LYING IN PRINT: THE LINGUISTIC PATTERNS OF DECEPTION IN THE FABRICATED JOURNALISM OF STEPHEN GLASS

JUSTINE SMITH
UNIVERSITY OF SUSSEX

ABSTRACT

Public trust in journalism is waning, yet research on disinformation has focused predominantly on non-institutional online sources such as social media or partisan websites. Far less is known about how deception can emerge within mainstream newswriting that appears to conform to professional journalistic norms. This study addresses that gap by investigating how fabricated journalism differs from verified reporting in its patterning of stance and evaluation using an Appraisal-informed corpus-assisted discourse studies (CADS) approach. The analysis draws on a purpose-built, matched-author corpus of 23 articles (12 fabricated, 11 verified) written by Stephen Glass for *The New Republic*. Frequency profiling, collocation, and concordance analysis reveal systematic contrasts between fabricated and verified journalism. Fabricated texts display lower lexical density, heavier use of verbs and pronouns, and a more personalised narrative style. Evaluative items such as *most*, *just*, and *not* occur more frequently and function to intensify claims, limit alternative interpretations, and manage authorial commitment. Collocational and attributional patterns further show that fabricated articles embed stance more frequently in the journalist's own voice, projecting confidence and sincerity while guiding reader alignment. By holding author, outlet, and register constant, the study ensures that any differences in stance and evaluative patterning are not attributable to genre or institutional context. Methodologically, it demonstrates how corpus tools can be integrated with discourse analysis to examine how deception is linguistically enacted through patterns of interpersonal meaning, with implications for media literacy and the detection of deceptive news.

CONTACT

Justine Smith, School of Media, Arts and Humanities, University of Sussex, Sussex House, Brighton, BN1 9RH, United Kingdom. js2374@sussex.ac.uk

ORCID

0009-0001-6329-7652

Lying in print: The linguistic patterns of deception in the fabricated journalism of Stephen Glass

Justine Smith

University of Sussex

1. Introduction

Public trust in journalism is declining as readers find it increasingly difficult to separate fact from fabrication (Fletcher et al., 2025; Newman, 2024; Savage, 2025). This erosion of confidence, together with falling news engagement (Reuters Institute, 2025), has intensified scrutiny of how language itself constructs credibility. While research on disinformation has largely focused on digital and social media contexts (e.g., Tandoc *et al.*, 2018; Torabi Asr & Taboada, 2019; Wardle, 2018), the linguistic signatures of fabrication within long-form journalism remain relatively under-explored. Examining how deception operates in written news offers a way to understand not only why readers are misled, but also how credibility is linguistically performed.

A revealing context for such analysis is the case of Stephen Glass, a former journalist for *The New Republic* whose fabrications were exposed in 1998. Glass authored both fabricated and verified articles for the same outlet, writing under a consistent editorial style and register. This rare control of author and genre enables direct linguistic comparison between fabricated and verified texts. Fabrication status was established using publicly documented findings from the 1998 internal and external investigations into Glass's reporting, with details provided in Section 3.2. The study draws on a purpose-built corpus of 23 articles (12 fabricated; 11 verified) to compare their linguistic patterns through frequency profiling, collocation, and concordance analysis.

The analysis adopts a corpus-assisted discourse studies (CADS) approach (Baker, 2023) to investigate whether fabricated journalism differs from verified reporting in its patterning of stance and evaluation, and if so, how. It examines variation in grammatical features (e.g., pronoun use, noun density), lexical items (e.g., intensifiers, negators, stance verbs), and their contextual deployment as evaluative resources. Register is treated as a fixed contextual frame, allowing differences in stance and evaluative patterning to be interpreted independently of genre or publication context.

By situating a historical case of deception within a consistent journalistic register, this paper extends linguistic research on disinformation (e.g., Jaworska, 2023, 2025; Sousa-Silva, 2022; Trnavac & Põldvere, 2024) to legacy print journalism. It examines how fabricated texts may simulate credibility through systematic patterns of stance and evaluation, realised through grammatical and lexical choices that project sincerity and authority. Using a matched-author corpus, the study examines how deception is enacted linguistically, highlighting implications for media literacy and corpus-assisted approaches to detecting deceptive news.

2.   Deception and disinformation in professional journalism

2.1.   From 'fake news' to Type II disinformation

The label 'fake news' has become so politically and rhetorically charged that it is often more effective at discrediting journalism than at describing text types (Lugea, 2021; Wardle, 2018). For analytical purposes, this study follows work that distinguishes *misinformation* (false or inaccurate information shared without intent to deceive) from *disinformation* (fabricated or manipulated content produced with the intention of misleading; Sousa-Silva, 2022; Tandoc *et al.*, 2018). Within Grieve and Woodfield's (2023) three-way distinction, Stephen Glass's articles fall clearly into Type II disinformation: fully fabricated news with deliberate intent to deceive.

Adopting a discourse-analytic perspective, the focus is therefore not on truth status per se, but on how deception is enacted linguistically. Specifically, how writers construct stance and evaluative meaning to simulate trustworthiness and journalistic authority. In this sense, Fairclough's (2010) observation that discourse is both socially shaped and socially shaping is relevant insofar as fabricated journalism must simultaneously conform to established institutional norms of newswriting while actively constructing a plausible version of social reality for readers. Fabricated reporting is thus treated here as strategic discourse, constrained by journalistic conventions yet exploiting them through patterned choices in stance and evaluation.

2.2.   Linguistic cues to deception in newsrooms

Prior work has attempted to distinguish fabricated from authentic news using computational approaches (e.g., Rashkin *et al.*, 2017; Torabi Asr & Taboada, 2019). However, such studies may blur the distinction between deception and register, meaning that classifiers capture patterns associated with genre differences rather than fabrication itself. (Lugea, 2021). In response, discourse-analytic research has called for tightly controlled comparisons that hold author, outlet, and register constant (Grieve & Woodfield, 2023).

2.3.   Evaluation, stance and Appraisal

Beyond lexis and syntax, deception in news discourse also operates through evaluation: the linguistic resources writers use to express commitment, authority, and alignment with readers. Evaluation concerns how strongly claims are presented, how certainty or doubt is signalled, and how readers are guided toward particular interpretations. Within this framework, stance refers to how writers position themselves in relation to claims and events, while engagement concerns how alternative viewpoints are acknowledged, attributed, or closed down (Hyland, 2005).

This study adopts an Appraisal-informed approach to analysing evaluation and stance. Developed within Systemic Functional Linguistics (Halliday & Matthiessen, 2004), the Appraisal Framework (Martin & White, 2005) provides tools for analysing how evaluative meaning is realised in discourse. While Appraisal distinguishes AFFECT,

GRADUATION, and ENGAGEMENT, the present analysis focuses primarily on GRADUATION and ENGAGEMENT, alongside patterns of attribution, as these are particularly relevant to journalistic writing and prior research on disinformation (Trnavac & Põldvere, 2024; White, 2012).

Hyland's (2005) model of stance and engagement likewise addresses how writers position themselves in relation to readers, but it differs in emphasis. Whereas Hyland's model is oriented toward rhetorical interaction, Appraisal offers more detailed tools for analysing how stance is built up through recurring lexical and grammatical choices. In this study, Appraisal is therefore used as an analytical lens, rather than as a competing theoretical framework.

Because Glass's fabricated and verified articles share an authorial voice, an Appraisal-informed analysis makes it possible to examine whether the same high-frequency lexical items perform different evaluative functions in context. Items such as *not*, *only*, and *most* are treated not as grammatical categories but as evaluative lexical resources, whose contribution to stance emerges through patterned use. For this reason, they are referred to descriptively as evaluative lexical items through which stance may be realised, reflecting their function in context rather than their part of speech.

2.4. Register control and matched-author design

One long-standing difficulty in deception research is that linguistic cues do not behave consistently across contexts. Some studies report fewer first-person pronouns in deceptive texts, others more; some find that liars produce shorter texts, others the opposite (Ali & Levine, 2008; DePaulo *et al.*, 2003). These inconsistencies suggest that apparent 'cues to deception' are often shaped by contextual factors such as genre, topic, or authorship rather than by deception itself.

As a result, there is growing agreement that linguistic patterns associated with deception become more interpretable when context, genre, and author are tightly controlled. Grieve and Woodfield's (2023) matched-author design offers a clear response to this problem. By comparing fabricated and verified articles written by the same journalist for the same outlet, their approach minimises genre- and publication-related variation, allowing differences associated with fabrication to be examined more directly.

The present study adopts the same approach. It follows a corpus-assisted discourse studies (CADS) approach (Baker, 2023), combining quantitative techniques such as frequency profiling and collocational analysis with qualitative, Appraisal-informed examination of concordance lines. Quantitative patterns are treated as prompts for closer discourse analysis, rather than as fixed or universal indicators of deception. Analytical decisions are guided by Brezina's (2018) principles of corpus transparency and interpretability, with particular attention to how linguistic features function in context.

The next section describes the corpus design and analytical procedures used to implement this approach.

3. Data and methods

3.1. Analytical stance: CADS and deception

The analysis adopts a corpus-assisted discourse studies (CADS) approach, in which quantitative findings are used to identify potentially meaningful patterns that are then examined through close qualitative analysis (Baker, 2023; Partington *et al.,* 2013). This approach is well suited to the study of disinformation, where relevant linguistic patterns are often subtle and emerge through systematic comparison and contextual interpretation.

In this study, frequency profiling and collocational analysis are used to guide concordance-based examination of stance and evaluative patterning in fabricated and verified reporting. Quantitative results are treated as indicators of where closer analysis is warranted, rather than as direct evidence of deception in themselves.

The analysis focuses on stance and evaluation as the primary dimensions through which fabricated and verified journalism are compared, interpreted using the Appraisal-informed framework outlined in Section 2.3. Evaluative lexical items are selected for closer examination because they recurrently contribute to the realisation of stance and evaluation in context.

As in much corpus-assisted discourse research, the methods employed here foreground differences between datasets while backgrounding areas of similarity (Taylor, 2013). The analysis therefore focuses on contrasting patterns of stance and evaluation without implying wholesale dissimilarity between fabricated and verified texts.

Finally, building on the view of disinformation as strategic and context-bound discourse (see Section 2), the study follows the controlled author-register approach proposed by Grieve and Woodfield (2023). By holding author, outlet, and register constant, it becomes possible to examine how patterns of stance and evaluation vary with fabrication status within a stable journalistic context, reducing the influence of genre- or source-related variation.

3.2. Corpus design and compilation

The corpus comprises articles written by Stephen Glass for *The New Republic* between 1995 and 1998, a rare and well-documented case of professional journalistic fabrication (Baranda, 2018; Reuters, 1998). *The New Republic* was selected because it represents a stable, high-prestige journalistic register comparable to US broadsheet and magazine journalism. This makes it suitable for comparison with existing research on fabricated texts, such as Grieve and Woodfield's (2023) analysis of Jayson Blair's work for *The New York Times.*

Unlike Blair's reporting, however, Glass's writing has not previously been examined from a linguistic perspective, despite its prominence in discussions of media ethics and professional misconduct (Ehrlich, 2005; Lasorsa & Dai, 2007; Spurlock, 2016). The cor-

pus therefore offers a valuable opportunity to analyse fabrication within a controlled, institutional journalistic context.

Glass published 41 by-lined feature articles during this period. To isolate cases of Type II disinformation (deliberate fabrication; see Section 2), articles were selected using the following criteria:

- Included: articles confirmed as fabricated, involving wholly or substantially invented people, events, or settings.
- Included: articles from the same period confirmed as verified texts.
- Excluded: cases involving plagiarism, minor factual error, or selective quotation only.
- Excluded: articles without reliable external confirmation of fabrication status.

This selection process resulted in a corpus of 23 articles: 12 fabricated and 11 verified.

Articles were retrieved from Nexis, cleaned to remove non-editorial material, and uploaded to Sketch Engine (Kilgarriff *et al.*, 2014) as two subcorpora (*Glass_Fabricated* and *Glass_Verified*). All computational analysis was conducted using Sketch Engine's tools. Article-level metadata (date, title, topic, and fabrication status) is provided in Appendix A.

The final corpus contains 55,786 tokens in total, with 41,114 tokens in the fabricated subcorpus and 14,672 tokens in the verified subcorpus. Sentence counts and unique word-form totals were generated in Sketch Engine and are presented in Table 1.

Subcorpus	Total tokens	Total sentences	Unique word forms
Glass_Fabricated	41,114	2,078	7,002
Glass_Verified	14,672	702	3,802

Table 1. Corpus statistics for Glass_Fabricated and Glass_Verified subcorpora

The imbalance reflects differences in article length: Glass's fabricated pieces were, on average, substantially longer than his verified articles. To ensure comparability across subcorpora, all frequency figures were normalised per 10,000 words. Normalisation is standard practice in corpus linguistics when comparing corpora of unequal size (Brezina, 2018). Because both subcorpora share author, outlet, and register, the design maintains a high degree of internal control despite differences in corpus size.

3.3. Analytical procedures

The analysis proceeded in three linked stages, (i) frequency profiling and feature selection, (ii) collocation and concordance analysis, and (iii) qualitative discourse analysis of evaluative patterns, with each stage informing the next.

3.3.1 Frequency profiling and feature selection

An initial frequency profile was generated for both subcorpora to identify systematic patterns of linguistic variation between fabricated and verified writing. Fourteen feature groups were selected on the basis of prior research on deception and journalistic style

(Grieve & Woodfield, 2023; Newman *et al.*, 2003; Sousa-Silva, 2022), alongside Appraisal-informed studies of evaluation (Martin & White, 2005; Trnavac & Põldvere, 2024). These comprised:

- Grammatical categories: nouns, verbs, adjectives, adverbs, conjunctions, overall pronouns, third-person pronouns, second-person pronouns, the copula *be*, and the pronoun *it*
- Evaluative lexical items whose recurrent use contributes to the realisation of stance (e.g., *but*, *not*, *only*), selected downtoners and emphatics (e.g., *just*, *most*, *very*), and a small set of stance-related markers (e.g., *say*, *seem*, *apparently*, *perhaps*)

The full list of evaluative lexical items included in these aggregated categories is provided in Appendix B.

Word lists were compiled and tested in Sketch Engine, with ambiguous items checked manually in concordance. All frequencies were normalised per 10,000 words. Log-likelihood (LL>3.84, *p*<0.05) was calculated using the UCREL wizard (Rayson, 2001) to identify features significantly over- or under-used in fabricated reporting (McEnery & Hardie, 2012). These frequency profiles are treated as diagnostic indicators, providing an empirical starting point for the qualitative analysis that follows.

3.3.2   Collocation and concordance analysis

For a subset of evaluative lexical items (*not, only, just, very, seem*, and the copula *be*), collocational patterns were extracted using Sketch Engine's Word Sketch with a +/-5-word span. These items were selected because they function as concrete resources for stance and evaluation, and because they are frequently associated with GRADUATION, ENGAGEMENT, and ATTRIBUTION in Appraisal-informed analyses.

To filter unstable associations, a minimum frequency threshold of two and the log-Dice statistic were applied, following Brezina *et al.* (2015). LogDice was chosen for its robustness across corpora of different sizes. These parameters were selected to balance salience and reliability in a medium-sized, specialised corpus.

Collocate sets were then examined through concordance analysis to assess how items functioned in context, for example whether *not* operated as simple negation or as dialogic denial, or whether *just* mitigated or intensified claims. A random sample of 30 concordance lines per item was selected for detailed coding. Speech-reporting verbs such as *say* were excluded because their strong association with quotation limits their relevance to authorial stance. The conjunction *but* was also excluded, as it functioned evaluatively in almost all cases across both subcorpora and therefore did not support contrastive analysis.

Each concordance line was coded for (i) evaluative status (evaluative vs. non-evaluative) and (ii) stance attribution (authorial, attributed, or vague), following the procedures outlined by Gillings and Mautner (2024). To assess reliability, a 20% random subsample of concordance lines was independently re-coded by the author at a later stage. Agreement exceeded 90% across evaluative and attributional categories, reflecting what Marchi (2024), drawing on Partington *et al.* (2013), describes as a 'culture of the counterexample'

(p. 301). All percentages reported in Section 4 are based on the full coded dataset; the re-coded subsample was used solely for reliability checking.

Coding outcomes were triangulated with frequency and collocational evidence to test for convergence. Collocational patterns then served as entry points for qualitative concordance analysis, linking statistical co-occurrence with the analysis of stance and evaluation presented in Section 4.

3.4. Scope and limitations

The corpus is limited to a single author, a single outlet, and a defined historical period. This high level of internal control supports close comparison of fabricated and verified journalism, but it also means that the findings should be interpreted as situated rather than universal. The analytical focus is restricted to stance and evaluation; other potentially informative dimensions, such as narrative organisation or reported speech framing, fall outside the scope of the present study but could be explored in future CADS replications.

4. Findings

4.1. Quantitative overview: Linguistic distributional patterns

This section reports the main quantitative findings comparing Stephen Glass's fabricated and verified articles. As described in Section 3.3.1, 14 linguistic features were selected on two grounds: their relevance in prior research on deception and stance-taking, and their reliable retrievability within Sketch Engine's search and extraction environment. These features include core grammatical categories (e.g., nouns, pronouns, conjunctions) and lexical items that recurrently realise stance through evaluation. Only the evaluative items were subsequently analysed using a reduced set of Appraisal-informed categories.

Table 2 presents raw frequencies, normalised percentages, and log-likelihood (LL) values for each feature. Together, these measures capture both grammatical variation (e.g., differences in nominal density) and lexical patterning (e.g., items that recurrently realise stance through evaluation). Similar distributional contrasts have been reported in earlier studies of deceptive discourse (Govaert *et al.*, 2020; Grieve & Woodfield, 2023; Rashkin *et al.*, 2017) and, in the context of fake news, in analyses of structural and lexical distinctiveness (Sousa-Silva, 2022). By profiling these features across matched-author corpora, the present analysis tests whether Glass's fabricated reporting exhibits comparable stylistic tendencies.

With the exception of nouns, all 14 features occur more frequently in the fabricated corpus. Nouns are notably less frequent in fabricated articles (24.86% vs. 26.80%) and yield the highest LL value (15.92), making this difference the most statistically salient. Overall, seven features exceed the conventional LL significance threshold (3.84, *p*<0.05), indicating systematic stylistic variation between Glass's fabricated and verified writing.

Linguistic feature	Fabricated (raw)	Fabricated (%)	Verified (raw)	Verified (%)	Log-likelihood
Nouns	10,219	↓24.86	3,932	↑26.80	**15.92**
Verbs	6,591	↑16.03	2,214	↓15.09	**6.13**
Adjectives	2,707	↑6.58	952	↓6.49	0.15
Pronouns (overall)	2,250	↑5.47	686	↓4.68	**13.39**
3rd person pronouns	1,674	↑4.07	503	↓3.43	**11.79**
Adverbs	2,077	↑5.05	697	↓4.75	1.99
Copula *be*	1,343	↑3.27	432	↓2.94	3.59
Conjunctions	984	↑2.39	303	↓2.07	**5.17**
Disclaimers	478	↑1.16	151	↓1.03	1.03
Stance markers	378	↑0.92	107	↓0.73	**4.67**
Pronoun *it*[1]	337	↑0.82	113	↓0.77	0.33
Emphatics	168	↑0.41	54	↓0.37	0.45
2nd person pronouns	123	↑0.30	36	↓0.25	1.13
Downtoners	113	↑0.27	21	↓0.14	**8.70**

Table 2. *Frequency (↑ = higher relative frequency; ↓ = lower relative frequency) and significance of selected linguistic features in Glass's fabricated vs. verified articles. Frequencies are shown as percentages of total corpus tokens. Note: log-likelihood values above 3.84 (shown in bold) indicate statistical significance at p<0.05. Counts for emphatics, downtoners, disclaimers, and stance markers are based on the word lists provided in Appendix B*

Figure 1 displays normalised percentages for each feature, while Figure 2 reports their statistical significance. The most pronounced contrasts are the lower frequency of nouns and the higher use of pronouns, particularly third-person forms, in the fabricated texts.

Taken together, these results point to a distinct distributional profile in Glass's fabricated writing. While the discursive interpretation of these patterns is developed in Section 4.3, two patterns are already evident: reduced noun frequency, indicating lower lexical density, and increased use of third-person pronouns, suggesting a shift toward more personalised, narrative-oriented discourse. In aggregate, fabricated articles rely less on nominal structures and more on verbs and pronouns than verified articles. The concordance and collocation analyses that follow examine how these quantitative differences operate rhetorically in context.

---

1   'Pronoun *it*' is a subset reported separately for interpretive interest.
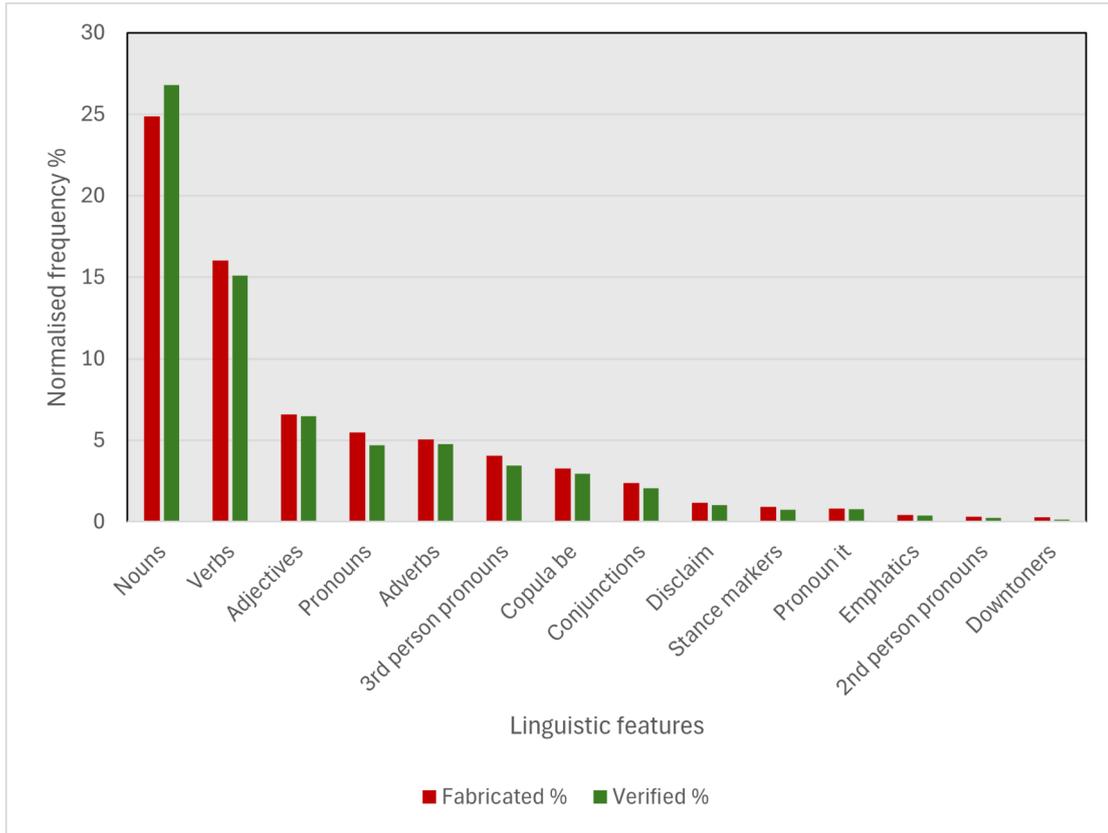
Figure 1: Normalised frequencies (%) of linguistic features in Glass's fabricated vs. verified articles
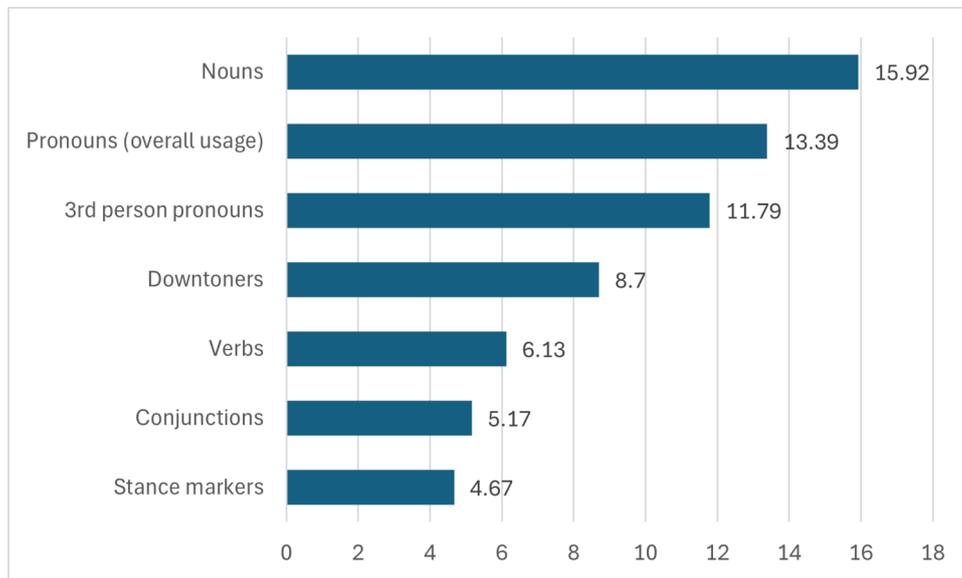

Figure 2: Log-likelihood (LL) values for linguistic features showing significant differences between Glass's fabricated and verified articles (LL>3.84; p<0.05)

Before turning to those analyses, brief examples illustrate how the observed distributional patterns surface in the texts themselves. Examples (1) and (2), both drawn from fabricated articles, exemplify this trend.

> (1) **He** *quit* **his** dollar-a-day job delivering ice and *became* a hack. Working tirelessly, **he** *saved* a small fortune. (Glass_Fabricated_Authorial)

In (1), individual agency is foregrounded through third-person pronouns (*he, his*) and dynamic verbs (*quit, became, saved*), consistent with the distributional tendencies reported in Table 2. The clause structure focuses on concrete actions and actors, with little recourse to abstract or nominalised expression.

> (2) '**I** *give* **them** whatever they *want*. **I** just *want* **my** life.' (Glass_Fabricated_Quoted)

Example (2) reinforces this pattern through immediate narration, personal pronouns (*I, them, my*), and active verbs (*give, want*). The phrase 'I just want my life' introduces explicit affect and alignment, privileging persuasive engagement over detached reportage.

By contrast, examples (3) and (4) from verified articles display greater lexical density and abstraction, shifting attention away from individual actors toward institutions, systems, and beliefs.

> (3) Eight years ago, Johansen retired from his job as a bookkeeper at a car dealership and turned his attention to **God's work** - converting the heathen to his **Baptist faith**. (Glass_Verified_Authorial)

Although narrative in form, example (3) includes abstract noun phrases (*God's work, Baptist faith*), aligning with the higher noun frequency observed in the verified corpus (Table 2) and with more institutional framing.

> (4) New York farmers were further helped by a bizarre **regulation** called 'the distance differential'. Under this rule, the **USDA** requires **processors** to pay farmers more the farther their cows graze from Eau Claire, Wisconsin. (Glass_Verified_Authorial)

In example (4), lexical density is realised through nominal structures (*regulation, rule, USDA, processors*), which foreground systems and processes rather than personal agency. This pattern is consistent with the more reportorial style of the verified texts.

These findings parallel those reported by Grieve and Woodfield (2023) in their analysis of Jayson Blair's fabricated journalism, which likewise showed reduced noun density alongside increased reliance on verbs, pronouns, and stance-related items. In Glass's case, this stylistic configuration similarly foregrounds personal voice and dynamic narration, reflecting the contrast Biber (1988, 1995) draws between informational writing and more informal, subjective registers.

Although direct quantitative comparison is not possible due to differences in corpus design and metrics, the directional tendencies observed in Glass's fabricated reporting broadly align with those reported by Grieve and Woodfield (2023) for Blair. In both cases, fabricated articles display reduced nominal density alongside increased use of verbs, pronouns, and stance-related modifiers, suggesting a shared tendency toward narrative

immediacy and personalisation in deceptive journalism. Read in this contextual sense, the results support the view that fabricated writing prioritises narrative dynamism over informational density.

The following section examines whether this stylistic divergence also extends to the lexical and evaluative level.

4.2.   Negation as a deceptive linguistic strategy

Although negation was excluded from the main frequency profile (Section 4.1), it was examined separately due to its established association with stance-taking and deceptive discourse (e.g., Govaert *et al.*, 2020; Newman *et al.*, 2003). Markers such as *not*, *no*, *never*, *none*, and *n't* were examined not only as grammatical operators, but as potential stylistic resources for realising stance, managing credibility, and guiding reader alignment.

Table 3 presents absolute and normalised frequencies for these markers across the subcorpora, while Figure 3 visualises their distribution. Overall, negation occurs more frequently in Glass's fabricated articles, with a combined log-likelihood (LL) of 9.28 indicating statistical significance.

Fabricated corpus			Verified corpus			Log-likelihood
Word	**Relative frequency per 10,000**	**Absolute frequency**	**Word**	**Relative frequency per 10,000**	**Absolute frequency**	
n't	↑45.48	187	n't	↓40.21	59	0.69
not	↑44.51	183	not	↓33.40	49	3.36
no	↑22.62	93	no	↓9.54	14	**11.13**
never	↓7.54	31	never	↑8.18	12	0.06
none	↑1.70	7	none	↓0.00	0	**4.27**
Total	↑121.86	501		↓91.33	134	**9.28**

Table 3. *Frequency and significance of selected negation markers in Stephen Glass's fabricated and verified articles. Note: log-likelihood values above 3.84 (shown in bold) indicate statistical significance at p<0.05. TOTAL log-likelihood (LL) computed from pooled counts of all listed negation tokens across corpora (not a sum of individual LLs) following Rayson (2001)*
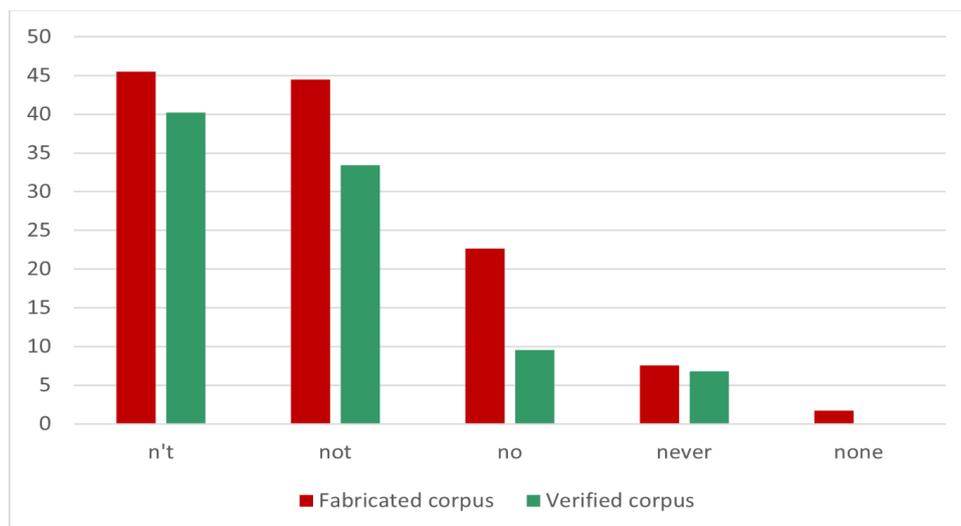
Figure 3: Normalised frequency of negation markers in Stephen Glass's fabricated and verified articles (per 10,000 words)

To examine how this distributional difference is realised in use, concordance lines were analysed qualitatively. Example (5) illustrates a recurrent pattern in the fabricated corpus:

> (5) There is **no** evidence of this in Young's report. There is **no** quote, positive or negative, from any union representative; there is **no** account of how the management treats the union representatives… (Glass_Fabricated_Authorial)

In this extract, the repeated use of negation (*no evidence, no quote, no account*) constructs a monoglossic frame of certainty, excluding alternative viewpoints and foreclosing contestation. In Appraisal Theory terms (Martin & White, 2005), this pattern can be classified as ENGAGEMENT: DENIAL, functioning to reject competing voices and assert authorial control over interpretation.

Read in this functional context, the higher frequency of negation in fabricated reporting suggests that negation operates not simply as logical contradiction, but as a stance resource that manages epistemic positioning. This interpretation aligns with findings by Govaert *et al.* (2020), who associate negation in fabricated journalism with discursive distancing and defensiveness, as well as with psychological research indicating that deceptive language often anticipates challenge or constrains interpretation through denial (De-Paulo *et al.*, 2003; Hauch *et al.*, 2015). In Glass's fabricated articles, negation thus appears to function interpersonally, positioning the journalist as authoritative while narrowing the space for alternative readings.

The following section (4.3) extends this analysis by examining whether evaluative expressions such as *most* and *just* differ across fabricated and verified reporting not only in frequency, but in their rhetorical deployment.

4.3. Concordance evidence: evaluative language in context

To examine how the distributional contrasts identified in Section 4.1 are realised in use, this section presents a qualitative analysis of selected evaluative lexical items through which stance may be realised. These items were chosen because, as shown in Section 4.1 and established through the frequency profiling outlined in Section 3.3.1, they occur more frequently in fabricated articles and have recognised potential to enact stance and guide reader alignment (Martin & White, 2005; Trnavac & Põldvere, 2024; Zhao & Fu, 2024).

Five high-frequency items (*most*, *just*, *not*, *only*, and *but*) were initially selected based on their evaluative potential and their occurrence across both subcorpora (see Appendix B), allowing for functional comparison under controlled conditions. However, as detailed in Section 3.3.2, *but* was evaluative in 100% of instances in both corpora and therefore did not show contrastive variation by truth status. It is therefore excluded from the coding results that follow, which focus on the remaining four items.

Final item selection followed the frequency profiling in Section 3.3, with reporting verbs such as *say* excluded because they primarily functioned to introduce reported speech.

Each occurrence of the four remaining items was manually coded for evaluative function (evaluative vs. non-evaluative) and source attribution (authorial vs. attributed), following the Appraisal-informed framework outlined in Section 3.3 (see Appendix C for operational coding criteria). This procedure enables close examination of how these resources are deployed in fabricated and verified journalism.

As shown in Figure 4, *most* and *just* are used evaluatively in 77% and 90% of instances, respectively, in the fabricated corpus, compared with 63% and 25% in verified articles. This indicates that, when fabricating, Glass relies more heavily on subtle intensifiers to frame certainty, constrain interpretation, and align readers with the narrative perspective. Evaluative uses of *not* and *only* are less frequent overall but show a slight increase in the fabricated texts.

These patterns align with findings by Trnavac and Põldvere (2024), who show that deceptive news discourse often uses evaluative language to intensify claims (GRADUATION) and guide reader alignment (ENGAGEMENT). In Glass's fabricated corpus, the frequent evaluative use of *just* and *most* reflects this dual role. Within Appraisal Theory (Martin & White, 2005), such items function as GRADUATION resources, adjusting the strength of authorial commitment, while not and only operate as ENGAGEMENT resources by restricting the range of interpretations made available to the reader. Together, these choices allow fabricated reporting to project confidence and immediacy without relying on explicit assertion.
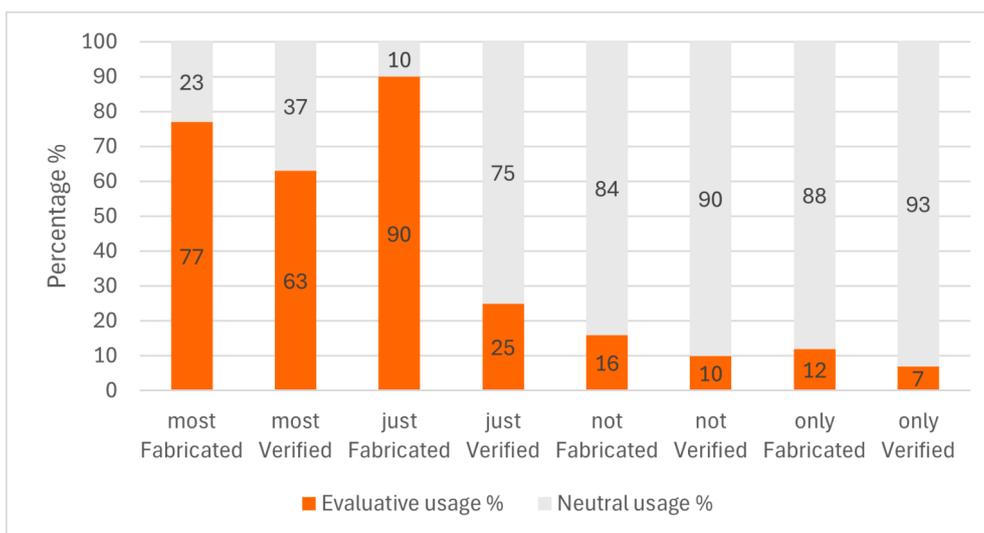
Figure 4: Proportion of evaluative vs. neutral usage for target words in Glass's fabricated and verified reporting. Note that but *was always evaluative in both subcorpora and therefore not included in this figure*

In Glass's fabricated articles, these evaluative devices co-occur more frequently with explicit authorial presence (see Section 4.3.4), with narration relying less on informational detachment. This pattern is consistent with findings by Baissa *et al.* (2025), whose Discursive News Values Analysis shows that deceptive news foregrounds evaluative dimensions such as credibility, rationality, morality, and emotion. Although Glass's texts belong to professional print journalism rather than digital disinformation, the evaluative clustering observed here similarly constructs epistemic authority and moral conviction as markers of credibility.

The following sections illustrate these contrasts through item-specific concordance analysis.

4.3.1 *Most*: evaluative generalisation vs. factual classification

The emphatic *most* appears frequently in Glass's fabricated corpus as an intensifier or generaliser, often conveying implied consensus or authority. Rather than requiring explicit evidential support, it broadens the scope of a claim, allowing it to appear widely shared or representative. To illustrate these tendencies, the following concordance lines in examples (6) and (7) show typical uses of *most* in each subcorpus:

(6) '**Most** of my peers want to know why they should drive a lot of rich white people everywhere they want to go', Slippy says. (Glass_Fabricated_Quoted)

(7) With little regulation and—unlike **most** major cities—no cap on the number of licenses, nearly anyone can get a hack permit. (Glass_Fabricated_Authorial)

In both examples, *most* operates evaluatively, either exaggerating commonality (6) or sharpening contrast (7). These uses frame claims as socially endorsed or normatively deviant, contributing to a persuasive rather than descriptive account. By contrast, in veri-

fied reporting, *most* tends to function more descriptively, as shown in examples (8) and (9):

> (8) The war between the two lactose lobbies is, like **most** congressional wars, about
>   government money. (Glass_Verified_Authorial)

> (9) So far the CAA's **most** debilitating provision is its requirement that congressional offices
>   comply with a 1938 labor law... (Glass_Verified_Authorial)

Here, *most* supports generalised classification without overt stance, specifying quantity rather than evaluation (8 and 9). This contrast indicates that Glass uses *most* primarily as an evaluative resource in fabricated contexts, but in a more neutral, classificatory role in verified writing. The next subsection examines whether comparable shifts are evident in the use of other high-frequency stance markers.

4.3.2 *Just*: hedging, emphasis and stance softening

In the fabricated corpus, just frequently functions as a downtoner, softening claims or presenting them as limited or incidental. This allows emotionally charged or controversial content to be introduced in a conversational register, reducing the apparent strength of commitment while maintaining persuasive force, as shown in examples (10) and (11):

> (10) 'The heading **only** says we spoke with them. Sometimes it **jus**t may have been **very, very**
>   briefly.' (Glass_Fabricated_Quoted)

> (11) Barbara Bush has a similarly compelling need to rewrite history. It isn't **just** her husband's
>   reputation that is on the line; it is her son's future. (Glass_Fabricated_Authorial)

In (10), *just* and *only* restrict the scope of involvement, while *very, very* intensifies the speaker's minimising stance. In (11), *just* redirects evaluative focus from individual reputation to collective legacy, increasing emotional and political salience. In both cases, *just* performs evaluative work, shaping alignment and constraining interpretation.

In verified texts, however, *just* appears less frequently in evaluative contexts, as shown in examples (12) and (13):

> (12) **Just** before Thanksgiving, dairy reform was removed from the reconciliation bill.
>   (Glass_Verified_Authorial)

> (13) '**Just** frame it there at the top of the commercial: How much can we really trust Bill
>   Clinton?' (Glass_Verified_Quoted)

Here, *just* functions as a temporal marker (12) or directive (13), contributing to informational clarity rather than stance-taking. This contrast supports the pattern observed in Section 4.1, whereby evaluative loading is more characteristic of Glass's fabricated reporting. In these texts, *just* functions as a low-salience but recurrent stance marker, shaping reader interpretation without overt emphasis.

More broadly, these patterns suggest that items such as *most, just,* and *not* acquire evaluative force through repeated patterned use rather than inherent meaning. This interpretation aligns with Partington and Diegoli's (2025) account of evaluative priming and evaluative prosody, whereby stable evaluative meanings emerge from recurrent cotextual association.

4.3.3 *Not* and *Only*: limited evaluative contrast, but recurrent rhetorical use

Although *not* and *only* can convey stance through denial or exclusivity, they were coded as explicitly evaluative in fewer than 20% of cases in either corpus. In most contexts, *not* functions as a clausal negator, while *only* restricts scope without overt authorial emphasis.

Yet frequency matters. As shown in Section 4.2, *not* and related markers appear significantly more often in the fabricated corpus. To illustrate how this difference is realised in context, the following concordance examples (14 and 15) are typical of the fabricated corpus:

(14) 'dare exposure does **not** produce any long-term prevention efforts on adolescent drug use rates.' (Glass_Fabricated_Quoted)

(15) no one—**not** parents, **not** educators and certainly **not** dare officials—wanted to hear the bad news. (Glass_Fabricated_Authorial)

In these examples (14) and (15), negation is used to specify what is excluded (e.g., *does not produce*; *no one—not parents, not educators*), narrowing the range of possible interpretations. In fabricated texts, this pattern asserts certainty by denying alternatives, reducing opportunities for challenge while keeping the evaluative framing within the journalist's own voice.

4.3.4 Source attribution: authorial stance and the projection of authority

Beyond lexical choice, stance is also shaped by source attribution, that is, whether evaluative expressions are voiced by the journalist or attributed to a quoted speaker. Each instance of the four analysed lexical items was coded for this distinction.

While both subcorpora contain evaluative language in authorial and attributed voice, a consistent difference emerges. In the fabricated corpus, Glass realises stance through evaluative language more frequently in his own voice (see Table 4). This pattern indicates that fabricated reporting relies more heavily on direct authorial evaluation than on attribution to external sources.

This finding is consistent with Grieve and Woodfield's (2023) observation that Blair's fabricated articles made less use of conventional quotation and attribution structures, favouring author-centred stance expression. It also aligns with Trnavac and Põldvere's (2024) account of persuasive discourse, in which direct evaluation through GRADUATION and ENGAGEMENT resources supports credibility claims by foregrounding the writer's perspective.

Word	Fabricated Authorial %	Fabricated Quoted %	Verified Authorial %	Verified Quoted %
most	↑87%	↓13%	↓20%	↑80%
just	↑88%	↓12%	↓8%	↑92%
not	↑92%	↓8%	↓0%	↑100%
only	↑90%	↓10%	↓0%	↑100%

Table 4. *Relative frequencies (%) of authorial and attributed evaluative uses of target words in fabricated and verified subcorpora*

In Glass's fabricated articles, evaluative meaning is embedded within the narrative voice, while evaluative language in verified reporting appears more often in quoted speech, maintaining clearer separation between journalist and source.

Comparable attributional patterns have been identified in digital contexts. Sousa-Silva (2022) and Jaworska (2023) show that reduced attribution and the use of evaluative or emotionally loaded language are common persuasive strategies, while Lebernegg *et al.* (2025) demonstrate how such practices can simulate objectivity despite strong stance-taking.

The present study extends these findings by showing that similar attributional cues operate not only in online or partisan disinformation, but also in mainstream professional journalism when reporting is fabricated.

Overall, the evidence points to a configuration of stylistic choices rather than isolated markers. Emphasis (e.g., *most, just*), denial (*no, not*), and increased reliance on authorial voice recur across the fabricated corpus, contributing to the presentation of credibility while reducing interpretive openness. The next section examines how these features co-occur locally through collocational analysis.

4.3.5 Collocation evidence: patterns of evaluative framing

As outlined in Section 3.3.2, collocate lists were generated for a subset of evaluative lexical items (*just, very, not, only, seem*, and the copula *be*). These items were selected on two grounds: their relevance to the Appraisal-informed categories of GRADUATION, ENGAGEMENT, and ATTRIBUTION, and their prominence in the frequency profile (Section 4.1). The analysis focuses on co-occurrence patterns, examining whether evaluative meaning is realised through recurrent lexical associations rather than individual items in isolation.

The analysis reveals systematic stylistic contrasts across the subcorpora. In fabricated texts, evaluative items frequently co-occur with intensifiers (*much, always*), hedgers (*just, briefly, merely*), and negation (*not, never, nothing*). These recurrent combinations form local clusters in which evaluation is intensified, moderated, or redirected. Compared with verified texts, the fabricated corpus displays a denser and more varied collocational pro-

file, suggesting that evaluative meaning is more often realised through layered patterning rather than isolated lexical choices.

To illustrate these differences, Tables 5a and 5b summarise the most frequent collocates and their observed discourse functions in the fabricated and verified subcorpora, respectively.

Evaluative item	Key collocates: Fabricated	Interpretation
just	go, be, do, have, not, as, simply	Persuasive downtoner. Softens speaker stance; often paired with vague verbs and negation.
very	good, briefly, much	Amplifies vague evaluation (e.g., very good, very briefly)
not	do, be, have, know, want, get, really, just, even, merely, only, much	Constructs denial or deflection. Adverb collocates shape authorial stance.
only	not, say	Used in limiting constructions (e.g., only say, not only); enhances persuasive framing.
seem	not, it, be	Indicates soft commitment; combines with negation to express doubt.
say	also, only, not, nothing, it, he, she, they, I	Blends authorial and attributed stance; negation and pronouns suggest distancing.
be	not, never, also, just, only, always, everywhere, really, all	Generalisation and moral emphasis. Marks emotional tone or absolutes.

Table 5a. *Key collocates and interpretations: Fabricated subcorpus. Note: All collocates met a minimum frequency threshold of ≥2. Interpretations reflect Appraisal-informed discourse patterns discussed in Sections 3.3.2*

These tables show that the same evaluative items pattern differently across the fabricated and verified corpora. For example, *just* appears in both corpora but exceeds the frequency threshold only in the fabricated corpus, where it co-occurs with high-frequency, semantically general verbs (*go, be, do, have*) and with negation (*not*). In these contexts, *just* functions as a downtoner, narrowing claims while maintaining persuasive force.

Similarly, *be* co-occurs with generalising or amplifying modifiers (*always, everywhere, really*) in fabricated texts, contributing to emotional emphasis or moral generalisation. In verified texts, by contrast, *be* more often collocates with temporal or factual modifiers (*still, now, precisely*), consistent with a more informational style. These contrasts align with research on evaluative clustering in deceptive discourse (Grieve & Woodfield, 2023; Trnavac & Põldvere, 2024) and support the view that stance is systematically realised through evaluative patterning in fabricated reporting.

Evaluative item	Key Collocates: Verified	Interpretation
just	N/A	Collocates do not reach frequency threshold; limited hedging role.
very	little, surprising	Used sparingly and often with understatement or irony (e.g., very little, very surprising).
not	think, want, do, be, have, get	Typically paired with cognitive verbs, indicating distancing or careful negation.
only	be, really	Indicates constraint; often paired with intensifiers to downplay commitment (e.g., only really).
seem	be, really	Used for cautious framing, less frequently combined with negation.
say	it, he, she, they, I	Primarily marks attribution. Fewer co-occurrences with evaluative terms suggest a conventional reporting function.
be	still, now, precisely, also, exactly, simple, about	Used for precision and temporality. Marks a measured, factual tone.

Table 5b. *Key collocates and interpretations: Verified subcorpus. Note: All collocates met a minimum frequency threshold of ≥2. Interpretations reflect Appraisal-informed discourse patterns discussed in Sections 3.3.2*

The following examples (16 and 17) illustrate how these collocational patterns operate in context:

(16) They **just** *don't* believe the dare program works. (Glass_Fabricated_Authorial)

(17) 'We **just** *do not* agree with one of the major findings.' (Glass_Fabricated_Quoted)

In both cases, *just* collocates with *do* and *not*, combining hedging and negation (*just don't believe; just do not agree*). This pattern reduces the force of disagreement, presenting dissent as measured rather than confrontational.

(18) 'I think it will *really* show that I **am** dedicated to my job…' (Glass_Fabricated_Quoted)

(19) 'I'**m** *not really* sure what else to say.' (Glass_Fabricated_Quoted)

(20) 'Just because he wasn't re-elected doesn't mean he **was***n't really* loved…' (Glass_Fabricated_Quoted)

These examples (18–20) show how *be* (realised as *am, I'm, wasn't*) co-occurs with *really*, including within quoted speech. In (18), *really* amplifies sincerity; in (19), *not really sure* softens uncertainty; and in (20), *wasn't really loved* mitigates denial. Across these in-

stances, intensification and mitigation co-occur, allowing evaluation to be calibrated even when stance is attributed rather than authorial.

Overall, the collocational evidence indicates that Glass's fabricated journalism relies less on isolated lexical cues and more on recurrent evaluative configurations. Through the repeated co-occurrence of intensification, hedging, and negation, evaluative meaning is layered across local discourse environments, contributing to the presentation of credibility while limiting interpretive openness. This supports the view that linguistic deception emerges from patterned language use, rather than from individual markers.

The final section draws these findings together to consider how such patterning contributes to the discursive construction of credibility in fabricated journalism.

5. Conclusion

This study set out to examine whether fabricated reporting can simulate credibility through grammatical, lexical, and evaluative choices that create an impression of sincerity and authority. Using a mixed-methods, corpus-assisted discourse analysis of Stephen Glass's fabricated and verified journalism, the findings suggest that deception is systematic rather than random, leaving recurring linguistic patterns even within the same register and publication context.

Across the three analytical stages (frequency profiling, collocation, and concordance analysis), fabricated texts diverged in consistent and interpretable ways. Quantitatively, they contained fewer nouns and more verbs, pronouns (especially third person and *it*), downtoners, and evaluative items. These grammatical tendencies reduce informational density while increasing immediacy and personalisation, producing prose that appears more vivid and engaging than detached or reportorial.

At the lexical level, fabricated articles relied more heavily on evaluative words such as *most, just*, and *not*. Concordance analysis showed that these items were frequently used to emphasise, soften, or signal certainty: functions that shape how events and characters are framed for the reader. In verified articles, the same items appeared more often in descriptive or classificatory roles, contributing less to the realisation of stance or the management of reader alignment.

Collocation analysis reinforced these distinctions. In fabricated texts, evaluative items clustered with intensifiers (*much, always*), hedges (*just, merely*), and negators (*not, never*), forming recurrent local patterns that heighten emotion, soften responsibility, or narrow interpretive possibilities. Verified reporting, by contrast, showed collocates expressing caution and precision (*very little, precisely, still*), consistent with a more informational style. Across analyses, these results indicate that fabricated journalism constructs credibility through repeated combinations of emphasis, hedging, and denial, rather than through isolated lexical choices.

Because these contrasts were observed in texts by the same author, within the same publication and register, they are unlikely to reflect house style or audience targeting. While the study does not make claims about authorial intent, some of the recurring pat-

terns, particularly in evaluative language, are consistent with the view that stylistic cues may emerge below conscious control (Newman *et al.*, 2003), reflecting both deliberate rhetorical choices and linguistic strain associated with fabrication (Hancock *et al.*, 2008).

More broadly, these findings show that fabricated journalism differs from verified reporting not only in content, but in how stance is realised through patterned evaluation to guide reader interpretation. Glass's fabrications move toward a mode of narration that privileges immediacy, affect, and authorial presence. This tendency aligns with research linking deception to reduced abstraction and increased reliance on evaluative resources (Biber, 1995; Grieve & Woodfield, 2023; Trnavac & Põldvere, 2024). Comparable strategies have also been identified in digital disinformation, where fabricated voices similarly rely on personalisation and conviction (Jaworska, 2025). These parallels point to a shared rhetorical tendency in deceptive journalism: a preference for directness and subjectivity over detachment and attribution.

Methodologically, the study demonstrates the value of combining frequency profiling, collocation, and concordance within an Appraisal-informed framework. This approach links quantitative distributional patterns to their local rhetorical functions and, by controlling for author and register, shows that the differences observed are closely associated with fabrication itself. In this respect, the findings support Hollmann and Gillings's (2025) argument that deception cues are most informative when categories are linguistically motivated and context-sensitive, rather than mechanically defined.

Although limited to one journalist and a specific historical context, the study contributes to wider debates on the linguistic enactment of deception. It shows that fabricated reporting may be distinguished from verified journalism through recurrent grammatical and evaluative patterning that collectively supports the presentation of credibility. Future research could apply this framework to contemporary digital disinformation or extend it through comparative analyses across journalists and outlets. In an environment of increasing media scepticism, understanding how deception operates linguistically, how it is patterned, perceived, and made persuasive, remains an important task for scholars, educators, and regulators of news discourse.

In short, deception may be better understood not as a collection of surface cues, but as a patterned way of writing in which stance is realised through evaluative language.

Acknowledgements

Competing interests

The author has no competing interests to declare.

Data accessibility

The data analysed in this study are publicly available and details of corpus composition are provided in the Appendices.

References

Ali, M., & Levine, T. (2008). The language of truthful and deceptive denials and confessions. *Communication Reports, 21*(2), 82–91. https://doi.org/10.1080/08934210802381862

Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). London: Bloomsbury.

Baranda, D. (2018, September 9). Still shattered: Revisiting the Stephen Glass case. *Medium*. Retrieved from https://medium.com/illumination/still-shattered-revisiting-the-stephen-glass-case-1a016d54c177

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511519871

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316410899

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139–173. https://doi.org/10.1075/ijcl.20.2.01bre

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Ehrlich, M. C. (2005). Shattered Glass, movies, and the free press myth. *Journal of Communication Inquiry, 29*(2), 103–118. https://doi.org/10.1177/0196859904272741

Fairclough, N. (2010). *Critical discourse analysis: The critical study of language* (2nd ed.). London: Routledge.

Fletcher, R., Andı, S., Badrinathan, S., Eddy, K. A., Kalogeropoulos, A., Mont'Alverne, C., Robertson, C. T., Ross Arguedas, A., Schulz, A., Toff, B., & Nielsen, R. K. (2025). The link between changing news use and trust: Longitudinal analysis of 46 countries. *Journal of Communication, 75*(1), 1–15. https://doi.org/10.1093/joc/jqae044

Gillings, M., & Mautner, G. (2024). Concordance lines: What are they and how can they be used to explore representation? In F. Heritage & C. Taylor (Eds.), *Analysing*

representation: A corpus and discourse textbook (pp. 12–26). London: Routledge. https://doi.org/10.4324/9781003350972-2

Govaert, C., Lagerwerf, L., & Klemm, C. (2020). Deceptive journalism: Characteristics of untrustworthy news items. *Journalism Practice, 14*(6), 697–713. https://doi.org/10.1080/17512786.2019.1637768

Grieve, J., & Woodfield, H. (2023). *The language of fake news.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009349161

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. *Discourse Processes, 45*(1), 1–23. https://doi.org/10.1080/01638530701739181

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review, 19*(4), 307–342. https://doi.org/10.1177/1088868314556539

Hollmann, W. B., & Gillings, M. (2025). Adjectives and deception: A view from linguistic theory. *Applied Corpus Linguistics, 5*(3), 1–10. https://doi.org/10.1016/j.acorp.2025.100150

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies, 7*(2), 173–192. https://doi.org/10.1177/1461445605050365

Jaworska, S. (2023). Fakespeak in 280 characters: Exploring the language of disinformation on Twitter through a comparative corpus-based approach. In S. M. Maci, M. Demata, M. McGlashan, & P. Seargeant (Eds.), *The Routledge handbook of discourse and disinformation* (pp. 103–117). London: Routledge. https://doi.org/10.4324/9781003224495-8

Jaworska, S. (2025). Engaging to deceive: Strategies of synthetic involvement in fake tweets. *Linguistics Vanguard: Multimodal Online Journal.* https://doi.org/10.1515/lingvan-2024-0124

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Lasorsa, D. L., & Dai, J. (2007). Newsroom's normal accident? An exploratory study of 10 cases of journalistic deception. *Journalism Practice, 1*(2), 159–174. https://doi.org/10.1080/17512780701275473

Lebernegg, N., Eberl, J.-M., Tolochko, P., & Boomgaarden, H. (2025). Do You Speak Disinformation? Computational Detection of Deceptive News-Like Content Using Linguistic and Stylistic Features. *Digital Journalism, 13*(8), 1373–1398. https://doi.org/10.1080/21670811.2024.2305792

Lugea, J. (2021). Linguistic approaches to fake news detection. In Deepak P., T. Chakraborty, C. Long, & S. Kumar G. (Eds.), *Data science for fake news: Surveys and perspectives* (pp. 287–302). Switzerland: Springer. https://doi.org/10.1007/978-3-030-62696-9_15

Marchi, A. (2024). Evaluating corpus and discourse work. In F. Heritage & C. Taylor (Eds.), *Analysing representation: A corpus and discourse textbook* (pp. 212–222). London: Routledge. https://doi.org/10.4324/9781003350972-14

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

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Newman, N. (2024, 17 June). Overview and key findings of the 2024 Digital News Report. *Reuters Institute digital news report.* Retrieved from https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024

Partington, A., & Diegoli, E. (2025). Lexical priming theory: Evolution, evaluation, extension. *International Journal of Corpus Linguistics, 30*(3), 352–375. https://doi.org/10.1075/ijcl.24030.par

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS).* Amsterdam: John Benjamins.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017, September). Truth of varying shades: Analyzing language in fake news and political fact-checking. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2931–2937). Copenhagen: Association for Computational Linguistics. https://doi.org/10.18653/v1/D17-1317

Rayson, P. (2001). *Log-likelihood and effect size calculator.* UCREL, Lancaster University. Available at http://ucrel.lancs.ac.uk/llwizard.html

Reuters. (1998, June 13). Magazine apologizes for fabricated articles. *Los Angeles Times*. Retrieved from https://www.latimes.com/archives/la-xpm-1998-jun-13-mn-59510-story.html

Reuters Institute for the Study of Journalism. (2025). *Digital news report 2025*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025

Savage, M. (2025, April 1). Outlets seek fresh strategies as UK poll shows 'news avoidance' on the rise. *The Guardian*. Retrieved from https://www.theguardian.com/media/2025/apr/01/outlets-seek-fresh-strategies-as-uk-poll-shows-news-avoidance-on-the-rise

Sousa-Silva, R. (2022). Fighting the fake: A forensic linguistic analysis to fake news detection. *International Journal for the Semiotics of Law, 35*(6), 2409–2433. https://doi.org/10.1007/s11196-022-09901-w

Spurlock, J. (2016). Why journalists lie: The troublesome times for Janet Cooke, Stephen Glass, Jayson Blair, and Brian Williams. *ETC: A Review of General Semantics, 73*(1), 71–76.

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

Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora 8*(1), 81–113. https://doi.org/10.3366/cor.2013.0035

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

Trnavac, R., & Põldvere, N. (2024). Investigating appraisal and the language of evaluation in fake news corpora. *Corpus Pragmatics, 8*, 107–130. https://doi.org/10.1007/s41701-023-00162-x

Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism, 6*(8), 951–963. https://doi.org/10.1080/21670811.2018.1502047

White, P. R. R. (2012). Exploring the axiological workings of 'reporter voice' news stories: Attribution and attitudinal positioning. *Discourse, Context & Media, 1*(2–3), 57–67. https://doi.org/10.1016/j.dcm.2012.10.004

Zhao, J., & Fu, C. (2024). Linguistic indicators for predicting the veracity of online health rumors. *Frontiers in Public Health, 11*, 1–10. https://doi.org/10.3389/fpubh.2023.1278503

Appendix A: Stephen Glass corpus metadata

All articles (11 verified, 12 fabricated) from *The New Republic*

Type	Date	Headline / Article Title	Topic	Word Count
Verified	Dec 25, 1995	Cheese Biz	Political lobbying / agricultural policy	999
Verified	Mar 18, 1996	Pat Speaks	Political rhetoric / right-wing extremism	648
Verified	Mar 25, 1996	Attack Dogs	Election politics / campaign strategy	862
Verified	Apr 15, 1996	Do As They Do	Congressional reform / workplace regulation	1280
Verified	May 13, 1996	Philadelphia Story	Urban politics / mayoral performance	1193
Verified	Jul 1, 1996	The Hall Monitor	Political reform / internal GOP conflict	1089
Verified	Nov 4, 1996	Deliverance	Consumer experience / corporate accountability	1164
Verified	Dec 9, 1996	Shalom, Y'All	Religion / proselytism and identity	1527
Verified	Dec 16, 1996	Ad and Subtract	Media / political image rehabilitation	1116
Verified	Jul 7, 1997	Gold Diggers	Political corruption / legislative manipulation	1489
Verified	Jul 28, 1997	Space Oddity	Science policy / international cooperation	1233
Total word count				**12600**
Fabricated	Aug 5, 1996	Taxis and the Meaning of Work	Consumer / corporate experience	4156
Fabricated	Jan 6 / Jan 13, 1997	Probable Claus	Religion / social identity	1717
Fabricated	Jan 27, 1997	Holy Trinity	Politics / governance	2437
Fabricated	Mar 3, 1997	Don't You D.A.R.E.	Science/ technology policy	7868
Fabricated	Mar 24, 1997	Writing on the Wall	Consumer/ corporate experience	1825
Fabricated	Mar 31, 1997	Spring Breakdown	Politics / governance	2195
Fabricated	June 9, 1997	Peddling Poppy	Media / image	4230
Fabricated	Jul 14 / Jul 21, 1997	Slavery Chic	Religion / social identity	1067

Type	Date	Headline / Article Title	Topic	Word Count
Fabricated	Sep 8 / Sep 15, 1997	The Young and the Feckless	Politics / governance	3876
Fabricated	Nov 17, 1997	Anatomy of a Policy Fraud	Politics / governance	3165
Fabricated	Apr 13, 1998	Monica Sells	Political youth culture / conservative movements	965
Fabricated	May 18, 1998	Hack Heaven	Consumer/ corporate experience	1073
Total word count				**34574**

Appendix B: Frequency table of evaluative lexical items

Note: highest counts are shown in bold; category labels such as 'stance marker' and 'stance verb' are used descriptively to group items by typical discourse function; stance itself is analysed as an effect realised through evaluative use in context (see Section 2.3)

	Fabricated corpus				Verified corpus		
Word	**Category**	**Relative frequency per 10,000**	**Absolute frequency**	**Word**	**Category**	**Relative frequency per 10,000**	**Absolute frequency**
very	Emphatic	9.00	37	very	Emphatic	6.82	10
really	Emphatic	7.05	29	really	Emphatic	4.77	7
totally	Emphatic	0.24	1	totally	Emphatic	0.68	1
absolutely	Emphatic	0.49	2	absolutely	Emphatic	0.68	1
completely	Emphatic	0.00	0	completely	Emphatic	0.00	0
especially	Emphatic	0.24	1	especially	Emphatic	1.36	2
utterly	Emphatic	0.24	1	utterly	Emphatic	0.00	0
perfectly	Emphatic	0.49	2	perfectly	Emphatic	1.68	1
most	**Emphatic**	**14.11**	58	**most**	**Emphatic**	**11.59**	17
always	Emphatic	3.89	16	always	Emphatic	2.04	3
indeed	Emphatic	2.19	9	indeed	Emphatic	3.41	5
certainly	Emphatic	1.46	6	certainly	Emphatic	2.04	3
exactly	Emphatic	1.46	6	exactly	Emphatic	2.73	4
		40.86	**168**			**36.80**	**54**
slightly	Downtoner	2.19	9	slightly	Downtoner	0.68	1

	Fabricated corpus				Verified corpus		
Word	Category	Relative frequency per 10,000	Absolute frequency	Word	Category	Relative frequency per 10,000	Absolute frequency
just	**Downtoner**	**16.78**	69	**just**	**Downtoner**	**10.91**	16
barely	Downtoner	0.73	3	barely	Downtoner	0.00	0
somewhat	Downtoner	1.46	6	somewhat	Downtoner	0.68	1
kind of	Downtoner	0.00	0	kind of	Downtoner	0.00	0
a bit	Downtoner	0.00	0	a bit	Downtoner	0.00	0
almost	Downtoner	5.35	22	almost	Downtoner	1.36	2
relatively	Downtoner	0.97	4	relatively	Downtoner	0.68	1
		27.48	113			14.31	21
not	**Disclaim**	**45.24**	186	**not**	**Disclaim**	**32.03**	47
never	Disclaim	7.54	31	never	Disclaim	8.18	12
only	**Disclaim**	**15.57**	64	**only**	**Disclaim**	**20.45**	30
but	**Disclaim**	**39.40**	162	**but**	**Disclaim**	**35.44**	52
however	Disclaim	2.68	11	however	Disclaim	4.09	6
nonetheless	Disclaim	0.24	1	nonetheless	Disclaim	0.00	0
although	Disclaim	3.41	14	although	Disclaim	1.36	2
yet	Disclaim	2.19	9	yet	Disclaim	1.36	2
		116.27	478			102.91	151
apparently	Stance Marker	0.73	3	apparently	Stance Marker	2.04	3
reportedly	Stance Marker	0.24	1	reportedly	Stance Marker	0.68	1
maybe	Stance Marker	0.49	2	maybe	Stance Marker	2.73	4
perhaps	Stance Marker	1.22	5	perhaps	Stance Marker	2.04	3
clearly	Stance Marker	0.49	2	clearly	Stance Marker	0.00	0
honestly	Stance Marker	0.00	0	honestly	Stance Marker	0.68	1
frankly	Stance Marker	0.00	0	frankly	Stance Marker	0.00	0
surely	Stance Marker	0.00	0	surely	Stance Marker	0.68	1

Fabricated corpus				Verified corpus			
Word	Category	Relative frequency per 10,000	Absolute frequency	Word	Category	Relative frequency per 10,000	Absolute frequency
say	Stance Verb	83.67	344	say	Stance Verb	59.30	87
seem	Stance Verb	5.11	21	seem	Stance Verb	4.77	7
		91.94	378			72.93	107
	TOTAL	276.55			TOTAL	226.95	

Appendix C: Evaluative status coding manual

This manual outlines the procedure used to determine whether a given lexical item in context was functioning evaluatively or neutrally, based on its discourse role, meaning, and grammatical context.

Step 1: Core definition
- Evaluative = contributes to the realisation of stance, subjectivity, contrast, intensity, or authorial attitude in context
- Neutral = functions only grammatically or factually, with no added stance or emotion

Step 2: Evaluation tests

Test Name	If True → Likely	Description
Substitution test	Evaluative	Can the word be removed or replaced with no change to factual meaning? If so, it adds emphasis.
Paraphrase test	Evaluative	Can a more neutral paraphrase be used instead? Signals stance or tone.
Grammatical function	Neutral	Is the word used structurally (e.g., not in 'don't')? Then likely neutral unless contrastive.
Literal category test	Neutral	Is the meaning literal (e.g., 'just now', 'most of the people')? Likely neutral.
Affect amplification	Evaluative	Does the word boost emotional or attitudinal intensity (e.g., 'just awful', 'most unfair')?
Contrast cue	Evaluative	Does it signal denial, limitation, or contrast (e.g., 'not', 'but', 'only')?
Narrative stance test	Evaluative	Does it reflect the speaker's viewpoint (e.g., doubt, emphasis, hedging)? Then evaluative.