ORIGINAL PAPER

Investigating Appraisal and the Language of Evaluation in Fake News Corpora

Radoslava Trnavac[1] · Nele Põldvere[2]

Abstract

The present corpus study, which is grounded in Appraisal Theory, investigates evaluative language use in fake news in English. The primary aim is to find out how and why, if at all, evaluative meanings are construed differently in fake news compared to genuine news. The secondary aim is to explore potential differences between types of fake news based on contextual factors. The data are from two carefully-designed corpora containing both fake and genuine news: a single-authored corpus and a multi-authored corpus. Both corpora contain false information that is meant to deceive, but they also differ from each other in terms of register, genre and the motivational goals of the authors. Through qualitative and quantitative analyses, we show that there are systematic differences in the occurrence of Appraisal expressions across fake and genuine news, with Appraisal being more common in the former. However, the exact nature of the affective, dialogic and modal expression of fake news is influenced by contextual factors that, so far, have largely been ignored in fake news research. Therefore, the study has important implications for the development of fake news detection systems based on data sources of different kinds, a task which is in grave need of the input of corpus linguists.

✉ Nele Põldvere
  nele.poldvere@ilos.uio.no

  Radoslava Trnavac
  rtrnavats@hse.ru

1   The School of Linguistics, The National Research University Higher School of Economics, Moscow, Russian Federation

2   Department of Literature, Area Studies and European Languages, University of Oslo, Oslo, Norway

🖄 Springer

Introduction

Fake news as a media phenomenon has been around for a long time. However, it is only recently, in the 21st century, that it has become a global threat affecting democracies around the world, largely due to technological advancements and the democratization of news media where anyone can post, share and like whatever is in their reach, often with little control or supervision by reliable authorities. Digital media platforms have also made it easier for journalists to fabricate news and to evade editorial control. Therefore, it is of utmost importance that linguists, and corpus linguists in particular, contribute to the description and explanation of the characteristic features of fake news, these features being integral to the development of automatic fake news detection systems. Early research in linguistics has already made important contributions, for instance, by identifying general linguistic differences between fake and genuine news based on carefully designed corpora (e.g., Asr & Taboada, 2019; Grieve & Woodfield, 2023; Sousa-Silva, 2022).

With a focus on deception and disinformation in fake news, the present study sets out to break new ground on the language of fake news in English by focusing on a specific linguistic phenomenon, namely, *evaluation*. The study is grounded in Appraisal Theory (Martin & White, 2005), a widely recognized framework in linguistics that has the potential to shed new light on how and why, if at all, evaluative meanings are construed differently in fake news compared to genuine news—the primary aim of this study. Consider the fake news headline in (1), taken from an American far-right news website, *The Gateway Pundit*.

(1) Germany: 96% of Latest Omicron Patients were <u>FULLY</u> Vaccinated—<u>Only</u> 4% Unvaccinated

Despite the presence of numbers, the headline still comes across as rather evaluative. By means of linguistic expressions of emphasis (*fully*) and counter-expectation (*only*), underlined, as well as capital letters, the headline clearly aligns itself with an anti-vaccine sentiment associated with right-wing groups to promote an ideological agenda on an online platform. In this way, it seems different from the way news is usually disseminated in traditional news media, that is, with a greater degree of objectivity (Tandoc et al., 2021). Note, however, that the features in (1) might not be the same across all types of fake news, or all types of genuine news, in the same way that language use in general is highly contextualized and influenced by a variety of factors in discourse. The use and function of Appraisal expressions in particular have been found to be sensitive to a variety of factors that influence the way in which speakers and writers evaluate events and situations in the world (Kaltenbacher, 2006; Fuoli, 2012; Põldvere et al., 2016; among others). Therefore, in this study we also pursue a secondary, more exploratory aim to find out whether the Appraisal expressions are the same across different types of fake news or whether they are influenced by contextual factors. The factors that we explore are *register*, *genre* and the *motivation* for the journalist to lie—all of which have largely been ignored in previous research in Natural Language Processing (NLP) where the bulk of research on the language of fake news

is from (see Section 'Previous research on evaluation in fake news'). We define and distinguish between register and genre in line with Biber and Conrad (2019). While register is understood as any language variety associated with a particular communicative purpose or situational context, genre focuses on the "conventional structures used to construct a complete text" within the language variety (Biber & Conrad, 2019: 2). This secondary aim is exploratory due to the relatively simple design of the study where corpora of different kinds are examined to shed early light on the potential evaluative differences between types of fake news.

For this, we analyse and compare two corpora of fake and genuine news which differ from each other along the parameters above. While one of them is collected from a traditional newspaper focused on news reporting where the author's motivation to lie appears to have been personal (e.g., reputational and material), in the other one the authors are presumed to have been driven by ideological motives for publication in online news websites with a focus on opinion-based journalism (the example in (1) above). We refer to the former corpus as the *single-authored corpus*, because all the news stories, fake and genuine, have been written by one author, Jayson Blair. The latter corpus, the *multi-authored corpus*, contains news stories by seven other authors, all of whom have written both fake and genuine news. The advantage of corpora of this kind is that they give us a controlled context for the study of fake news (see Section 'The corpora'). On the one hand, then, the corpora differ from each other in terms of register (news reporting vs. opinion), genre (newspaper vs. news website) and the specific motivational goals of the authors (personal vs. ideological), which together may influence the choice and occurrence of evaluative features in the texts.[1] On the other hand, both corpora contain fake news in its general, widely recognized sense in NLP in terms of veracity (false information) and honesty (intent to deceive), namely, *disinformation* (Tandoc et al., 2018). Therefore, they may be characterized by a core set of evaluative resources to reflect the journalists' general communicative purposes (see Section 'Defining fake news' for a view of fake news as a register in its own right).

To the best of our knowledge, Appraisal Theory has not been applied to fake news research so far, so the present study has the potential to add considerably to explanations for the differences between fake and genuine news, as well as to the description of the characteristic features of fake news for practical applications in automatic detection model development. In addition, it fills an unfortunate gap, and paves the way for further research, in our understanding of the role of contextual factors such as register, genre and motivation, or a combination of these, thus calling into question the empirical validity of broad definitions of fake news in current research.

After the Introduction, Section 'Background' presents the background to the study, including definitions of fake news, previous research on evaluation in fake news, and a presentation of Appraisal Theory as a suitable framework for a study of the language of fake news. In Section 'Data and methods', we describe the main similarities and differences between the single-authored (Jayson Blair) and

---

[1] Note that the names of the corpora, single-authored and multi-authored, do not reflect these contextual differences. Instead, they make reference to the sizes of the corpora in terms of the number of authors.
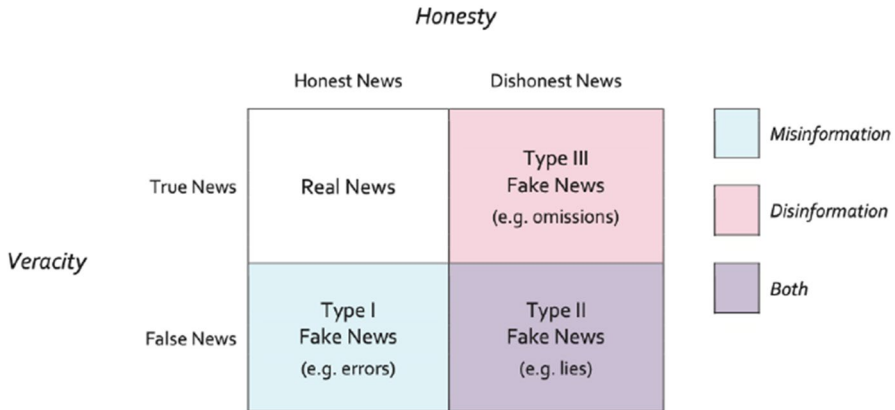
Fig. 1 Typology of fake news (Grieve & Woodfield, 2023: 13)

multi-authored corpora, followed by a more detailed overview of the decisions made in the annotation of the corpora for Appraisal expressions. Section 'Results and discussion' compares the results within and between the corpora, along with a general discussion of what the results might mean for fake news research. Finally, Section 'Conclusion' summarizes the study.

Background

Defining Fake News

Despite the relatively young age of the field, fake news has been defined and operationalized in a variety of ways, which makes it difficult to propose a standard definition. A useful starting point is the typology of fake news in Grieve and Woodfield (2023); see Figure 1. The typology is organised around two main dimensions: veracity and honesty. The notion of veracity has been a key topic of fake news research within NLP, where machine learning systems are trained to automatically distinguish between true and false news (Conroy et al., 2015; Potthast et al., 2018; Rashkin et al., 2017; Volkova et al., 2017; among others). These studies have been less concerned with the notion of honesty, or the author's *belief* in the veracity of the news item. It is the combination of these two dimensions that gives us four types of news. Typical news is when the news story is true and the author believes it to be so. Divergence from this ideal situation results in three types of fake news: (i) Type I Fake News is unintentional false news, such as errors, (ii) Type II Fake News is intentional false news, such as lies, and (iii) Type III Fake News is news that is true but produced with the intent to deceive, such as omissions, selective reporting and propaganda (see Tandoc et al., 2018).

In this study, we are primarily concerned with Type II Fake News. As Figure 1 shows, this type of fake news shares features with misinformation, or Type I Fake News, in that both contain false information, but an important difference between

them is the deceptive element in Type II Fake News. We refer to Type II Fake News as 'disinformation', in line with much of previous research. In fact, it is deceptive news that Grieve and Woodfield (2023) consider to be of greatest societal concern, due to the difficulty for humans to identify it. Moreover, such news stories share with each other a particular configuration of communicative and situational characteristics related to veracity and honesty, as well as aspects of 'newsness' such as the lack of objectivity (cf. Tandoc et al., 2021). Together, these characteristics are reflected in the linguistic features which are conventionally associated with the news stories and which are 'functionally adapted to the communicative purposes and situational contexts' of the texts from that register (Biber & Conrad, 2019: 2; see Section 'Previous research on evaluation in fake news' for examples). Therefore, in line with Grieve and Woodfield (2023), we understand of fake news as a news *register* that contains false information and where the communicative intent of the speaker/writer is to deceive. Register variation is highly pervasive (Li et al., 2022) and so it is possible that there is a great deal of homogeneity in how fake news is expressed linguistically across contexts of use.

Previous Research on Evaluation in Fake News

As previously mentioned, much of what we know about the language of fake news originates from NLP, which aims to develop automatic fake news detection systems based on linguistic and other features. The same applies to studies that have focused on the evaluative features of fake news in particular. The classification of evaluation in NLP is based on the underlying intuition that spreaders of fake news use 'emotional communication, judgment or evaluation of affective state' (Conroy et al., 2015; see also Hancock et al., 2011). For example, Rashkin et al. (2017) set out to investigate a range of evaluative features in a number of lexicons to distinguish between trusted and fake news sources. The lexicons were Linguistic Inquiry and Word Count (LIWC), a widely used general lexicon in social sciences, a sentiment lexicon with subjective words, a lexicon for hedges, and an intensifying lexicon from Wiktionary. The authors found fake news to be characterized by more subjective language use (e.g., *brilliant, very clear*), modal stance adverbs (e.g., *inevitably*), superlatives (e.g., *most*), hedges (e.g., *claims*) and negative markers (e.g., *nothing*), many of which are words that can be used to exaggerate (for similar results, see Volkova et al.'s (2017) study of suspicious and verified Twitter news accounts).

In linguistics, the topic of fake news is still in its infancy, which means that the studies that have been conducted so far have been rather broad, covering a wide range of linguistic resources and patterns rather than those marking evaluation more specifically. However, these studies provide important early insight into the affective, dialogic and modal expression of fake news based on linguistic theory. Sousa-Silva (2022), for example, considers agency to be a powerful resource of fake news disseminated for ideological purposes because it 'allows positive actions to be directly attributed to us, while negative actions are attributed to our opponents, the *other*' (emphasis in original).

In their analysis of fake and genuine news written by *The New York Times* journalist Jayson Blair (see Section 'The corpora' for details), Grieve and Woodfield (2023) found stance to play an important role in explaining Blair's linguistic strategies in each case. When he was telling the truth, Blair tended to write more densely and with greater conviction, reflected by 'a more persuasive and confident stance towards the information he is reporting and his sources' (Grieve & Woodfield, 2023: 56). Some of the lexico-grammatical features that contributed towards this style were suasive verbs (e.g., *allow, decide, determine*), prediction modals (e.g., *will, would*), *by*-passives (e.g., *Mr. Malvo's court-appointed guardian … was rebuffed by both the police and prosecutors*) and public verbs (e.g., *agree, report, suggest*). Within public verbs, specifically, by far the most common verb in Blair's genuine news was *to say* to introduce quotations and reported speech, but only in the past tense form. The present tense form, illustrated in (2), was almost exclusively used in his fake news.

(2) At moments, Ms. Anguiano says, she can picture her son in an Iraqi village, like the ones she has seen on television, surrounded by animals and the Iraqi people he has befriended. (Grieve & Woodfield, 2023: 61)

According to Grieve and Woodfield, this may be due to an element of doubt in the present tense form, which along with downtoners (e.g., *nearly*, *slightly*), and *wh*-relatives used to provide additional information about anonymous non-experts (e.g., *Muhammed was shown the soldier by a friend who was a doctor at the hospital*), were some of the ways in which Blair expressed greater uncertainty and lower specificity in his fake news stories. In addition, the authors found a greater number of attributive adjectives (e.g., *exculpatory evidence*), adverbs and emphatics (e.g., *very*, *extremely*) in Blair's fake news. These types of words usually have evaluative function and often are avoided in newspaper writing 'because they tend to add inconsequential information while risking weakening or editorialising statements' (Grieve & Woodfield, 2023: 54).

By grounding their research in Register Analysis and the understanding that differences in communicative purpose are reflected in linguistic structure (Biber, 1988; Biber & Finegan, 1989), Grieve and Woodfield have been able to provide, not only a comprehensive account of the variation in Blair's fake and genuine news, but a theoretical basis for the interpretation of the results. Specifically, there are differences in Blair's fake and genuine news because of his intent to deceive and to inform, respectively. However, it would be difficult to extend this theoretical premise to cases where the news stories may be false but where there is no intent to deceive, because the communicative purpose of both fake and genuine news is then the same (i.e., the distinction is along the dimension of veracity rather than honesty). The situation gets even more muddied when the fake news stories are identified based on source reputation, as is often the case in NLP and linguistics research. Not all fake news sources disseminate mis- or disinformation all the time, thus adding further variability to the findings. Furthermore, there is a tendency for fake news researchers to conflate different types of news contexts (e.g., newspapers, news websites, social media posts, debates) in search of a core set of linguistic features that characterize fake news as a whole (e.g., Rashkin et al., 2017). While fake news is a news register with its own particular configuration of communicative and situational characteristics and

purposes (cf. Section 'Defining fake news'), it is nevertheless the case that there are many other contextual factors that might influence the way in which people express themselves when they lie. Other aspects of register and genre such as where the news article is published, its style and intended audience are equally important to consider, as are the specific motivational goals of the speaker/writer (see also Potthast et al., 2018 on the role of hyperpartisanship, which, however, is beyond the scope of this study). Therefore, we follow Grieve and Woodfield in their methodological rigor, while at the same time zooming in on the role of evaluation, and contextual factors, in how fake news is disseminated in English.

Appraisal Theory

Appraisal Theory (Martin & White, 2005) is a powerful tool for engaging with discourse analysis, since it is one of the most comprehensive linguistic frameworks for studying grammatical and lexical resources which convey explicit or implicit evaluative meanings. Authors frequently express approval or disapproval for things, people, behaviour or ideas in the news discourse and a great deal of the Appraisal framework was first developed in relation to that (see Shizhu & Jinlong, 2004). Over the years, Appraisal has been refined by Bednarek (2008), White (2012), Ngo and Unsworth (2015), and it has grown gradually, providing comprehensive descriptions of academic language, film reviews and business discourse, among others (e.g., Carretero & Taboada, 2014; Fuoli, 2012; Hood, 2010). Extension to fake news research is therefore a natural next step.

Appraisal Theory has the three main systems: Attitude, Engagement and Graduation. Firstly, Attitude is concerned with emotional (Affect), ethical (Judgment), and social or aesthetic evaluation (Appreciation). As pointed out by Martin (2003: 173), Judgment and Appreciation might be interpreted as institutionalizations of Affect—Judgment as Affect recontextualized to control behaviour (what we should and should not do), Appreciation as Affect recontextualized to manage taste (what things are worth).

Secondly, Engagement is informed by Bakhtin's/Voloshinov' widely influential notions of heteroglossia and dialogism under which all verbal communication is a response to what has been said/written before and an anticipation of what will be said/written after (Voloshinov, 1973; Bakhtin, 1981; Martin & White, 2005: 92). Monoglossic assertions contrast with heteroglossic options in the sense that they do not overtly recognize alternative positions. Heteroglossic resources can be divided into two broad categories according to whether they are 'dialogically expansive' or 'dialogically contractive' in their intersubjective functionality. The distinction draws on the degree to which an utterance actively makes allowances for dialogically alternative positions (dialogic expansion) or acts to challenge their scope (dialogic contraction). Figure 2 shows a fine-grained classification of Engagement resources along with examples. According to the Appraisal framework (Martin & White, 2005: 97–98), Disclaim and Proclaim are instances of dialogic contraction in the sense that they contain either propositions that are rejecting some contrary position (Disclaim) or propositions that are highly warrantable,
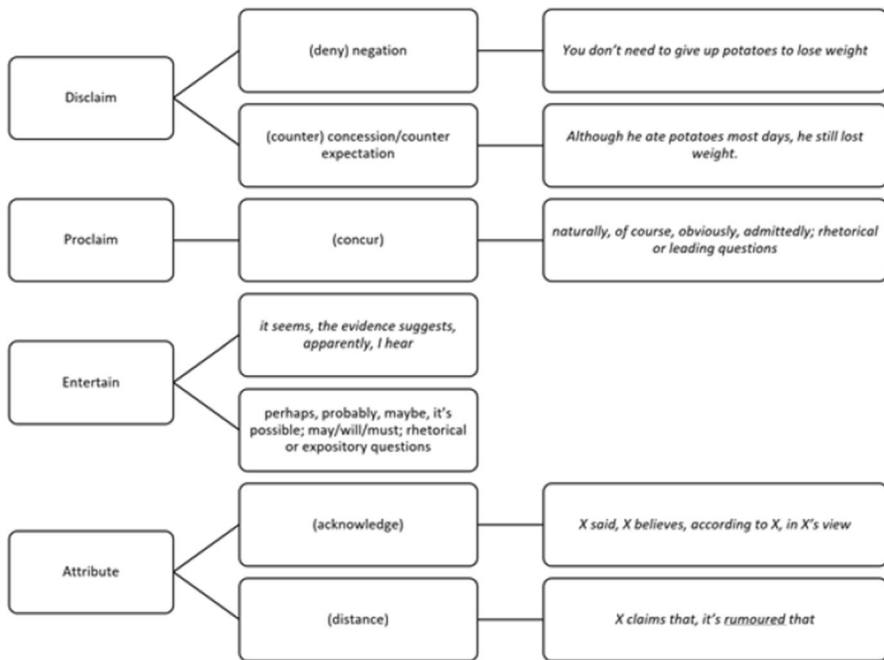
Fig. 2 Taxonomy of Engagement resources (Trnavac & Taboada, 2023: 239)

well-founded and plausible (Proclaim). Examples of the former are markers of denial (negation) and counter-expectation. Entertain and Attribute are instances of dialogic expansion. Entertain contains propositions that are grounded in their own subjectivity, represented as one of a range of possible positions (e.g., epistemic and evidential modality, rhetorical or expository questions), while Attribute represents propositions grounded in the subjectivity of an external voice which invokes dialogic alternatives. Those may either acknowledge the external voice or distance themselves from it, depending on the choice of framer.

Finally, according to Martin and White (2005: 135–201), Graduation operates across two axes of scalability—that of grading according to intensity or amount (Force), and that of grading according to prototypicality (Focus). Force applies most typically to the scaling of qualities and processes, which may be raised through the use of Emphasizers (e.g., *a very long meeting*) or lowered through the use of Downtoners (e.g., *somewhat upset*). Focus typically refers to categories which, when viewed from an experiential perspective, are not scalable. Under Focus, it is possible to up-scale, or Sharpen (e.g., *a real thing*), or to downscale, or Soften (e.g., *they are kind of nice*). In this study, we draw on all three Appraisal systems (Attitude, Engagement, Graduation), and a selection of the subcategories (see Section 'Annotation procedures' for details), to achieve a comprehensive understanding of the evaluative resources in fake news.

Data and Methods

The Corpora

The data used in this study are based on two corpora: the single-authored corpus and the multi-authored corpus (see Põldvere et al. 2023). While the single-authored corpus contains news by one author only, namely, Jayson Blair, the multi-authored corpus contains news by several authors. The latter was modelled on the basis of the former to achieve a larger dataset where both the fake and genuine news articles have been produced by the same authors. This means that the corpora are controlled for well-known confounding variables in fake news research such as register, genre, authorship, etc. (Grieve & Woodfield, 2023). Another important similarity between the corpora is that, as far as we can see, the fake news articles in the corpora are instances of disinformation. This is certainly the case in the single-authored corpus in which case a large-scale investigation was launched by the newspaper where the articles were published, followed by the author's own admissions of guilt in his autobiography (Blair, 2004). The extent to which deception played a role in the multi-authored corpus is less clear; however, there are two main reasons why we believe that it did. Firstly, the articles in the corpus originate from news outlets that are partisan in nature and that tend to disseminate ideological messages to legitimize the actions of their party, and to delegitimize those of others—a fertile breeding ground for at least some level of bias and deception. Secondly, none of the articles contains a correction, which is common when the author has made an inadvertent mistake. Therefore, the two corpora share important similarities with each other, which explains why we have considered them together in this study. In what follows, we present a few differences, along with a more detailed description of each corpus.

The single-authored corpus based on news by Jayson Blair is the larger of the two corpora, with 55,785 words. It contains 36 fake news articles and 28 genuine news articles, 34,741 and 21,044 words, respectively. Blair was caught fabricating news for *The New York Times* (*NYT*) in the early 2000s. While *NYT*'s and Blair's accounts of why he lied differed, it was clear at the time that it was due to personal issues rather than to advance an ideological agenda (see Grieve & Woodfield, 2023 for details). It turned out that Blair rarely left New York City, including to cover one of his most famous cases, the D.C. sniper attacks of 2002, and the stories of parents whose sons and daughters had gone missing, or were reported dead, in the Iraq War. We used LexisNexis, a large archive of newspapers and periodicals, to collect the articles from 25 October 2002 to 29 April 2003, the period covered by *NYT*'s investigation. The fake news articles were identified based on the presence of corrections by the newspaper.[2]

The multi-authored corpus contains news by seven other authors, with 13,391 words for the whole corpus: 8,185 words and 19 texts of fake news and 5,206 words and 10 texts of genuine news. Therefore, the corpus is considerably smaller than

---

[2] A breakdown of each text in the single-authored corpus is available online at https://osf.io/en3f5/?view_only=a37279fa4fab400ab5721278a42e2c5f

the single-authored one, largely due to the difficulty of finding authors who have produced both fake and genuine news. However, the corpus complements well the more focused design of the single-authored corpus by including more authors. The authors were identified based on existing fake news datasets such as MisInfoText (Asr & Taboada, 2019) and the PolitiFact–Oslo Corpus (Põldvere et al. 2023). Both corpora rely on well-known fact-checking websites for their data, including Politi-Fact.com, a non-profit service operated in the USA. Both MisInfoText and the Politi-Fact–Oslo Corpus give access to the full texts of the news items, comprehensive metadata about them, and the veracity labels assigned by the fact-checkers. To build a corpus based on criteria in the single-authored corpus, we followed strict guidelines: (i) the authors had produced both fake and genuine news, regardless of quantity, (ii) in situations of co-authorship, the author in question was listed first, (iii) the veracity label was either (mostly) true or (mostly) false, with no indeterminate labels (e.g., half true), and (iv) the fake and genuine news articles by the same journalist were published in the same online news outlet. The period covered by the corpus is 21 November 2011—24 April 2022. The news outlets vary from left-leaning, left-wing (e.g., *Mediaite*) to right-leaning, right-wing (e.g., *The Gateway Pundit*) websites operated in the USA. The major topics are COVID-19 and the 2016 and 2020 US presidential elections.[3]

Therefore, the single- and multi-authored corpora differ from each other along three factors: (i) register, (ii) genre, and (iii) the authors' motivation to lie. Firstly, while Blair's articles fit the bill for traditional journalism where facts are relayed in a straightforward way to inform the reader about current events, the other authors' articles are more similar to personal commentaries where current events are interpreted and written about through the lens of the authors' own thoughts and experiences (compare the headlines *Sniper Suspects Linked To Yet Another Shooting* vs. *My Vote for Ann Patterson*, respectively). Secondly, the corpora differ from each other in terms of genre, that is, the textual and linguistic conventions of the news articles. Specifically, the procedural competence needed to produce a traditional newspaper article is quite different from that of news articles written for online websites, with the latter following a less clear structure. Finally, there is a difference in the specific motivational goals of the authors. Since the articles in the multi-authored corpus originate from news outlets that are partisan in nature, they tend to promote viewpoints and political ideologies that align with their own, while portraying negatively those that do not. This is illustrated by the following short headline in the far-right news website *The Gateway Pundit*: 'A STOLEN ELECTION' (in reference to President Biden's win during the 2020 US presidential election). The ideological motivation of these authors is, of course, very different from Blair's personal agenda to improve his own reputation and material well-being, for example.

We view the three factors above as being closely related to each other and as potentially having a combined influence on the use of Appraisal expressions in the data, in addition to the difference between fake and genuine news. For instance, the

---

[3] See https://osf.io/en3f5/?view_only=a37279fa4fab400ab5721278a42e2c5f for a breakdown of each text in the multi-authored corpus.
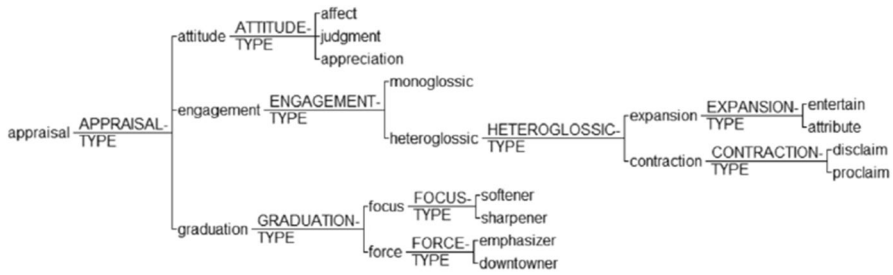
Fig. 3 Annotation scheme of Appraisal

ideological desire to deceive and persuade readers to adopt a certain worldview is clearly facilitated by the largely undefined conventions associated with online news websites. It is the combination of these factors that makes up a hugely influential news source that people today turn to for critical information. The same applies to traditional newspaper writing but with its own set of characteristic features. Therefore, in this study we compare the distribution of Appraisal expressions both within and between the single- and multi-authored corpora to shed new light on the linguistic choices that people make when they lie or tell the truth under different sets of circumstances.

Annotation Procedures

The annotation of the Appraisal expressions involved two tasks, firstly, the identification of the expressions and, secondly, their classification into Appraisal (sub)categories. For these two tasks, we used specialized software for qualitative and quantitative data analysis: UAM CorpusTool (O'Donnell, 2016) and MAXQDA (VERBI Software, 2023). The annotation tasks were divided up equally between the first and the second author, who worked separately on the corpus texts assigned to them. In the first stage of the annotation process, we restricted the segments to explicitly evaluative lexical items, leaving aside the constituents of the syntactic units as well as implicit or invoked items.[4] In (3), for example, only the adjective *important* was annotated.

(3) Among the important pieces of evidence in the case are police records that track law enforcement encounters with the 1990 Chevrolet Caprice in the months of sniper shootings and during them.

The focus was on the body texts of the news articles, not the headlines, in line with Grieve and Woodfield (2023).

The classification task involved assigning each segment in the corpora an Appraisal label, based on the three main systems—Attitude, Engagement,

---

[4] Within Appraisal, implicit or invoked items are 'seemingly neutral wordings that imply or invite a positive or negative evaluation' (Fuoli, 2018: 235) rather than expressing evaluation explicitly (i.e., inscribed evaluation); they are notoriously difficult to identify reliably in discourse.

Graduation—and their subcategories. Figure 3 shows the Appraisal (sub)categories that were included in our analysis. Note that, within Engagement, we only annotated for heteroglossic voices and viewpoints, rather than monoglossic, and that the four subcategories (Entertain, Attribute, Disclaim, Proclaim) were considered in their entirety without further categorization (e.g., denial vs. counter-expectation within Disclaim). Such details are accounted for in our qualitative analysis in Section 'Results and discussion'.

Since annotating Appraisal is a complex and highly subjective task, we created a short document with the main annotation decisions to ensure that the authors followed the same guidelines. For example, in the identification task we decided to annotate separately words that belong to the same Appraisal system and are positioned next to each other. Thus, combinations of Entertain and Disclaim (e.g., *couldn't*), Entertain and Attribute (e.g., *seemed to suggest*), and Entertain and Proclaim (e.g., *the evidence would show*) were marked as two different segments since they refer to different aspects of Engagement. Also, we decided not to include capitalized letters as a criterion of Force (e.g., 'A STOLEN ELECTION'), although we admit that it would be an interesting feature to explore in the future.

Following AUTHOR 1 and co-author (2023), we classified conditionals as instances of Engagement, even though originally they were not part of Appraisal. This is because conditionals point to some possibility or irrealis situation that corresponds to the pragmatic space of the Entertain subcategory. Reporting verbs which rule out alternative positions (*declare, announce*) were classified as instances of Proclaim, while verbs that entertain the possibility of dialogic alternatives (*say, speak, describe*) were given the label Attribute. We also included large numbers which determine noun phrases (e.g., *thousands of families*) in the Graduation subcategory of Force, as Emphasizers, while the notions of Proximity/Distribution of time and space (e.g., *recent, distant*) were not incorporated into our scheme since they do not directly refer to intensity or amount.

To assess the reliability and replicability of the guidelines, we carried out a series of inter-rater reliability tests based on ~10 per cent of the data. The tests were followed by discussion sessions where disagreements between the authors were discussed and resolved together. The discussion sessions served to refine the annotation guidelines and to revise the rest of the annotations accordingly. For both the identification and classification tasks, we report the agreement scores obtained prior to reconciliation. In the first instance, we used *precision* (PRE), *recall* (REC) and *F measure* (F-score) scores. According to Fuoli and Hommerberg (2015), these scores are more appropriate for the identification of what to mark than *kappa* scores (Cohen, 1960), which are more suited for the labelling of units of fixed length (see below). Taking the annotations of the second author as the gold standard (i.e., the 'correct' annotations), (i) PRE indicates the units identified by the first author that are relevant or 'correct', (ii) REC indicates the units that have been successfully identified by the first author, and (iii) F-score provides a synthetic measure of PRE and REC (see Fuoli & Hommerberg, 2015 for details on the calculations). The inter-coder agreement for the identification task was moderate, with an overall mean F-score of 0.61. However, during reconciliation many of the problems that we had encountered were resolved to ensure greater reliability and accuracy in the rest of the corpus texts. We

Table 1 Distribution of main Appraisal systems across fake and genuine news in the single-authored corpus (normalised frequencies per 10,000 words with raw frequencies in parentheses; LL = Log Likelihood, LR = Log Ratio)

	Fake news	Genuine news	LL	$p < .05$?	LR
Attitude	806 (2799)	757 (1593)	3.97	Yes	0.09
Engagement	639 (2221)	641 (1348)	0.00	No	−0.00
Graduation	212 (735)	149 (313)	28.46	Yes	0.51

used chance-corrected *kappa* scores for the classification task, along with observed agreement. The scores were calculated based on the subcategories, so on a relatively high level of granularity. Despite this, they were very high, with 90.9% observed agreement and an 'almost perfect' level of agreement ($k = 0.885$), according to the *kappa* scale of Landis and Koch (1977).

The quantitative, statistical analysis of the annotated data was conducted in RStudio (RStudio Team, 2022) to generate the plots (package *ggplot2*) and via the UCREL log-likelihood and effect size calculator (https://www.socscistatistics.com/tests/chisquare/) to determine statistical significance (Log-likelihood) and effect size (Log Ratio). The significance level was set to 5% (critical value = 3.84)

Results and Discussion

The Single-Authored Corpus: Jayson Blair

The single-authored corpus featuring news articles by Jayson Blair contains 9009 Appraisal expressions. Between the fake and genuine news samples, there is a statistically significant difference in the use of these expressions (Log-likelihood = 9.92, Log Ratio = 0.10), with the fake news sample (1657 per 10,000 words; 5755 occurrences) containing relatively more instances than the genuine news sample (1546 per 10,000 words; 3254 occurrences). Table 1 shows further differences in the distribution of the main Appraisal systems.

As can be seen in Table 1, both Attitude and Graduation are more common in Blair's fake news compared to his genuine news, while Engagement did not reach statistical significance. Blair's tendency to make greater use of Attitude expressions in his fake news may be linked to the circumstances around his deception. Considering that Blair lied about being on site and had to concoct scenes to keep his job, it is possible that the Attitude expressions were used to compensate for his lack of access to basic information about the news events (i.e., the *who*, *what*, *when* and *where* of journalism). The greater use of Graduation expressions will be easier to understand when we consider the subcategories separately, since Graduation contains a variety of meanings that often occupy opposite ends of the scales of intensification and prototypicality. Next, we zoom in on the distribution of the subcategories to further unpack Blair's linguistic strategies.

Figure 4 illustrates the distribution of the Attitude subcategories.

Fig. 4 Distribution of Attitude subcategories
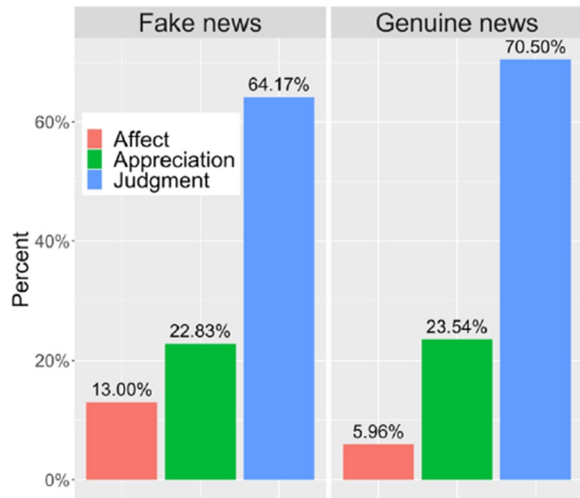
In both types of news, the most common way to express attitude is through Judgment, possibly to characterize the behaviour of many of the people that Blair wrote about, followed by Appreciation and then Affect. However, there are differences in the extent to which this is the case. The subcategory that shows a statistically significant difference between fake and genuine news is Affect (Log-Likelihood = 61.90, Log Ratio = 1.21), which occurs more frequently in the fake news sample (105 per 10,000 words; 364 occurrences) than in the genuine news sample (45 per 10,000 words; 95 occurrences). The rest of the subcategories did not reach significance.[5] An example of Affect from the fake news sample is given in (4), by means of the adjectives *bored* and *crazy*, underlined, to describe the state of mind of Lee Malvo, a teenager charged in the D.C. sniper attacks.

(4) In Mr. Malvo's first few weeks in jail, he was so <u>bored</u>, Mr. Petit said, that he tore up a business card and used it to play checkers with himself. "He had nothing else for entertainment other than to count the cinder blocks on the wall," Mr. Petit said. "That kind of isolation can drive anyone <u>crazy</u>."

For comparison, example (5) is from the genuine news sample. It describes the behaviour of Lee Malvo during his trial, without revealing the defendant's real or attributed emotions.

(5) At one moment, Mr. Malvo did not break his pose as a sheriff's deputy used scissors to cut open a box containing the Bushmaster rifle. As Mr. Franklin stared at the weapon that investigators say was used to kill his wife, Mr. Malvo briefly looked at the ground and then returned his attention to the proceedings.

The greater use of Affect expressions in Blair's fake news may be explained by the author's dramatization of the scenes that he wrote about, and his attempts to

---

[5] Complete results are available online at https://osf.io/en3f5/?view_only=a37279fa4fab400ab5721278a42e2c5f

Table 2 Sources of Affect (raw frequencies with percentages in parentheses)

	Fake news	Genuine news
Authorial	121 (42%)	23 (23%)
Non-authorial	170 (58%)	75 (77%)

Fig. 5 Distribution of Engagement subcategories
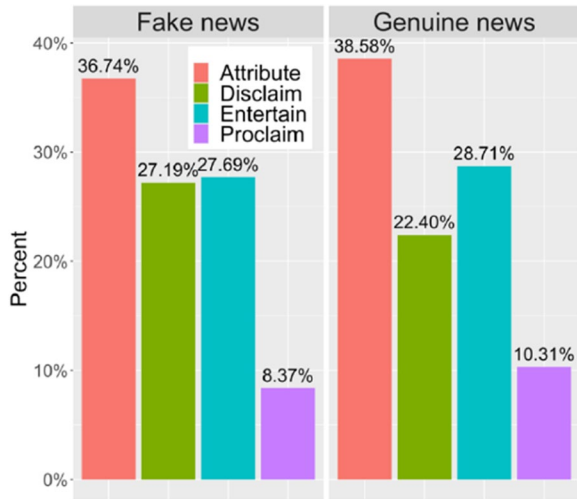
compensate for the lack of basic information with details about emotions that characterize scene participants. Additionally, emotional language in Blair's fake news is possibly there to influence the affective side of readers on the topics of crime, war and race relations in the USA at the time. To find out the sources of the affective states, we carried out additional analysis of the data based on a distinction between authorial (i.e., Blair's own evaluations) and non-authorial (i.e., those expressed by third parties) sources; see Table 2.

Table 2 shows the proportions of authorial and non-authorial sources of Affect across fake and genuine news. In both cases, non-authorial sources are more common (see (4) above). However, the proportions are almost equal in Blair's fake news (42% vs. 58%), which suggests an alleged presence of the author during the events and creates an impression of vividness in the stories. Example (6), taken from a news story about a woman whose son had gone missing in war, illustrates an authorial source.

(6) She finds herself jolted out of the numbness at moments, realizing that she has been staring blankly at a framed photograph of her eldest son, Cpl. Michael Gardner II, a Marine scout in southern Iraq.

This contributes further to Blair's reliance on his own assessment of what happened rather than what was relayed to him by actual scene participants, an inevitable outcome of false narratives.

Figure 5 shows the distribution of the Engagement system. As can be seen in the figure, the distribution of the subcategories is roughly the same across Blair's

Fig. 6 Distribution of Graduation subcategories
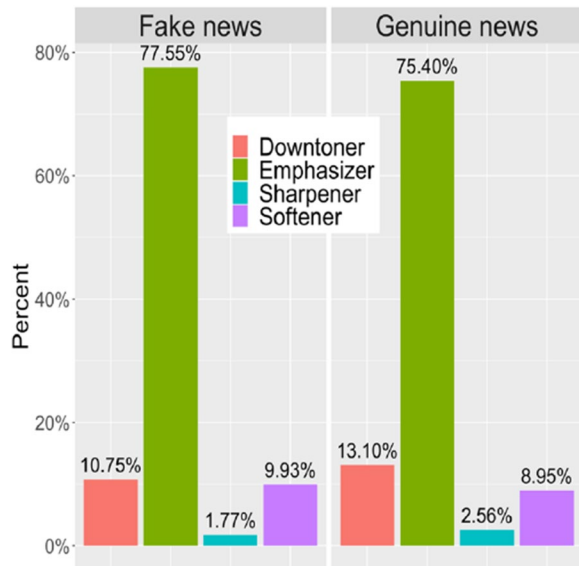
fake and genuine news, with Attribute being the most frequent one, followed by Entertain, Disclaim and Proclaim, in that order. The prevalence of markers of Attribute to quote or report is expected in newspaper writing. The only subcategory that shows significant difference across the fake and genuine news samples is Disclaim (Log-Likelihood = 7.56, Log Ratio = 0.28), which is more frequent in Blair's fake news (174 per 10,000 words; 604 occurrences) than in his genuine news (144 per 10,000 words; 302 occurrences).

The function of Disclaim resources in discourse is to restrict the scope of dialogically alternative positions, making it difficult to question the author's perspective. This is illustrated by several markers of denial and counter-expectation in fake news in (7), where Blair describes the legal situation of Lee Malvo and his accomplice John Muhammad. Collectively, the markers serve to project on the reader particular beliefs and expectations of the legal system, both by the legal experts and by extension Blair himself. The use of the conjunction *but* also seems to strengthen the understanding that the opinion of legal experts was indeed sought out by Blair.

(7) The decision to try Mr. Muhammad and Mr. Malvo in separate Virginia counties was almost certainly driven by a desire to get around the quirk in that state's law that contemplates the death penalty in serial murder cases only where the defendant was the gunman and not an accomplice, legal experts said. But legal experts said today that Mr. Malvo's admissions to Mr. Meyers's killing could be used in Mr. Muhammad's trial to bolster arguments that he did not fire the fatal bullet.

In fake news, the use of denial and counter-expectation markers such as the ones in (7) and *only* in (1) in Section 'Introduction' above may be useful for closing down the communicative space for dialogic alternatives from readers who might otherwise start to question the information provided. Even though alternative positions are

Table 3 Distribution of main Appraisal systems across fake and genuine news in the multi-authored corpus (normalised frequencies per 10,000 words with raw frequencies in parentheses; LL = Log Likelihood, LR = Log Ratio)

	Fake news	Genuine news	LL	$p < .05$?	LR
Attitude	936 (766)	920 (479)	0.09	No	0.02
Engagement	617 (505)	442 (230)	18.31	Yes	0.48
Graduation	272 (223)	246 (128)	0.86	No	0.15

recognized, they are nevertheless held not to apply, a powerful tactic for convincing the readers of the validity of one's arguments (Martin & White, 2005: 118–119).

Figure 6 illustrates the distribution of the subcategories of Graduation.

Of these, Emphasizer is by far the most common choice in both fake and genuine news, with the rest being relatively uncommon in both types of news (Downtoner, Softener, Shapener, in that order). The prevalence of Emphasizers in Blair's news is somewhat surprising considering that generally they are associated with more informal and interactive contexts (Biber, 1988). But similar to Grieve and Woodfield (2023) who found emphatics to be used at a considerably higher rate in Blair's fake news, we observed a significant difference in the use of Emphasizers (Log-Likelihood = 25.35, Log Ratio = 0.55), with more instances in fake news (164 per 10,000 words; 570 occurrences) than in genuine news (112 per 10,000 words; 236 occurrences). This result also provides support for Rashkin et al.'s (2017) finding that words used to exaggerate, such as subjective words, superlatives and modal adverbs, are typically used in fake news sources compared to trusted news sources. In Blair's case, what is often being exaggerated is the affective state of scene participants, as conveyed by expressions of Affect, which together with Emphasizers of different kinds are used to convey inconsequential information when basic information about the events and situations is missing. Example (8) illustrates the use of the superlative *most* (see also (1) in Section 'Introduction' for *fully*).

(8) The officers say the <u>most</u> emotional moments come when they find themselves in the position of surrogate parents, helping children with homework, or playing with dolls or throwing a baseball with them.

Finally, there is a statistically significant difference in the use of Softeners in Blair's news (Log-Likelihood = 4.49, Log Ratio = 0.66): they occur with greater frequency in his fake news (21 per 10,000 words; 73 occurrences) than in his genuine news (13 per 10,000 words; 28 occurrences). This suggests a relationship between telling a lie and having to be vague about details, although the low frequencies in our case make it difficult to draw definite conclusions at this point.

The Multi-Authored Corpus: Other Authors

The multi-authored corpus featuring news articles by seven other authors contains 2331 Appraisal expressions. As in the single-authored corpus, Appraisal is significantly more common (Log-Likelihood = 8.74, Log Ratio = 0.18) in fake news (1825 per 10,000 words; 1494 occurrences) compared to genuine news (1608 per 10,000

Table 4 Sources of Attribute (raw frequencies with percentages in parentheses)

	Fake news	Genuine news
Specified	51 (69%)	53 (81%)
Non-specified	23 (31%)	12 (19%)

words; 837 occurrences). Table 3 illustrates the distribution of the main Appraisal systems.

As can be seen in the table, Engagement is the only system that shows a significant difference between the fake and genuine news samples, while the differences between the two types of news with respect to Attitude and Graduation did not reach significance. This is the opposite to the single-authored corpus which showed a greater use of Attitude and Graduation but not Engagement (see Section 'Comparing the corpora' for further comparisons).

The greater use of Engagement resources in fake news in the multi-authored corpus reflects one of the main communicative functions of the ideology-driven nature of this type of news: to convince the reader to adopt the author's opinion by means of linguistic expressions of the author's commitment to the truth of the described events and situations. Based on qualitative observations of the corpus texts, the two Engagement subcategories that stood out in this regard are Disclaim and Entertain. On the one hand, markers of Disclaim such as of denial (negation) and counter-expectation are often used to convince readers to adopt a negative opinion of their opponents and their competences. In (9), this is achieved through *not* (negative marker) and *just* (countering).

(9) Then I read the mail put out by Marie Corfields opponent in the 16th legislative District Assembly special election. Two negative pieces which did <u>not</u> address any differences in issues, <u>just</u> made negative and ridiculous points.

On the other hand, Entertain resources are used when authors try to be cautious about their responsibility for the propositions that they advance, presumably to avoid any legal repercussions that might arise. As illustrated in (10), this is often done through the use of evidential markers that express indirect evidence without attribution to a precise source, while at the same time toning down the author's commitment to the reported event.

(10) The union is <u>reportedly</u> not allowing drivers to remove goods from the Port of San Juan.

Another prominent type of Entertain resources is found among obligation markers, suggesting that authors prescribe or direct the way readers should think, in accordance with the ideology of the author. This is illustrated in (11) where the author gives an authorial directive in the kind of attitude that they want readers to share about the US senator Ted Cruz. The sentence in (11) follows a false statement that the author made about Cruz banning dildos as Texas solicitor general.

(11) Thus it <u>should</u> come as no surprise that as Texas solicitor general, a post he held from 2003 to 2008, Cruz defended Conservative Christian causes like the inclusion of under God in the pledge of Allegiance, the display of the Ten Commandments on the grounds of the state capitol, and a ban on late-term abortions.

Table 5 Distribution of main Appraisal systems across fake and genuine news for each author in the multi-authored corpus (italics are used to indicate which news register is more common within each system; normalised frequencies per 100 words with raw frequencies in parentheses)

	Attitude		Engagement		Graduation	
	Fake	Genuine	Fake	Genuine	Fake	Genuine
Author 1	*9 (26)*	5 (31)	5 (14)	5 (31)	1 (2)	1 (5)
Author 2	11 (107)	*14 (141)*	*4 (36)*	3 (26)	*4 (35)*	3 (28)
Author 3	6 (7)	*12 (41)*	9 (10)	2 (8)	*4 (5)*	1 (4)
Author 4	*13 (60)*	8 (115)	*6 (27)*	5 (77)	1 (3)	*2 (30)*
Author 5	*12 (219)*	5 (34)	*8 (148)*	4 (31)	*3 (51)*	2 (11)
Author 6	8 (82)	8 (74)	*7 (75)*	3 (29)	2 (22)	*4 (39)*
Author 7	8 (265)	*12 (43)*	6 (195)	*8 (28)*	3 (105)	3 (11)

Since Engagement resources seem to be a prominent feature of fake news, we carried out an additional analysis of how markers of Attribute manifest themselves in the multi-authored corpus, following Grieve and Woodfield's (2003) qualitative observations of the single-authored corpus. Specifically, we were interested in the sources of Attribute, which may be either specified (i.e., concrete references) or non-specified (vague references, e.g., *some people*, *legal experts*, *investigators*). Table 4 shows the outcome of that analysis.

The sources in both fake and genuine news are mainly specified, meaning that authors tend to attribute propositions to people or entities that, if needed, could be validated. However, it is also the case that the fake news sample contains a greater proportion of non-specified sources (31% vs. 19%), which make it easier for authors to attribute false propositions to people without being found out, while at the same time lending credibility to the message that they are advancing. The same observations were made by Grieve and Woodfield (2023). As noted by Rashkin et al. (2017: 2933) who observed a greater use of hedges in fake news sources, these results are in line with psychology theories whereby deceivers show more 'uncertainty and vagueness' and 'indirect forms of expression'.

Below, we briefly observe the individual differences between the seven authors. Table 5 shows the distribution of the main Appraisal systems separately for each author. As can be seen in the table, for the most part the results reflect those for the whole corpus: Engagement is clearly more common in fake news, while Attitude and Graduation are more evenly distributed across fake and genuine news.

However, there is one author whose fake news stories are characterized by all three systems, namely, Author 5. This author uses Appraisal extensively to convey highly subjective, opinionated, and ideologically-charged stance. This is illustrated in (12) where various evaluative resources come together to construct a false narrative of US–China agreements on cybersecurity.

(12) As usual, government is holding itself to a different, lower standard. That is unacceptable. We must send a clear signal to our adversaries that repeated, aggressive hacking will not be tolerated. That starts with the development of a serious strategy. But it also means that we have to be willing to push back. This President

has <u>not</u> done that. When we learned that China had <u>hacked</u> into our systems, President Obama had a state visit. When we held recent economic dialogues with China, we agreed on <u>over</u> 100 different things including wildlife <u>trafficking</u> and volcano research. <u>None</u> of these 100-<u>plus</u> points of agreement addressed cybersecurity. China <u>sees</u> this and <u>knows</u> that they <u>can push</u> <u>ahead</u> with an <u>aggressive</u> economic and military agenda because they are <u>not</u> being <u>challenged</u>.

These resources range, among other things, from negative judgments of behavior (the idea of an incompetent US government and an aggressive China), rejections of alternative viewpoints via negative markers (e.g., *none of these 100-plus points of agreement addressed cybersecurity*) to resources for intensification to further drive home the message about the US government's oversight about cybersecurity (*over 100 different things, 100-plus points*). Interestingly, Author 7 shows a different trend, where it is the genuine news stories that are more attitudinal and engaged with dialogic alternatives. It is therefore clear that there is variability in how fake news is expressed linguistically, both across individual authors and contexts of use. The next subsection elaborates on the latter.

Comparing the Corpora

The main similarity between the single- and multi-authored corpora is their greater use of evaluative resources in fake news, a finding which has been pointed out previously (e.g., Conroy et al., 2015; Hancock et al., 2011; Sousa-Silva, 2022). However, through a detailed analysis of the distribution of Appraisal expressions in the corpus texts, we have been able to provide a more comprehensive description of evaluation in deceptive discourse. We showed that, while Jayson Blair in the single-authored corpus made greater use of Attitude and Graduation expressions in fake news, the same type of news by seven other authors in the multi-authored corpus was instead characterized by Engagement expressions. We tentatively suggest that this is due to the combination of contextual factors such as register, genre and the motivation of the authors to deceive readers for personal reasons in the single-authored corpus and for ideological reasons in the multi-authored corpus. On the one hand, then, the function of the evaluative resources in the single-authored corpus is to evoke a sense of vividness in the stories to compensate for the lack of access to basic information that is required by the register of news reporting in traditional journalism. On the other hand, the evaluative resources in the multi-authored corpus, which contains personal commentaries on online news websites, is to convince readers to adopt the author's ideology. However, it is important to note that we do not know which factor (genre, register, motivation) contributes most strongly to these differences, which needs to be determined in future research.

These linguistic differences between the corpora do not come as a surprise to those who subscribe to functionalist approaches to language where language use in general is regarded as highly contextualized, but surprisingly it is a view that has not been taken seriously in much of previous research on the language of fake news within NLP. This is evidenced by the tendency of such research to conflate different registers, genres and contexts of use in search of a core set of features of fake

news. These features are then fed into automatic detection systems which are meant to identify fake news in its general sense (false information, intent to deceive) but whose performance may be negatively affected by a range of other factors. By showing that evaluation does not behave in the same way across registers, genres and motivational goals, we call on fake news researchers to take into consideration these factors in future research, for example, by developing different detection systems for different purposes (see Põldvere et al. 2023 for early attempts). Corpus linguists would make a valuable contribution to these efforts due to their knowledge and training both in linguistic theory and quantitative, statistical methods.

Despite these differences at the higher level of Appraisal systems, we identified two specific features that were the same across the two corpora and that therefore may be more resistant to contextual factors. Firstly, in fake news authors tend to tone down their commitment to and responsibility for the truth of the propositions that they advance. This can be observed through markers of Entertain such as evidentials and obligation markers in the multi-authored corpus, and non-specified sources of Attribute in both corpora (cf. Grieve & Woodfield, 2023). The author may want to be vague about describing events that did not happen or citing sources that did not exist, for fear of being found out and/or taken legal action against. Secondly, based on their prevalence in both the single- and multi-authored corpora, fake news is characterized by markers of Disclaim such as of denial and counter-expectation that function to close down the communicative space for dialogic alternatives, making it difficult for readers to question the writer's perspective. Such markers are used to convince the readers of the validity of one's arguments, an integral part of telling a lie. At first sight, then, the two features seem quite contradictory in terms of intersubjective functionality: the former is dialogically expansive, while the latter is contractive. However, it is possible that the resources become relevant at different points in discourse and therefore they complement each other well when authors navigate the complex web of deception.

Conclusion

This study had two aims: one primary and the other secondary. Firstly, it has investigated the distribution of Appraisal expressions in two corpora—Jayson Blair's single-authored corpus and a multi-authored corpus—with the aim to shed new light on how and why, if at all, evaluative meanings are construed differently in fake news compared to genuine news. The findings have provided evidence in support of previous research that the greater use of evaluative expressions may indicate "fakeness" of news. By drawing on Appraisal Theory, a powerful model of evaluation in linguistics, we have been able to show in a systematic way exactly what types of evaluative resources are instances of fake news, based on qualitative and quantitative evidence. Dialogically, they are both expansive and contractive, and they tend to be vague with regards to the source of information. Secondly, we have explored the idea that evaluative language use in fake news may be influenced by contextual factors such as the combination of register, genre and the motivational goals of the authors, by using two very different corpora. The fact that there were noticeable

differences in the Appraisal systems across the corpora (Attitude and Graduation in the single-authored corpus and Engagement in the multi-authored corpus) suggests to us that people lie in different ways depending on their circumstances.

These results have important implications for fake news research both in linguistics and NLP: rather than looking for a core set of linguistic features that characterize fake news as a whole, we should instead shift our focus to the many ways in which fake news manifests itself to develop more effective detection systems. For corpus linguists, in particular, there are great opportunities to contribute with robust data collection procedures based on which detailed investigations of the role of contextual factors can be carried out. In this study, we have presented two corpora that, albeit relatively small and limited in topical scope, have been carefully designed and controlled for well-known confounding variables in fake news research. The corpora have already served as an inspiration for larger corpus compilation projects across different languages, registers, genres and contexts of use (e.g., Põldvere, Kibisova & Alvestad, 2023, Põldvere et al. 2023).

Author contributions All authors contributed to the study conception and design. Data collection and analysis were performed by both authors, R.T. and N.P. The first draft of the manuscript was written by both authors and both authors commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

Data availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval No ethical approval is required. This is an observational study.

Consent for publishing The authors give the permission to *Corpus Pragmatics* to publish the article *Investigating Appraisal and the language of evaluation in fake news corpora*.

References

Asr, F. T., & Taboada, M. (2019). Big Data and quality data for fake news and misinformation detection. *Big Data & Society, 6*(1), 1–14. https://doi.org/10.1177/2053951719843310

Bakhtin, M. (1981). Discourse in the novel. In M. Holquist (Ed.), *The Dialogic Imagination: Four Essays by M. M. Bakhtin* (pp. 259–422). University of Texas Press.

Bednarek, M. (2008). Emotion Talk across Corpora. *Palgrave Macmillan*. https://doi.org/10.1558/lhs.v3i3.399

Biber, D. (1988). Variation across Speech and Writing. *Cambridge UniversityPress*. https://doi.org/10.1017/CBO9780511621024

Biber, D., & Conrad, S. (2019). Register, Genre, and Style. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511814358

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*(1), 93–124. https://doi.org/10.1515/text.1.1989.9.1.93

Blair, J. (2004). *Burning down my Master's House: My Life at The New York Times*. New Millennium Press.

Carretero, M., Taboada M. (2014). Graduation within the scope of Attitude in English and Spanish consumer review of books and movies. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in Context* (pp. 221–239). John Benjamins. https://doi.org/10.1075/pbns.242.11car.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *ASIST Proceedings, 52*(1), 1–4. https://doi.org/10.1002/pra2.2015.145052010082

Fuoli, M. (2012). Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports. *Discourse & Communication, 6*(1), 55–81. https://doi.org/10.1177/1750481311427788

Fuoli, M. (2018). A stepwise method for annotating APPRAISAL. *Functions of Language, 25*(2), 229–258.

Fuoli, M., & Hommerberg, C. (2015). Optimising transparency, reliability and replicability: Annotation principles and i66666nter-coder agreement in the quantification of evaluative expressions. *Corpora, 10*(3), 315–349. https://doi.org/10.3366/cor.2015.0080

Grieve, J., & Woodfield, H. (2023). *The Language of Fake News* (Elements in Forensic Linguistics). *Cambridge University Press*. https://doi.org/10.1017/9781009349161

Hancock, J., Woodworth, M., & Porter, S. (2011). Hungry like a wolf: A word pattern analysis of the language of psychopaths. *Legal and Criminological Psychology, 18*(1), 102–114. https://doi.org/10.1111/j.2044-8333.2011.02025.x

Hood, S. (2010). *Appraising Research: Evaluation in Academic Writing*. Palgrave.

Kaltenbacher, M. (2006). Culture related linguistic differences in tourist websites: The emotive and the factual. A corpus analysis within the framework of Appraisal. In G. Thompson & Hunston, S. (Eds.), *System and Corpus: Exploring Connections* (pp. 269–292). Equinox. https://doi.org/10.1558/EQUINOX.19152.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Li, H., Dunn, J., & Nini, A. (2022). Register variation remains stable across 60 languages. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2021-0090

Martin, J. R., & White, P. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Martin, J. R. (2003). Introduction, Special issue on Appraisal. *Text, 23*(2), 171–181. https://doi.org/10.1515/text.2003.007

Ngo, T., & Unsworth, L. (2015). Reworking the appraisal framework in ESL research: refining attitude resources. *Functional Linguistics, 2*(1), 1–24. https://doi.org/10.1186/s40554-015-0013-x

Põldvere, N., Fuoli, M., & Paradis, C. (2016). A study of dialogic expansion and contraction in spoken discourse using corpus and experimental techniques. *Corporation, 11*(2), 191–225. https://doi.org/10.3366/cor.2016.0092

Põldvere, N., Kibisova, E., & Alvestad, S. S. (2023). Investigating the language of fake news across cultures. In S. M. Maci, M. Demata, M. McGlashan & P. Seargeant (Eds.), *The Routledge Handbook of Discourse and Disinformation* (pp. 153–165). Routledge. https://doi.org/10.4324/9781003224495-11

Põldvere, N., Uddin, Z., & Thomas, A. (2023). The PolitiFact-Oslo Corpus: A new dataset for fake news analysis and detection. *Information, 14*, Article 627. https://doi.org/10.3390/info14120627

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J. & Stein, B. (2018). A stylometric inquiry into hyper-partisan and fake news. *ACL Proceedings 1*, 231–240. https://doi.org/10.48550/arXiv.1702.05638

Rashkin, H., Choi, E., Jang, J. Y. , Volkova, S. & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. *EMNLP Proceedings*, 2931–2937.

RStudio Team. (2022). *RStudio: Integrated Development for R* (Version 2022.12.0+353). RStudio. https://rstudio.com

Shizhu, L., & Jinlong, H. (2004). Appraisal system in news discourse. *Technology. Enhanced Foreign Language Education, 4*, 17–21.

Sousa-Silva, R. (2022). Fighting the fake: A forensic linguistic analysis to fake news detection. *International Journal for the Semiotics of Law, 35*, 2409–2433. https://doi.org/10.1007/s11196-022-09901-w

Tandoc Jr, E. C., Thomas, R. J. & Bishop, V. (2021). What is (fake) news? Analyzing news values (and more) in fake stories. *Media and Communication 9* (1), 110–119. https://doi.org/10.17645/mac.v9i1.3331

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Trnavac, R., & Taboada, M. (2023). Engagement and constructiveness in online news comments in English and Russian. *Text & Talk, 43*(2), 235–262. https://doi.org/10.1515/text-2020-0171

VERBI Software. (2023). *MAXQDA 2020*. VERBY Software. https://www.maxqda.com.

Volkova, S., Shaffer, K., Jang, J. Y. & Hodas, N. (2017). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. *ACL Proceedings 2*, 647–653. https://doi.org/10.18653/v1/P17-2102.

Voloshinov, V. N. (1973). *Marxism and the Philosophy of Language* (L. Matejka and I. R. Titunik, Trans.). Harvard University Press. https://doi.org/10.1163/9789047408482.

White, P. R. R. (2012). Attitudinal meanings, translational commensurability and linguistic relativity. *Revista Canaria De Estudios Ingleses, 65*, 147–159.