when these are attributed to the interested party. This is the case, in particular, when negative actions are attributed to the establishment. This emphasis on agency of establishment for negative acts and backgrounding of that agency in the case of positive facts, e.g. by using passive or active voice, contributes to constructing an anti-establishment stance, which is often associated with disinformation [48].

Overall, the features analysed—text statistics, structure, and discourse—concur in demonstrating that fake news texts employ identical identifying patterns. This

suggests, based on the overlapping patterns in the English and Portuguese fake news texts, that some of these patterns are language-agnostic.

8 Final Remarks

This research was conducted over a small, DIY corpus of texts collected online from publicly identified sources of disinformation and mainstream media (with a focus on quality newspapers). The findings are promising and have a significant potential to contribute to scholarship, not only by enabling a streamlined development of computational detection systems, but more importantly by permitting the forensic linguistics expert to assist criminal investigations and give evidence in court. As Brennen [6] strongly argues, “[i]n our postmodern world, it is important to consider the value of evidence, particularly in light of an understanding that in a socially constructed reality interpretation is always intertwined with fact” (p. 301). This claim gains particular relevance in a world of alternative facts.

This research, however, has some limitations. The first one is the corpus size. Although texts have been added to the corpus since the preliminary study, it is still rather small to allow for inferences to be made; a continuous enlargement of the corpus will allow an analysis of the statistical significance of the findings to be made. One practical problem underlying the corpus collection is that fake news outlets tend to be very ephemeral, or not to be available worldwide or permanently. Since the beginning of the corpus collection, several disinformation outlets changed their name, URL, or were no longer made available.

An additional issue that is worth considering in future work is the consistency of the corpus texts. In the current version of the CONSPIRACY corpus an attempt has been made to ensure that the sub-corpora were as consistent as possible, hence its being structured into four sub-corpora: two corpora of disinformation and two corpora of mainstream news, two in English and two in Portuguese. All texts being equal, this procedure is appropriate. Nevertheless, it is now evident that not all texts are equal, and whereas most of the texts published by fake news media are poorly written, there are also exceptions where the texts are written with a degree of sophistication that is likely to exceed the readers’ expectations.

Likewise, disinformation pieces can hardly be considered a consistent text genre, rather the contrary; this analysis confirms previous studies that considered disinformation as a ‘genre blending’ [53], where aspects of traditional journalism are combined with behaviours that are alien to mainstream journalists: the lack of facticity, objectivity and neutrality. This genre blending also reflects on the language used in the disinformation pieces. For example, it is common to find entire paragraphs of disinformation reports that are plagiarized from mainstream media texts. This not only influences the text statistics, but also contributes to the complexity of the genre blending.

Additionally, in the future it will also be necessary to investigate each group of fake news texts in relation to its particularly society, since, as is argued by Friedman [29], legal systems are interconnected with the society where they are enforced. This will be important to assess what the author refers to *impact*: “behavior that is tied

causally, in some way or other, to some particular law, rule, doctrine, or institution” (p. 2).

Notwithstanding, this study undeniably demonstrates the worth of a forensic linguistic analysis to identify the main features of disinformation and which route to follow in the future.

Acknowledgements I would like to thank the anonymous reviewers for their positive feedback and, in particular, for their comments and suggestions, which certainly contributed to an improved version of this article.

Funding This publication was supported by Project UIDB/00022/2020, FCT-Fundação para a Ciência e Tecnologia, Portugal, through national funds.

References

1. Ainsworth, Janet. 2010. Miranda Rights: Curtailing Coercion in Police Interrogation: The Failed Promise of *Miranda v. Arizona*. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard and Alison Johnson, 111–125. London: Routledge.
2. Ainsworth, Janet. 2020. The Toxic Proliferation of Lies and Fake News in the World of Social Media: Is it Time for the Law to ‘unfriend’ Facebook? In *Social Media in Legal Practice*, ed. Vijay K. Bathia and Girolamo Tessuto. London: Routledge.
3. Austin, John Langshaw. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.
4. Bednar, Peter M., and Christine Welch. 2008. Bias, Misinformation and the Paradox of Neutrality. *Informing Science: The International Journal of an Emerging Transdiscipline* 11: 85–106. <https://doi.org/10.28945/441>.
5. Bowman, Shayne, and Chris Willis. 2003. We Media: How Audiences are Shaping the Future of News and Information. *We Media*.
6. Brennen, Bonnie. 2009. The Future of Journalism. *Journalism: Theory, Practice and Criticism* 10 (3): 300–302. <https://doi.org/10.1177/1464884909102584>.
7. Buil-Gil, David, Fernando Miró-Llinares, Asier Moneva, Steven Kemp, and Nacho Díaz-Castaño. 2021. Cybercrime and Shifts in Opportunities During COVID-19: A Preliminary Analysis in the UK. *European Societies* 23: S47–S59. <https://doi.org/10.1080/14616696.2020.1804973>.
8. Butters, Ronald R. 2010. The Forensic Linguist’s Professional Credentials. *International Journal of Speech, Language and the Law* 16 (2): 237–252. <https://doi.org/10.1558/ijsl.v16i2.237>.
9. Butters, Ronald R. 2012. Forensic Linguistics: Linguistic Analysis of Disputed Meanings: Trademarks. In *The Encyclopedia of Applied Linguistics*, ed. Carol A. Chapelle. Oxford: Wiley-Blackwell.
10. Butters, Ronald R. 2021. Trademarks: Language that One Owns. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 364–381. London: Routledge.
11. Ceron, Wilson, and Mathias-Felipe de-Lima-Santos and Marcos G. Quiles. 2021. Fake News Agenda in the Era of COVID-19: Identifying Trends Through Fact-Checking Content. *Online Social Networks and Media* 21: 100116. <https://doi.org/10.1016/j.osnem.2020.100116>.
12. Chiang, Emily, and Tim Grant. 2017. Online Grooming: Moves and Strategies. *Language and Law/Linguagem e Direito* 4 (1): 103–141.
13. Coulthard, Malcolm, and Alison Johnson. 2007. *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.
14. Coulthard, Malcolm, and Rui Sousa-Silva. 2016. Forensic Linguistics. In *What are Forensic Sciences? Concepts, Scope and Future Perspectives*, ed. Ricardo Jorge Dinis-Oliveira and Teresa Magalhães. Pactor: Lisboa.
15. Coulthard, Malcolm, Alison Johnson, Krzysztof Kredens, and David Woolls. 2010. Four Forensic Linguists’ Responses to Suspected Plagiarism. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard and Alison Johnson, 523–538. London: Routledge.
16. Coulthard, Malcolm, Alison May, and Rui Sousa-Silva. 2021. *The Routledge Handbook of Forensic Linguistics*, 2nd ed. London: Routledge.

17. Coulthard, Malcolm. 2004. Author Identification, Idiolect and Linguistic Uniqueness. *Applied Linguistics* 25 (4): 431–447.
18. Coulthard, Malcolm. 2010. In My Opinion. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard and Alison Johnson, 473–486. London: Routledge.
19. Darius, Philipp, and Michael Urquhart. 2021. Disinformed Social Movements: A Large-Scale Mapping of Conspiracy Narratives as Online Harms During the COVID-19 Pandemic. *Online Social Networks and Media* 26: 100174. <https://doi.org/10.1016/j.osnem.2021.100174>.
20. Dumas, Bethany K. 2010. Consumer Product Warnings: Composition, Identification, and Assessment of Adequacy. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard and Alison Johnson, 365–377. London: Routledge.
21. Eades, Diana. 2016. Erasing Context in the Courtroom Construal of Consent. In *Discursive Constructions of Consent in the Legal Process*, ed. Susan Ehrlich, Diana Eades, and Janet Ainsworth, 71–90. Oxford: Oxford University Press.
22. Fairclough, Norman, and Ruth Wodak. 1997. Critical Discourse Analysis. In *Discourse Studies: A Multidisciplinary Introduction—Discourse as Social Interaction*, ed. Teun A. van Dijk, 258–284. London: SAGE.
23. Fairclough, Norman. 1995. *Critical Discourse Analysis: The Critical Study of Language*. London: Longman.
24. Fairclough, Norman. 2001. *Language and Power*. London: Longman.
25. Figueira, Álvaro., and Luciana Oliveira. 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>.
26. Finegan, Edward, and Benjamin T. Lee. 2021. Legal Writing: Attitude and Emphasis: Corpus Linguistic Approaches to ‘legal language’: Adverbial Expression of Attitude and Emphasis in Supreme Court Opinions. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 48–63. London: Routledge.
27. Flynn, D.J., Brendan Nyhan, and Jason Reifler. 2017. The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics. *Political Psychology* 38: 127–150. <https://doi.org/10.1111/pops.12394>.
28. Fraser, Helen. 2019. The Reliability of Voice Recognition by ‘ear witnesses’: An Overview of Research Findings. *Language and Law/Linguagem e Direito* 6 (2): 1–9. https://doi.org/10.21747/21833745/lanlaw/6_2a1.
29. Friedman, Lawrence M. 2016. *Impact: How Law Affects Behavior*. Cambridge: Harvard University Press.
30. Gaines, Philip. 2018. Discourse Processes and Topic Management in False Confession Contamination by Police Investigators. *International Journal of Speech Language and the Law* 25 (2): 175–204. <https://doi.org/10.1558/ijssl.34951>.
31. Gales, Tammy, and Lawrence M. Solan. 2017. Witness Cross-Examinations in Non-Stranger Assault Crimes: An Appraisal Analysis. *Language and Law/Linguagem e Direito* 4 (2): 108–139.
32. Gibbons, John. 1999. Language and the Law. *Annual Review of Applied Linguistics* 19: 156–173. <https://doi.org/10.1017/S0267190599190081>.
33. Grant, Tim, and Nicci MacLeod. 2016. Assuming Identities Online: Experimental Linguistics Applied to the Policing of Online Paedophile Activity. *Applied Linguistics* 37 (1): 50–70. <https://doi.org/10.1093/applin/amv079>.
34. Grant, Tim, and Nicci MacLeod. 2018. Resources and Constraints in Linguistic Identity Performance—A Theory of Authorship. *Language and Law/Linguagem e Direito* 5 (1): 80–96.
35. Grant, Tim. 2021. Text Messaging Forensics—Txt 4n6: Idiolect-Free Authorship Analysis? In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 558–575. London: Routledge.
36. Hahn, Udo, Katrin Tomanek, Elena Beisswanger, and Erik Faessler. 2010. A Proposal for a Configurable Silver Standard. In *Proceedings of the Fourth Linguistic Annotation Workshop*, 235–242. <https://aclanthology.org/W10-1838>.
37. Halliday, Michael A. K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
38. Hermida, Alfred. 2012. Tweets and Truth: Journalism as a Discipline of Collaborative Verification. *Journalism Practice* 6 (5–6): 659–668. <https://doi.org/10.1080/17512786.2012.667269>.
39. Herring, Susan C. 2004. Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In *Designing for Virtual Communities in the Service of Learning*, ed. Sasha Barab, Rob Kling, and James H. Gray, 338–376. Cambridge: Cambridge University Press.

40. Holt, Thomas J., and Adam M. Bossler. 2014. An Assessment of the Current State of Cybercrime Scholarship. *Deviant Behavior* 35 (1): 20–40. <https://doi.org/10.1080/01639625.2013.822209>.
41. Hughes, Danny, Paul Rayson, James Walkerdine, Kevin Lee, Phil Greenwood, Awais Rashid, Corinne May-Chahal, and Margaret Brennan. 2008. Supporting Law Enforcement in Digital Communities Through Natural Language Analysis. *Computational Forensics*. https://doi.org/10.1007/978-3-540-85303-9_12.
42. Jang, S. Mo., Tieming Geng, Jo-Yun Queenie Li, Ruofan Xia, Chin-Tser Huang, Hwalbin Kim, and Jijun Tang. 2018. A Computational Approach for Examining the Roots and Spreading Patterns of Fake News: Evolution Tree Analysis. *Computers in Human Behavior* 84: 103–113. <https://doi.org/10.1016/j.chb.2018.02.032>.
43. Kilgo, Danielle K., and Vinicio Sintá. 2016. Six Things You Didn't Know About Headline Writing: Sensationalistic Form in Viral News Content From Traditional and Digitally Native News Organizations. *#ISOJ* 6 (1): 111–130.
44. Lewandowsky, Stephan, Werner G. K. Stritzke, Alexandra M. Freund, Klaus Oberauer, and Joachim I. Krueger. 2013. Misinformation, Disinformation, and Violent Conflict: From Iraq and the “War on Terror” to Future Threats to Peace. *American Psychologist* 68 (7): 487–501. <https://doi.org/10.1037/a0034515>.
45. Lotero-Echeverri, Gabriel, Luis M. Romero-Rodríguez, and Maria Amor Pérez-Rodríguez. 2018. ‘Fact-Checking’ vs. ‘Fake News’: Periodismo de Confirmación como Recurso de la Competencia Mediática contra la Desinformación. *Index Comunicacion* 8 (2): 295–316.
46. MacLeod, Nicci, and Tim Grant. 2017. “go on cam but dnt be dirty”: Linguistic Levels of Identity Assumption in Undercover Online Operations Against Child Sex Abusers. *Language and Law/Linguagem e Direito* 4 (2): 157–175.
47. Maia, Belinda. 1997. Do-It-Yourself Corpora ... With a Little Bit of Help from Your Friends! In: *PALC '97 Practical Applications in Language Corpora*, ed. Barbara Lewandowska-Tomaszczyk and Patrick James Melia. Lodz: Lodz University Press, pp 403–410.
48. Marantz, Andrew. 2019. *Anti-Social: Online Extremists, Techno-utopians, and the Hijacking of the American Conversation*. New York: Viking.
49. Martin, J.R., and Michele Zappavigna. 2016. Exploring Restorative Justice: Dialectics of Theory and Practice. *International Journal of Speech Language and the Law* 23 (2): 215–242. <https://doi.org/10.1558/ijsl.v23i2.28840>.
50. Matoesian, Gregory M., and Kristin Enola Gilbert. 2021. Multimodality in Legal Interaction: Beyond Written and Verbal Modalities. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 245–264. London: Routledge.
51. Monzó-Nebot, Esther, and Javier Moreno-Rivero. 2020. Jurilinguistics: Ways Forward Beyond Law, Translation, and Discourse. *International Journal for the Semiotics of Law: Revue Internationale de Sémiotique Juridique* 33 (2): 253–262. <https://doi.org/10.1007/s11196-020-09721-w>.
52. Moura, Ricardo, Rui Sousa-Silva, and Henrique Lopes Cardoso. 2021. Automated Fake News Detection Using Computational Forensic Linguistics. In *EPIA2021—20th EPIA Conference on Artificial Intelligence: Lecture Notes in Artificial Intelligence (LNAI)*, ed. Goretí Marreiros, Francisco Melo, Nuno Lau, Henrique Lopes Cardoso and Luís Paulo Reis. Springer.
53. Mourão, Rachel R., and Craig T. Robertson. 2019. Fake News as Discursive Integration: An Analysis of Sites That Publish False, Misleading, Hyperpartisan and Sensational Information. *Journalism Studies* 20 (14): 1–19. <https://doi.org/10.1080/1461670X.2019.1566871>.
54. Potthast, Martin, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A Stylometric Inquiry into Hyperpartisan and Fake News. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 231–240. <https://doi.org/10.18653/v1/P18-1022>.
55. Rock, Frances. 2021. Witnesses and Suspects in Interviews: Collecting Oral Evidence: The Police, the Public and the Written Word. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 112–126. London: Routledge.
56. Ross, Andrew S., and Damian J. Rivers. 2018. Discursive Deflection: Accusation of “Fake News” and the Spread of Mis- and Disinformation in the Tweets of President Trump. *Social Media and Society* 4 (2): 1–12. <https://doi.org/10.1177/2056305118776010>.
57. Sarmento, Luís, Belinda Maia, and Diana Santos. 2004. The Corpógrafo—A Web-Based Environment for Corpora Research. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

58. Schapals, Aljosha Karim. 2018. Fake News: Australian and British Journalists' Role Perceptions in an Era of "alternative facts." *Journalism Practice* 12 (8): 976–985. <https://doi.org/10.1080/17512786.2018.1511822>.
59. Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
60. Shuy, Roger W. 1993. *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom*. Oxford: Blackwell.
61. Shuy, Roger W. 2018. Telling Our Stories: Inside and Outside of Court. *Language and Law/Linguagem e Direito* 5 (1): 1–18.
62. Shuy, Roger W. 2021. Terrorism and Forensic Linguistics: Linguistics in Terrorism Cases. In *The Routledge Handbook of Forensic Linguistics*, ed. Malcolm Coulthard, Alison May, and Rui Sousa-Silva, 445–462. London: Routledge.
63. Singer, Jane B. 2018. Fact-Checkers as Entrepreneurs: Scalability and Sustainability for a New Form of Watchdog Journalism. *Journalism Practice* 12 (8): 1070–1080. <https://doi.org/10.1080/17512786.2018.1493946>.
64. Solan, Lawrence M. 2012. Linguistic Issues in Statutory Interpretation. In *The Oxford Handbook of Language and Law*, ed. Peter M. Tiersma and Lawrence M. Solan, 87–99. Oxford: Oxford University Press.
65. Sousa-Silva, Rui. 2013. Investigating Academic Plagiarism: A Forensic Linguistics Approach to Plagiarism Detection. *International Journal for Educational Integrity* 10 (1): 31–41.
66. Sousa-Silva, Rui. 2019. Plagiarism Across Languages and Cultures: A (Forensic) Linguistic Analysis. In *Handbook of the Changing World Language Map*, ed. Stanley D. Brunn and Roland Kehrein, 1–21. Springer. https://doi.org/10.1007/978-3-319-73400-2_191-1.
67. Sousa-Silva, Rui. 2021. Linguística Forense no Combate e Prevenção do Cibercrime. In *Cibercriminalidade: Novos Desafios, Ofensas e Soluções*, ed. Inês Sousa Guedes and Marcus Alan de Melo Gomes. Pactor: Lisboa.
68. Tandoc, Edson C., Zheng Wei Lim, and Richard Ling. 2018. Defining "Fake News." *Digital Journalism* 6 (2): 137–153. <https://doi.org/10.1080/21670811.2017.1360143>.
69. Tannen, Deborah, and James E. Alatis. 1986. *Languages and Linguistics: The Interdependence of Theory, Data, and Application*. Washington, DC: Georgetown University Press.
70. Tiersma, Peter M. 2002. The Language and Law of Product Warnings. In *Language in the Legal Process*, ed. Janet Cotterill, 54–71. London: Palgrave Macmillan UK.
71. Tiersma, Peter M., and Lawrence M. Solan. 2012. The Language of Crime. In *The Oxford Handbook of Language and Law*, ed. Peter M. Tiersma and Lawrence M. Solan, 340–353. Oxford: Oxford University Press.
72. Turell, Maria Teresa. 2008. Plagiarism. In *Dimensions of Forensic Linguistics*, ed. John Gibbons and Maria Teresa Turell, 265–299. Amsterdam: John Benjamins.
73. Turell, Maria Teresa. 2010. The Use of Textual, Grammatical and Sociolinguistic Evidence in Forensic Text Comparison. *The International Journal of Speech, Language and the Law* 17 (2): 211–250.
74. Wall, David S. 2001. Cybercrimes and the Internet. In *Crime and the Internet*, ed. David S. Wall, 1–17. London: Routledge.
75. Walters, Ryan M. 2018. How to Tell a Fake: Fighting Back Against Fake News on the Front Lines of Social Media. *Texas Review of Law and Politics* 23 (1): 111–179.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.