

Article

Rape, Words, and Power: Corpus Stylistics of Victimhood in Nigerian Newspapers

Chisom Paula Ogamba¹, Olubunmi Idowu², Chibuzo Onunkwo³, Nwaugo Goodseed Ochulor⁴

^{1, 2 & 4} Department of Languages and Literary Studies, Faculty of Education and Humanities, Babcock University, Ilishan-Remo, Ogun, Nigeria; ³Department of English and Literary Studies, University of Nigeria, Nsukka

Correspondence: Chibuzo Onunkwo, chibuzo.onunkwo@unn.edu.ng

Abstract

This study examines how Nigerian newspapers linguistically construct rape narratives and victim identities through a Corpus-Assisted Critical Stylistics (CACS) framework. It addresses the gap in large-scale, data-driven analyses of rape reportage in African media by integrating Corpus Linguistics with Critical Stylistics, Feminist Critical Discourse Analysis, and Framing Theory. Using a corpus of 893 newspaper articles (2017–2025), the study employs keyword and collocation analyses in AntConc to identify patterns of prioritisation, presupposition, and epistemic hedging. The findings reveal that rape reportage foregrounds emotive narratives, moral evaluations, and vulnerable child victims, while sexual assault reportage emphasises institutional and legal discourse. Across both corpora, a hierarchical model of victimhood emerges, privileging young female children as "perfect victims" while marginalising adult women and male survivors. Crucially, the study shows that the epistemic marker allegedly functions not merely as a legal safeguard but as an ideological tool that distributes doubt unevenly, binding most strongly to explicit sexual acts and to the most vulnerable victim categories, thereby undermining victim credibility precisely where claims are most serious. These patterns are consistent with the reproduction of rape myths and with the obscuring of structural explanations of sexual violence. The study concludes that media discourse plays a critical role in sustaining gendered inequalities and recommends the adoption of more inclusive, structurally informed, and ethically responsible reporting practices.

Keywords: Rape narratives, Corpus linguistics, Critical stylistics, Media framing, Victim hierarchy, Epistemic hedging

Introduction

Language is recognised as a key instrument in structuring news narratives and influencing how audiences perceive the individuals or groups mentioned in them (Andrew & Alexopoulos, 2018). Because the language used in news reporting plays a crucial role in shaping narratives and guiding audience perceptions, media framing in reports on gender-based violence significantly influences public attitudes and understanding (Akano & Afolabi, 2024). This framing is realised through specific linguistic choices and discursive strategies, including vocabulary selection, grammatical structures such as transitivity processes and voice, and rhetorical devices, which collectively convey overt and covert ideological messages (Oisanwo & Ojo, 2022).

Consequently, it is important to critically examine discourse during framing processes, since linguistic choices shape how issues such as sexual violence are represented and understood. The importance of examining linguistic framing lies in the premise that language is not neutral (Ujevic, 2015) but reflects and sustains deeply seated gender inequalities and patriarchal social orders. For this reason, the focus of this study lies on the framing of rape narratives and victims in Nigeria, a patriarchal context in which sexual violence continues

to rank among the most widespread manifestations of gender inequality, oppression, and violations of human rights (Akano & Afolabi, 2024).

This study examines how linguistic patterns in Nigerian newspaper rape reportage construct and sustain a hierarchy of victimhood through processes of lexical prioritisation and collocational meaning-making. It specifically investigates how the discursive construction of the "Perfect Victim" is produced, reinforced, and naturalised, and how epistemic markers contribute to a culture of doubt surrounding rape allegations. In doing so, the study demonstrates that media language functions not merely as a vehicle of reporting, but as an active site of ideological construction that shapes the perception, legitimacy, and hierarchy of victims.

When Nigerian newspapers repeatedly cast Female Rape Victims (FRVs) as the "Goal" in material processes, they utilise verb voice and sentence structure to foreground the victim's passivity and the perpetrator's agency, or conversely use passive constructions to obscure the "Actor" and eliminate explicit responsibility (Osisanwo & Ojo, 2022). AntConc 4's keyword analysis is a fundamental method for identifying statistically salient words that define the "aboutness" of a text and its specific style (Al-Rawi, 2017), highlighting thematic foci such as *raped*, *girl*, *victim*, and *raping*. Keywords are units claimed to hold special status because they express important evaluative social meanings or play unique roles in a specific text type. Identifying them uses statistical measures such as log-likelihood to compare a target corpus against a reference corpus and identify words of unusual frequency. From the perspective of Jeffries' Critical Stylistics, these keywords function within the tool of "Naming, Describing and Prioritising," where the choice of specific nouns (e.g., *victim*, *girl*, *rapist*, *monster*) constitutes an ideological act that "packages" information to influence reader perception. These significant keywords therefore serve the function of prioritising, exploiting information structures to emphasise certain social actors over others (Tabbert, 2013).

While keywords indicate frequency, collocation analysis reveals the "aura of meaning" created by the company a word keeps. These collocations can be grouped into semantic fields—sets of semantically related words denoting a particular segment of reality. Consequently, the critical stylistics function of "Assuming and Implying" can be examined through collocation tables that reveal moral and evaluative associations, showing how victim terms are bundled with meanings such as "innocence," "vulnerability," or "disability," thereby pragmatically constructing the "Nigerian Perfect Victim." Such collocations trigger existential and logical presuppositions, packaging information (e.g., "incest rapist," "virgin victim") in ways that are unlikely to be questioned. The integration of AntConc's quantitative tools with Jeffries' Critical Stylistics provides a systematic toolkit for examining the "lexical architecture" of Nigerian rape narratives.

To achieve these aims, the study is guided by three research questions:

1. **RQ1:** How do keywords establish the critical stylistics function of prioritising?
2. **RQ2:** How do keywords help identify the "aboutness" of reportage?
3. **RQ3:** How do collocation profiles engender the critical stylistics function of assuming?

These questions are answered directly in the **Discussion** and summarised in **Table 1**.

Theoretical Framework

This study adopts a Corpus-Assisted Critical Stylistics approach to the analysis of rape narratives in Nigerian media, as it provides a multi-layered methodological framework that bridges quantitative data patterns and qualitative ideological critique.

Corpus Linguistics (CL) involves the use of computational tools to analyse large quantities of real-life language data, enabling the systematic and objective identification of statistically significant linguistic patterns that may be missed in manual analysis (Tabbert, 2013). Through tools such as word lists, concordances, and collocations, CL reveals frequent linguistic features, word choices, and semantic associations that contribute to framing (Turkewitz, 2010). Because rape reporting can be nuanced and biased, a methodological approach combining Corpus Linguistics and Critical Stylistics is appropriate for identifying embedded ideologies and linguistic constructions (Tabbert, 2013). CL facilitates the objective identification of predominant linguistic features, such as frequently used words and grammatical structures, providing a systematic and replicable analytical toolkit capable of revealing hidden ideologies without imposing preconceived interpretations.

Furthermore, linguistic features such as lexical choices, metaphors, rhetorical devices, and grammatical structures contribute to the construction of causality within a frame (Simonsen, 2024). Journalists' lexical choices reflect their attitudes, values, and beliefs (Barber, 2022), while evaluative language plays a key role in implicit or explicit blame attribution by foregrounding certain elements and backgrounding others (Simonsen, 2024). Framing thus operates by including and excluding voices, ideas, and interests, functioning as a tool for projecting ideology (van Hulst et al., 2024). Corpus-Assisted Critical Stylistics enables the identification of statistically significant patterns that reveal how language constructs and sustains hegemonic patriarchal ideologies in rape narratives, while also mitigating the subjectivity often associated with purely qualitative Critical Discourse Analysis through the use of computational tools such as AntConc (Akano & Afolabi, 2024).

Within this framework, the present study operationalises these tools to systematically uncover how lexical prioritisation and collocational patterning function as mechanisms through which victim hierarchies are discursively constructed and maintained in Nigerian rape reportage. To bridge social theory with linguistic analysis, the study synthesises Lazar's Feminist Critical Discourse Analysis (FCDA) (Lazar, 2014) and Entman's Framing Theory (Entman, 1993) with Jeffries' Critical Stylistics (Jeffries, 2010), operationalised through AntConc's corpus analysis tools (Anthony, 2004). This integration is justified by the need for methodological synergy in reducing researcher bias (Muchnik-Rozanov & Tsybulsky, 2021). While traditional CDA has been criticised for beginning with predefined conclusions (Tabbert, 2013), the use of AntConc allows statistically significant patterns to emerge from the data prior to interpretation. This quantitative foundation guides the qualitative analysis within a broader social critique, producing a "bottom-up" and "top-down" approach that grounds findings in empirical data.

Methodology

This study employs an explanatory sequential mixed-methods design, utilising a Corpus-Assisted Critical Stylistics (CACS) framework. This approach facilitates methodological synergy by integrating the quantitative rigour of Corpus Linguistics (CL) with the qualitative depth of Jeffries' (2010) Critical Stylistics, Lazar's (2014) Feminist Critical Discourse Analysis (FCDA), and Entman's (1993) Framing Theory.

Corpus Compilation

The data were purposively sampled from the digital archives of six major Nigerian national newspapers: *The Guardian*, *Punch*, *The Nation*, *Vanguard*, *The Sun*, and *Premium Times*. The sampling period spans from 2017 to September 2025, a timeframe selected to capture media discourse in the post-#MeToo movement era. All six outlets are national dailies; five are headquartered in Lagos. The corpus comprises **893 newspaper articles**, totalling approximately **548,649 tokens**.

The corpus was organised into two sub-corpora, used interchangeably as target and reference sets:

Sub-corpus	Articles	Approx. tokens
Target corpus: rape-related reports	516	~322,000
Reference corpus: sexual assault-related reports	377	~226,000

Following standard corpus linguistics protocols, all retrieved texts were converted into plain text format (.txt) and cleaned of non-relevant metadata to ensure compatibility with AntConc 4. The two sub-corpora are periodically swapped throughout the analysis to identify "keyness" – the statistically unusual frequency of words in one sub-corpus relative to the other. This genre-specific comparison ensures that identified keywords and collocations are salient to the discourse of rape rather than reflecting general journalistic style.

A critical boundary condition: "rape" and "sexual assault" are not fully parallel categories. "Rape" denotes a specific criminal act, whereas "sexual assault" functions as a broader legal-bureaucratic umbrella that in many formal contexts subsumes rape. This categorical asymmetry is itself analytically relevant: it helps explain why "sexual assault" collocates with institutional, statutory terminology (NAPTIP, DSVA, SARCs, petition, charge), since the term is used across legal-administrative registers. The study treats this asymmetry as an interpretive constraint – findings are framed as *relative* differences between the two sub-corpora, not as absolute properties of each discourse type.

Analytical Procedure

Data analysis was conducted in three stages using AntConc 4 (Anthony, 2004), comprising two computational phases interpreted through the critical stylistic functions of Prioritising and Assuming and Implying.

Statistical parameters. Keyness was calculated using the log-likelihood ratio, and effect size (Dice coefficient) was computed to measure the strength of collocational association independent of raw frequency. For keyword analysis, a significance threshold of $p < 0.05$ was applied (log-likelihood > 3.84). For collocation analysis, a minimum span of L5–R5 was used, with a minimum collocate frequency threshold of 10 to avoid chance associations, and effect size (Dice) was the primary measure of association strength. Log-likelihood and effect size are reported together because they answer different questions: log-likelihood reflects statistical salience (inflated by high frequency), whereas effect size reflects association strength.

Stage 1: Keyword Analysis (RQ1, RQ2) – the function of Prioritising. The Keyword List tool identifies statistically salient words in rape reportage by comparing the target and reference sub-corpora using the log-likelihood ratio. RQ1 ("prioritising") is answered by examining *which* social actors and themes the lexical architecture foregrounds or backgrounds. RQ2 ("aboutness") is answered by showing that these keywords collectively define the thematic focus of each reportage type. Within Critical Stylistics, this stage interrogates how Nigerian newspapers exploit information structures to foreground specific social actors or thematic concerns (e.g., foregrounding *girl* or *child*), establishing a hierarchy of importance that reinforces the "Perfect Victim" frame.

Stage 2: Collocation Analysis (RQ3) – the function of Assuming and Implying. The Collocation tool maps the "aura of meaning" around victim terms. Node words within a span of L5–R5 are analysed to generate collocation tables, enabling identification of evaluative and moral associations such as "innocence," "vulnerability," or "allegation." **Node-word selection rationale:** the five nodes analysed (*boy, child, girl, woman, allegedly*) were selected because (a) the keyword analysis identified them as statistically salient victim-category terms, and (b) together they span the full range of victim identities – male/female and child/adult – thereby allowing direct testing of the victim-hierarchy hypothesis. *Allegedly* was selected separately as the principal epistemic marker identified in Stage 1. These patterns are interpreted through the Critical Stylistics function of Assuming and Implying, identifying existential and logical presuppositions that package ideological constructs as naturalised facts.

Stage 3: Qualitative Synthesis. The quantitative findings are synthesised using Lazar's FCDA to critique discourses that sustain the patriarchal social order and Entman's Framing Theory to explain how the selection and salience of specific keywords and collocations construct the "Perfect Victim" versus "Imperfect Victim" binary. To safeguard against interpretive bias, the qualitative coding of collocation patterns into thematic categories (e.g., "age," "legal," "innocence") was conducted by two coders, with disagreements resolved through discussion to reach consensus (inter-coder agreement $> 85\%$).

Discussion of Findings

Stage 1: Keyword Analysis – The Function of Prioritising (RQ1, RQ2)

In the rape-versus-sexual-assault comparison, "rape" emerges as the most statistically frequent and relevant keyword in the rape corpus (log-likelihood = 609.918), with a substantially higher frequency, normalised frequency, and distributional range than in the

sexual assault corpus. This indicates that rape reportage explicitly names the act rather than subsuming it under general sexual violence categories, and that *raped* and *raping* reinforce a framing focused on the event itself. However, this explicit naming coexists with discursive patterns that shift attention towards surrounding circumstances rather than victims' lived experiences.

RQ1: How keywords perform the function of Prioritising

Prioritising is performed through the consistent foregrounding or backgrounding of social actors. Four patterns are salient.

(1) Lexical choices and moral surveillance. Personal pronouns centred on the victim ("I," "me," "my," "you") rank highly, producing an experiential framing that incorporates first-person survivor narratives. While this may appear to centre victim voices, from a critical stylistics perspective it also individualises rape experiences, potentially detaching them from structural explanations. Repeated foregrounding of personal pronouns frames rape as a series of personal tragedies rather than a pervasive social problem.

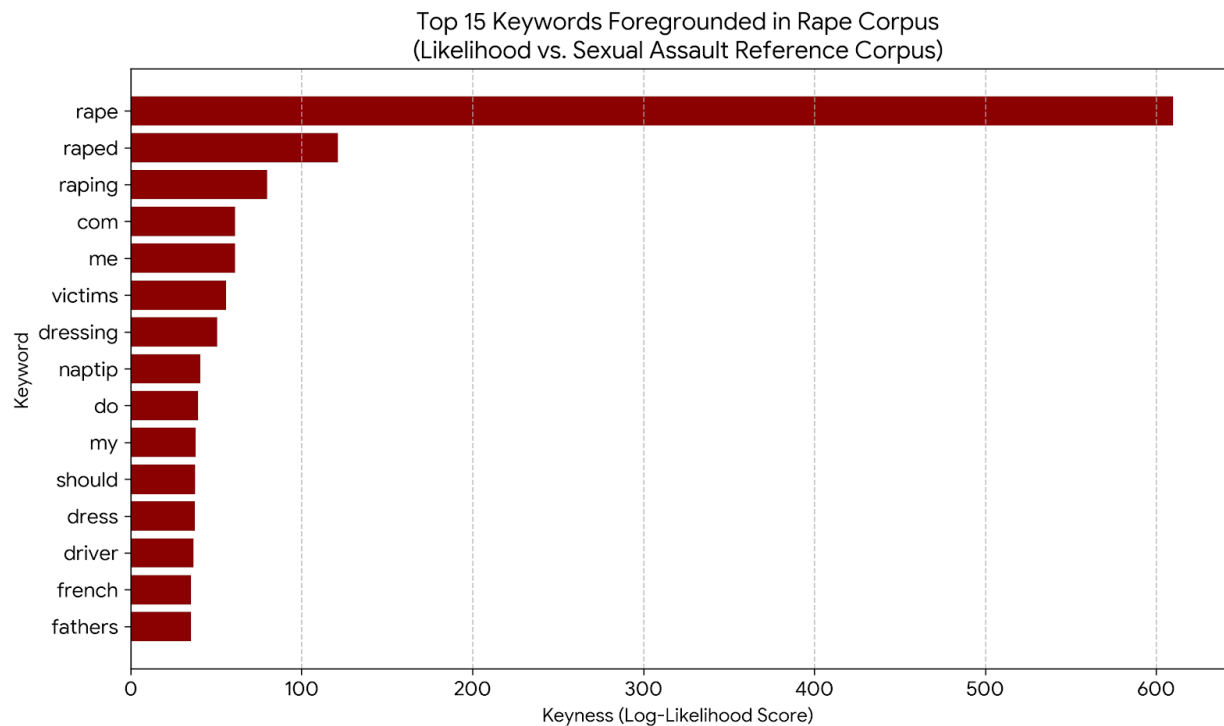
One of the most ideologically revealing findings is the statistical prominence of lexical items related to appearance and behaviour, such as "dressing" and "dress." Although low in raw frequency, their high log-likelihood values indicate they are disproportionately associated with rape reportage, creating moral surveillance and victim-blaming discourses. This foregrounding reflects moral evaluative framing in which victims' clothing and conduct are implicitly positioned as relevant explanatory factors, resonating with rape myth ideologies and reinforcing the construction of a conditional "perfect victim" whose credibility is tied to perceived respectability (Dhawan, 2018). Similarly, "parents," "permission," and "sleep" point towards narratives that scrutinise victims' actions, mobility, or supervision in cases involving girls and children.

(2) Institutional and punitive framing. A cluster of institutional and punitive terms – "NAPTIP," "deterrent," "punishment," "death," "castration," "jury," "penalty" – indicates that rape reportage frequently foregrounds punitive responses and state intervention. While this may signal seriousness, it simultaneously shifts narrative focus away from victim recovery and systemic prevention towards sensationalised punishment discourse, aligning with moral panic narratives and spectacle over structural critique.

(3) Gender, age, and victim hierarchy. The high frequency and range of "girl" in the rape corpus shows that rape reportage disproportionately centres young female victims, consistent with Christie's (1986) "ideal victim." References to "fathers," "uncle," and other familial relations contextualise rape within domestic and kinship settings while amplifying vulnerability as a criterion for sympathy. In contrast, adult women and male victims are less lexically prioritised, reinforcing a victim hierarchy in which youth, innocence, and femininity mark legitimate victimhood. Notably, terms related to patriarchy, gender inequality, rape culture, or systemic violence are absent among the most frequent keywords, indicating that rape reportage backgrounds structural and ideological explanations, favouring episodic over thematic framing.

(4) The "vanishing" actor. The high keyness of "victims" relative to the absence of perpetrator-focused keywords in the top 15 suggests a linguistic focus on the receiver of the act rather than the doer – a finding that prefigures the collocation analysis in Section 4.2.

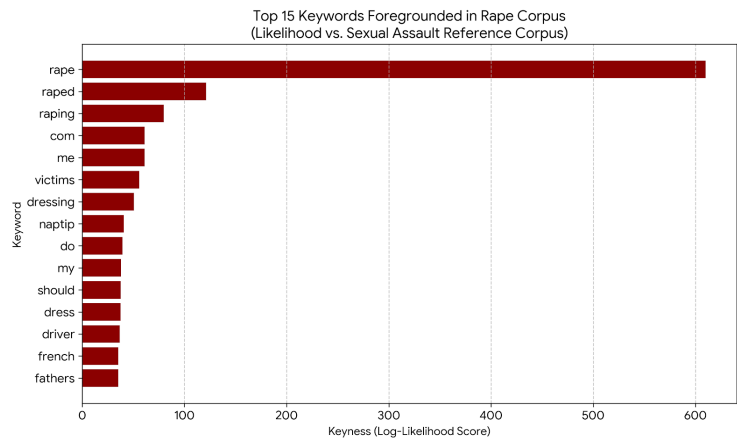
FIGURE 1



This chart illustrates the lexical items that are statistically more significant in “Rape” reports compared to general “Sexual Assault” reports. The presence of words like “dressing” and “dress” highlights “moral surveillance”. It suggests that rape reportage is more likely to foreground the victim’s attributes or perceived “innocence,” thereby engendering the “Perfect Victim” construct.

The “Vanishing” Actor: The high keyness of “victims” compared to the absence of perpetrator-focused keywords in the top 15 suggests a linguistic focus on the receiver of the act rather than the doer.

FIGURE 1 – Keyword salience in rape versus sexual assault reportage.



A horizontal bar chart (top 15 keywords by log-likelihood). High-salience items include *rape, girl, victims, raped, child, allegedly, rape, dressing, dress, NAPTIP, incest*. A call-out annotation above "dressing/dress" reads: "Moral surveillance: victim attributes foregrounded." A second annotation above "victims" reads: "Vanishing actor: focus on the receiver, not the doer."

This keyword analysis demonstrates that rape reportage prioritises explicit naming of the act, personal and experiential narratives, moral evaluation of victims' behaviour and appearance, punitive institutional responses, and young female children as primary subjects of sympathy. Conversely, it backgrounds structural gender inequalities, long-term consequences for survivors, and the cultural conditions sustaining rape culture.

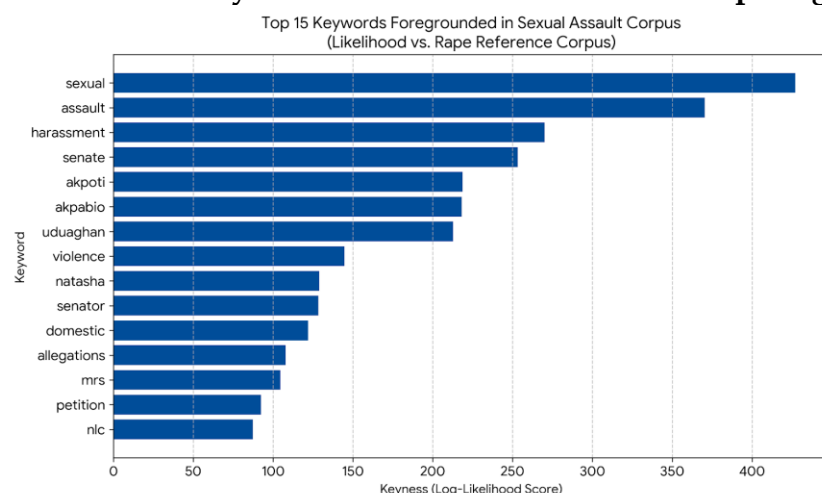
RQ2: How keywords define the "aboutness" of reportage

The keyword analysis of the sexual assault corpus reveals a distinct "aboutness." Whereas rape reportage is about *the act and the individual victim*, sexual assault reportage is about *the institution, the legal process, and the high-profile actor*.

The keywords with the highest keyness in the sexual assault corpus are "sexual" and "assault" (log-likelihood = 427.115 and 370.396 respectively), framing incidents through formal, legalistic terminology rather than experiential or narrative language. This institutional "aboutness" is reinforced by "petition," "committee," "charge," "evidence," "appellant," "prosecutrix," "rules," "ethics," and "board," which embed sexual assault within legal proceedings, administrative investigations, and bureaucratic processes. The victim's experience is thereby discursively mediated through institutional mechanisms.

A striking feature is the prominence of politically associated lexical items—"senate," "senator," "president," "committee," "elections," and named political actors (Akpabio, Akpoti, Uduaghan, Natasha). **A necessary caveat:** high keyness among proper nouns often reflects topical saturation from one or a few recurring high-profile cases (notably the Akpoti-Uduaghan/Senate controversy within the sampling period) rather than a general stylistic property. In treating these as evidence of political framing, the study acknowledges that person-name keyness is partly a single-case artefact; what is generalisable is the *institutional* register in which such cases are reported.

FIGURE 2— Keyword salience in sexual assault reportage.



A horizontal bar chart (top 15 keywords). Dominant items include *sexual, assault, senate, petition, allegations, Akpoti, NLC, committee, survivor, Lagos*. Call-out annotations read: "Political/bureaucratic framing" and "Depersonalised frame: institutional authority over survivor narrative."

In contrast to the rape corpus, the sexual assault corpus foregrounds "survivor," "gender," "support," "referral," "centre," "DSVA," "SARCs," and "SGBV"—a discourse aligned with advocacy, policy, and support frameworks. The use of "survivor" rather than "victim" indicates a shift towards empowerment associated with NGO, legal-reform, and institutional narratives. However, this terminology also abstracts individual experience into professionalised categories; victims become "cases" processed through centres, referrals, and committees, distancing readers from the embodied realities of sexual violence.

Lexical items such as "Lagos," "Ikeja," "Guardian," "news," "centre," and "devreporting" indicate that sexual assault reportage is geographically and institutionally concentrated—tied to urban centres, court locations, and media organisations—reinforcing a formalised style that prioritises location, authority, and verification over narrative depth. Notably, an absence of appearance/morality keywords (dress, dressing) indicates that sexual assault reportage backgrounds moral surveillance and overt victim blaming. However, this does not necessarily equate to greater empathy; rather, it reflects a shift from moral evaluation to *bureaucratic processing*, which risks depersonalising survivors and obscuring the gendered power relations underpinning sexual violence.

Answering RQ1 and RQ2 directly: Prioritising (RQ1) is realised through foregrounding young female child victims and the act, and backgrounding structural causes and adult/male victims; the "aboutness" (RQ2) of rape reportage is *individual moral violation*, whereas that of sexual assault reportage is *institutional-legal process*.

Stage 2: Collocation Analysis – The Function of Assuming and Implying (RQ3)

The nodes analysed were the nouns *boy, child, girl, woman*, and the adverb *allegedly*.

Collocation Profile of "Boy"

The collocation profile of *boy* reveals a framing centred on age and vulnerability. High-ranking collocates such as "old," "year," and "girl" foreground chronological youth, reinforcing representation of the victim as a minor rather than an autonomous social actor. The co-occurrence of "old" and "year" constructs age as a defining identity marker, situating the boy within a discourse of childhood dependency.

The collocation with "girl" suggests a discursive alignment of male child victims with female victims, implicitly feminising vulnerability and reinforcing rape myths that associate victimhood primarily with femininity and childhood, while excluding adult male victimhood from prominence. Institutional collocates such as "ward" reinforce the boy's dependent status, framing him as under protection rather than as an individual with agency.

Some collocates introduce sensational dimensions: "actor" and "Netflix," despite lower frequencies, suggest media-driven narratives linked to celebrities that may amplify newsworthiness while shifting focus from the violence to surrounding spectacle.

In the sexual assault corpus, the profile retains age-centring but introduces more bodily and sexualised collocates – "sexually," "assaulting," "anus" – foregrounding physical violation, alongside "his" and "mother," which introduce familial proximity and elicit moral outrage by highlighting violations within trusted spaces.

Across both corpora, *boy* constructs male child victims through overlapping discourses of age, innocence, and vulnerability. While these representations legitimise victimhood, they also reproduce restrictive assumptions that equate sexual victimisation with childhood and dependency, leaving adult male victimhood discursively constrained and engendering silences around male survivors.

Collocation Profile of "Child"

The profile of *child* across both corpora demonstrates framing rooted in legal discourse and protection. In the rape corpus, high-frequency function words (*of*: FreqLR = 246, Range = 119, Likelihood = 26.466; *a*: FreqLR = 211, Range = 90, Likelihood = 83.918) indicate embedding within formulaic noun phrases. Content words reveal the semantic prosody: *sexual* (Likelihood = 130.757, Effect = 1.996), *rape* (Likelihood = 16.881), and *abuse* (Likelihood = 266.300, Effect = 3.779) position the child within sexual offence narratives. Legal collocates – *rights* (Effect = 3.404), *protection* (Effect = 4.767), *act*, *law* – frame the child as an object of statutory concern, with high effect sizes for *protection* and *molestation* (Effect = 4.435) reinforcing a safeguarding orientation.

Age and vulnerability are emphasised through *girl* (Effect = 2.419) and *vulnerable* (Effect = 2.868), feminising and framing child victimhood through susceptibility. *Alarming* (Effect = 3.327) intensifies moral judgement, linking child sexual violence with urgency and public concern.

In the sexual assault corpus, *sexual* and *assault* replace *rape* as primary descriptors, but *abuse* remains salient (Effect = 3.799). Collocates such as *indecent* (Effect = 5.090), *treatment* (Effect = 5.015), and *neglect* (Effect = 6.215) emphasise ongoing harm and welfare conditions, while *welfare* (Effect = 4.603) and *safeguarding* (Effect = 4.810) reinforce a shift towards institutional responsibility.

Across both corpora, *child* is constructed as a passive subject requiring intervention. While this framing legitimises sympathy and protection, it marginalises the child's voice and experiential agency, positioning child victims as objects of rescue rather than narrators of harm.

Collocation Profile of "Girl"

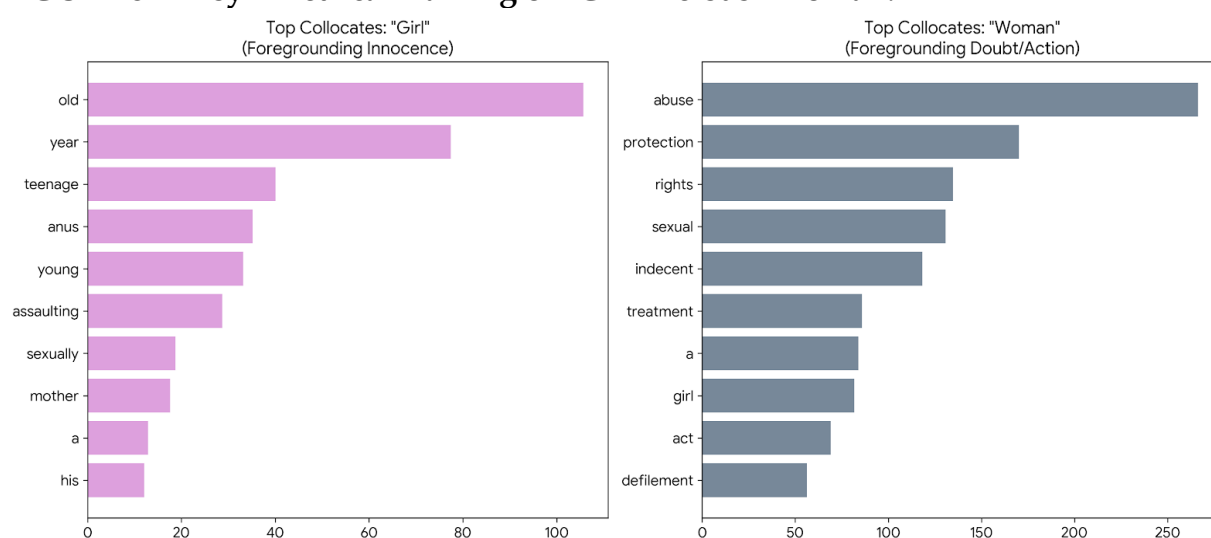
The collocations of *girl* foreground age, sexual violation, and moral innocence. In the rape corpus, *girl* is strongly framed through age specification – *old* (Likelihood = 974.139, Effect = 4.001) and *year* (Likelihood = 823.223, Effect = 3.762) – indicating age is a defining feature reinforcing vulnerability through youth. Sexual violence is made explicit through *raping* (Effect = 3.441), *raped* (Effect = 2.425), *defiling* (Effect = 3.900), and *defiled* (Effect = 3.699), positioning *girl* as the grammatical object of violent acts. Adjectival collocates – *little* (Effect = 4.306), *young*, *innocent* (Effect = 3.391), *teenage* (Effect = 4.528) – intensify fragility and moral purity. The adverb *allegedly* (Effect = 3.270) introduces epistemic distance alongside

emotive framing. Familial collocates (*father, mother*) embed the girl within domestic contexts, reinforcing betrayal narratives.

In the sexual assault corpus, the pattern is similar but with reduced lexical variety. *Old* and *year* dominate, and sexual violence is expressed through broader terms (*assault, sexually*), though *defiled, teenage*, and *little* retain high effect values.

Across both corpora, *girl* is constructed through a dense network of youth, innocence, and violation. While this invites sympathy, it reproduces a narrow victim archetype in which girls are passive and defined by vulnerability. The recurring *allegedly* complicates this by introducing epistemic caution that shifts interpretive responsibility.

FIGURE 3 – Asymmetrical framing of "Girl" versus "Woman."



A two-panel comparison. Left panel: *girl* shown with dense collocate lines to *old, year, minor, innocent, defiled, teenage* – labelled "Innocence Frame (Perfect Victim)." Right panel: *woman* shown with collocates *consent, unlawful, carnal, without, married, raped* – labelled "Procedural Frame (Juridical/Consent)." Annotation: "Girl → emphasis on being a child; Woman → emphasis on the act and legal definition."

Collocation Profile of "Woman"

The profile of *woman* shifts from innocence-based vulnerability to legality, consent, and adult sexual relations, yet remains oriented towards sexual violation with agency largely backgrounded. In the rape corpus, *woman* co-occurs with determiners such as *a* (FreqLR = 228, Range = 114, Likelihood = 284.140, Effect = 1.933) and *her* (FreqLR = 56, Range = 42, Likelihood = 36.558, Effect = 1.337), individualising women while defining them through relational reference. Age remains relevant but less central: *year* (FreqLR = 42, Likelihood = 62.366, Effect = 2.190) and *old* (FreqLR = 39, Likelihood = 58.609, Effect = 2.207) show lower values than for *girl*.

Sexual violence remains dominant: *raping* (FreqLR = 28, Range = 21, Likelihood = 84.156, Effect = 3.470), *raped* (Likelihood = 19.618, Effect = 1.816), and *assaulting* (Effect = 3.621) represent women as objects of sexual acts. However, the lower effect sizes relative to *girl* suggest a comparatively weaker moral shock value attached to adult female victimhood.

Legality and consent are prominent: *against* (Effect = 1.493), *without* (Likelihood = 30.103, Effect = 2.656), *consent* (Likelihood = 38.406, Effect = 3.264), *unlawful* (Likelihood = 30.286, Effect = 3.755), *carnal* (Effect = 3.737), *intercourse* (Effect = 2.816), and *knowledge* (Effect = 3.212) frame rape involving women as a juridical violation rather than an inherent moral outrage.

Notably, *man* (FreqLR = 26, Effect = 2.619) and *another* (Effect = 2.339) are salient, explicitly positioning the perpetrator relative to the woman—a contrast to child-focused representations where perpetrators are often backgrounded. *Married* (Effect = 3.138) suggests adult women's rape is frequently contextualised within marital or relational frames, subtly invoking discourses of sexual availability and legitimacy. *Alleged* (Effect = 1.937) introduces epistemic caution, though its lower effect value than for *girl* suggests doubt is more systematically attached to women's claims than to those involving minors.

In the sexual assault corpus, *woman's* environment becomes more abstract and generalised. Determiners (*a*: FreqLR = 51, Likelihood = 51.600, Effect = 1.711) and quantifiers (*any*: Effect = 3.549; *every*: Likelihood = 22.980, Effect = 4.112) indicate a shift from individual narratives to generalised statements about women as a category. *Sexually* (Effect = 3.013) and *raping* (Likelihood = 29.808, Effect = 4.964) retain strong associations; notably, *raping* has a higher effect size here than in the rape corpus, suggesting the act is foregrounded more starkly than the narrative context. Conversely, *and* shows a negative effect (Effect = -1.320), indicating routine narrative chaining rather than evaluative emphasis.

Across both corpora, *woman* is constructed less through innocence and more through legality, consent, and relational context. The lower likelihood and effect values for emotive descriptors indicate that adult female victimhood is normalised rather than sensationalised. The prominence of consent-based legal terminology subtly shifts interpretive responsibility onto definitions of sexual legitimacy, potentially enabling scrutiny of women's behaviour, relationships, or marital status. This pattern sustains a hierarchical model of victimhood in which girls are rendered morally pure and unquestionable victims, while women are positioned within juridical and relational frameworks that invite evaluation and doubt.

Collocation Profile of "Allegedly"

The adverb *allegedly* functions as a modal marker of epistemic uncertainty, hedging responsibility and pre-empting legal liability. Its collocation behaviour reveals how doubt is selectively distributed. **It is first necessary to acknowledge a grammatical constraint:** as an adverb, *allegedly* collocates with verbal forms by grammatical necessity – adverbs modify verbs. This must be distinguished from the *ideological* finding, which concerns *which* verbs and objects it attaches to and how strongly.

In the rape corpus, *allegedly* collocates with age markers—*old* (FreqLR = 127, Range = 71, Likelihood = 469.883, Effect = 3.985) and *year* (FreqLR = 118, Range = 73, Likelihood = 400.911, Effect = 3.755)—showing it is embedded in age-indexed constructions (e.g., "a 12-year-old girl allegedly..."). Even cases involving minors are discursively framed with procedural caution rather than moral certainty.

Sexual violence verbs show the strongest associations: *raping* (FreqLR = 65, Range = 46, Likelihood = 306.289, Effect = 4.759), *raped* (FreqLR = 68, Likelihood = 234.702, Effect =

3.808), *defiling* (Likelihood = 127.935, Effect = 5.576), and *defiled* (Likelihood = 87.927, Effect = 5.109). The very high effect sizes demonstrate that epistemic hedging is concentrated where harm is most explicit, softening the force of the allegation at the point of maximum moral gravity.

Victim nouns show strong collocation: *girl* (FreqLR = 46, Range = 36, Likelihood = 127.608, Effect = 3.282), *daughter* (Effect = 2.848), *minor* (Effect = 3.030), *teenager* (Effect = 3.732), *pupil* (Effect = 4.192), and *lady* (Effect = 3.677), indicating *allegedly* frequently accompanies references to female or child victims, surrounding victimhood itself with uncertainty. Function words (*the*: Effect = -0.423; *to*: Effect = -1.048; *of*: Effect = -1.272) show negative effect values, confirming structural but not ideological significance.

More revealing, *rape* appears with a **negative effect** (Effect = -2.240). While partly a grammatical artefact (an adverb binds to inflected verb forms rather than the bare noun), the ideological point is that hedging attaches at the level of the *individual alleged act* (raped, raping) rather than the *categorised crime* (rape). Where the noun form dominates, the offence is treated as a stable legal fact; where the verb form dominates, it is an individual, contestable act.

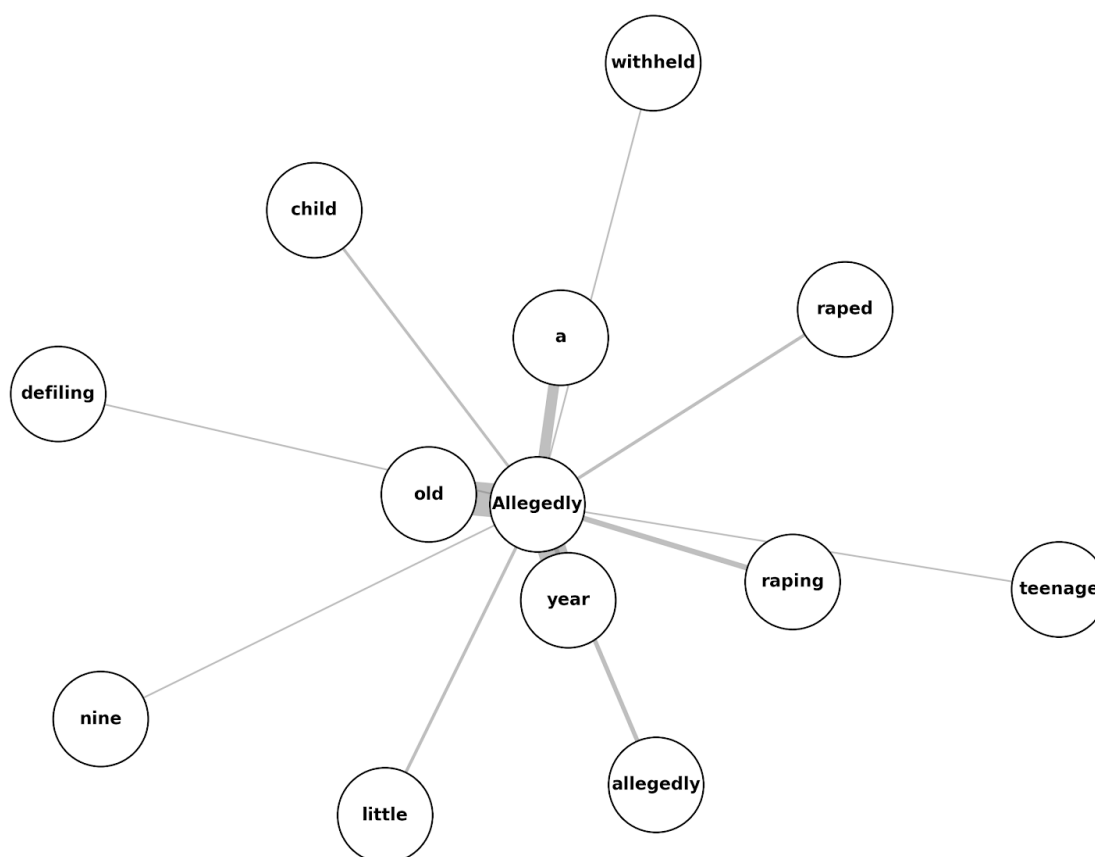
Perpetrator collocates – *man* (FreqLR = 24, Effect = 2.579), *suspect* (FreqLR = 16, Effect = 2.819), *neighbour* (Effect = 4.079), *gang* (Effect = 3.704) – appear less frequently than victim-focused ones, indicating epistemic caution is concentrated around the act and the victim rather than the perpetrator. Institutional collocates – *arrested* (FreqLR = 16, Effect = 2.766), *remanded* (Effect = 3.120), *remands* (Effect = 4.637), *correctional* (Effect = 4.137), *centre* (Effect = 2.510) – show that *allegedly* persists even within criminal justice contexts, suggesting legal process does not displace hedging but coexists with it.

On the sexual assault corpus comparison: *allegedly* does not register as a statistically distinctive collocate when the sexual assault corpus is treated as target. This is interpreted with caution. In that corpus, the related adjective *alleged* is common within legal-bureaucratic phraseology (e.g., "alleged assault," "alleged offence"), meaning the modal is dispersed across institutional contexts rather than concentrated on specific victim-act pairs. The interpretive claim is therefore not simply "sexual assault discourse uses less hedging" – which, given overall frequencies, may be an artefact of normalisation – but rather that *in the rape corpus the hedging binds with exceptional strength to vulnerable victims and explicit violent acts*, whereas in the sexual assault corpus it is diffused across legal-formulaic uses.

Overall, *allegedly* functions as a discursive buffer, with its strongest ties to age markers, explicit sexual acts, and female or child victims. The high likelihood and effect values for *raping*, *defiling*, and *old/year* indicate this pattern is structurally embedded rather than incidental. The routine use of *allegedly* contributes to the reproduction of rape myths – the notion that rape claims are inherently disputable and require constant qualification. Even in cases involving minors or familial perpetrators, the adverb sustains a climate of doubt that may symbolically undermine victim credibility. Read alongside *claimed* and *assault*, *allegedly* completes an evaluative continuum in which rape reportage moves from uncertainty (*allegedly*), through reported speech (*claimed*), to legal categorisation (*assault*).

FIGURE 4 – Network map of "allegedly."

Network Analysis of "Allegedly": The Grammar of Epistemic Doubt



A network/force-directed diagram with *allegedly* at centre, connected by weighted edges (edge thickness = log-likelihood) to *raped*, *defiling*, *defiled*, *old*, *year*, *girl*, *minor*, *pupil*, *remanded*, *arrested*. Dense, thick edges connect *allegedly* to the sexual-violence verbs and age/victim nodes; thin edges to institutional nodes. Annotation: "Epistemic shifting: media transforms a factual event into a negotiable claim. Strongest links = most serious harms."

Answering RQ3: Collocation profiles engender the function of Assuming (RQ3) by bundling victim terms with presupposed meanings – innocence and vulnerability (for *girl*, *child*, *boy*) versus consent and legality (for *woman*) – and by attaching epistemic uncertainty to the most vulnerable victims, thereby presupposing that rape claims are inherently contestable rather than established facts.

Stage 3: Qualitative Synthesis

The quantitative findings were synthesised through the combined lenses of FCDA and Framing Theory. This synthesis reveals that the linguistic patterns identified via Corpus Linguistics are not merely stylistic tendencies but function as discursive mechanisms that reproduce patriarchal hierarchies and sustain epistemic inequalities in representations of sexual violence.

From an FCDA perspective, the data demonstrate that media discourse systematically privileges certain victim identities, reinforcing gendered power relations. The repeated foregrounding of *girl* and *child*, alongside collocational patterns emphasising innocence, vulnerability, and moral purity, constructs the "Perfect Victim": young, female, passive, and deserving of sympathy. Conversely, *woman* is mediated through legalistic and relational frameworks centred on consent, marital status, and sexual legitimacy, positioning adult female victims within a discursive space that invites scrutiny, doubt, and conditional recognition. Male victimhood, particularly beyond childhood, is marginalised through limited lexical and collocational visibility, reinforcing heteronormative assumptions about vulnerability.

Framing Theory further elucidates how these patterns operate through selection and salience. Rape reportage is characterised by **episodic framing**, foregrounding individual cases, emotive narratives, and moral evaluations while backgrounding structural explanations such as patriarchy, rape culture, and systemic inequality. In contrast, sexual assault reportage adopts a more **thematic and institutional framing**, emphasising legal process, policy discourse, and elite actors. While this reduces overt moral judgement, it simultaneously abstracts and depersonalises survivor experiences, repositioning sexual violence as a bureaucratic issue rather than a lived social reality.

The collocation analysis of *allegedly* reveals that epistemic uncertainty is unevenly distributed. Its strong association with explicit sexual acts and vulnerable victim categories demonstrates that linguistic caution is disproportionately applied at points of maximum harm, contributing to a discursive environment in which victim credibility is implicitly destabilised.

A note on interpretive scope. The corpus method demonstrates *patterns of representation and meaning-potential*—the lexical architecture that *permits, encourages, or is consistent with* particular readings. It cannot by itself establish that readers actually internalise these framings or that media discourse "causes" rape myths. The claims made here are therefore framed as *discursive affordances*: Nigerian newspaper language is *consistent with the reproduction of rape myths* and is *conducive to* sustaining scepticism, rather than as proven effects on audience cognition. This is the standard evidentiary boundary of corpus-assisted critical stylistics.

Table 1. Research Questions and Where They Are Answered

Research Question	Addressed in	Key Finding
RQ1: How do keywords establish the function of Prioritising?	Section 4.1.1	Rape reportage foregrounds young female child victims and the act itself; backgrounds structural causes and adult/male victims.
RQ2: How do keywords help identify "aboutness" reportage?	the Section of 4.1.2	"Aboutness" of rape = individual moral violation; of sexual assault = institutional-legal process and elite actors.

Research Question	Addressed in	Key Finding
RQ3: How do collocation profiles engender the function of Assuming?	Section 4.2	Victim terms are bundled with presupposed meanings (innocence for girl/child; consent for woman); <i>allegedly</i> attaches epistemic doubt to the most serious harms and most vulnerable victims.

Conclusion

This study set out to investigate how linguistic choices in Nigerian newspaper reportage construct and sustain hierarchical representations of rape victims. Through a Corpus-Assisted Critical Stylistics approach, integrating quantitative corpus analysis with qualitative ideological critique, the study demonstrates that media language is not a neutral medium of representation but an active site of meaning-making that shapes public understanding of sexual violence.

The findings reveal that rape reportage prioritises explicit naming of the act, emotive and experiential narratives, and moral evaluations of victims, while sexual assault reportage foregrounds institutional, legal, and political discourse. Across both corpora, linguistic patterns systematically construct a "Perfect Victim" archetype centred on youth, innocence, and femininity, while adult women and male victims are positioned within frameworks that invite scrutiny, abstraction, or marginalisation.

Furthermore, the study shows that epistemic markers such as *allegedly* function not merely as legal safeguards but as ideological tools that distribute doubt unevenly, often undermining victim credibility – binding most strongly to explicit sexual acts and to the most vulnerable victim categories. The absence of structural terminology relating to patriarchy, rape culture, and systemic inequality indicates a broader tendency towards episodic framing that individualises sexual violence and obscures its socio-cultural roots.

Conclusively, the study demonstrates that Nigerian media discourse plays a significant role in reproducing rape myths, sustaining gendered hierarchies of victimhood, and shaping the boundaries of empathy and legitimacy in public narratives of sexual violence. Media discourse is *consistent with and conducive to* these outcomes – a discursive contribution best addressed through deliberate, ethically responsible reporting.

Limitations

Several limitations contextualise the findings:

1. **Corpus scope.** The six newspapers, five headquartered in Lagos, skew towards urban, southern, English-language journalism. Findings may not generalise to regional, vernacular, or broadcast media.
2. **Political-event confounds.** The sampling window (2017–2025) includes specific high-profile controversies (e.g., the Akpoti-Uduaghan Senate case). Person-name and political keywords partly reflect single-case saturation rather than stable discourse patterns, as discussed in Section 4.1.2.

3. **Category asymmetry.** As noted in Section 3.1, "rape" and "sexual assault" are not fully parallel categories; findings are relative to this comparison.
4. **Genre-relative keyness.** Because the sub-corpora are swapped as target/reference, all keyness and collocation values are *relative to the paired corpus*, not to general English or to a broader Nigerian media norm. A keyword salient only "in rape vs. sexual assault" might not be salient against a general-reference corpus.
5. **Interpretive boundary.** Corpus methods identify patterns of representation and meaning-potential, not reader effects (see Section 4.3). Claims about the social "reproduction" of rape myths are critical interpretations grounded in these patterns, not direct measurements of audience influence.

Contribution to Knowledge

This study:

1. Demonstrates how linguistic features such as keywords and collocations function as discursive tools for constructing hierarchical victim identities in media discourse.
2. Establishes Corpus-Assisted Critical Stylistics as a systematic methodological framework—integrating quantitative keyness/collocation measures with critical stylistic, feminist, and framing-theoretic interpretation—for uncovering the ideological implications of language use in representations of sexual violence.
3. Advances a data-driven account of the "Perfect Victim" construct and of *epistemic hedging* (*allegedly*) as an unevenly distributed ideological mechanism in African media.

Recommendations

1. **Media practitioners** should adopt linguistic choices that minimise epistemic hedging and avoid evaluative lexis that implicitly questions victim credibility—especially in cases involving minors and in descriptions of the most serious harms, where *allegedly* is currently concentrated.
2. **Reporting** should incorporate more inclusive and structurally oriented language that foregrounds systemic factors rather than individualising or depersonalising sexual violence narratives, and that recognises adult female and male survivors as legitimate victims.
3. **Editorial policy** should reflect an awareness of the victim hierarchy documented here, actively countering the "Perfect Victim" bias by reporting on the full range of victim identities with equal seriousness.
4. **Future research** should extend the framework to regional and vernacular media, incorporate readership/reception studies to test the interpretive claims about audience effects, and apply the CACS toolkit to other genres (e.g., broadcast, social media) and other sites of gendered violence.

References

- Akano, R., & Afolabi, A. (2024). Extending the rape narrative: A corpus-assisted discourse analysis of the construction of rape in selected Nigerian dailies. *Legon Journal of the Humanities*, 35(1), 164–191. <https://doi.org/10.4314/ljh.v35i1.5>
- Al-Rawi, M. K. (2017). Using AntConc, a corpus-based tool, to investigate and analyse the keywords in Dickens' novel *A Tale of Two Cities*. *International Journal of Advanced Research*, 5(3), 366–372.
- Andrew, A., & Alexopoulos, C. (2018). Framing blame in sexual assault: An analysis of attribution in news stories about sexual assault. *Violence Against Women*, 25(2), 1–20. <https://doi.org/10.1177/1077801218801111>
- Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *Proceedings of IWLeL 2004: An Interactive Workshop on Language E-Learning*, 7–13.
- Barber, K. (2022). *(Re)framing rape: A sociocognitive discourse analysis of sexual violence at the intersection of white and male supremacy* [Doctoral dissertation]. University of Hawai'i at Mānoa.
- Christie, N. (1986). The ideal victim. In E. A. Fattah (Ed.), *From crime policy to victim policy: Reorienting the justice system* (pp. 17–30). Macmillan.
- Dhawan, V. (2018). *(De)constructing the "perfect rape victim": An analysis of sexual assault and survivors discourses in the Canadian criminal justice system* [Master's thesis]. University of Waterloo.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Falohun, B., & Imam, M. (2019). Discourse constructions of gender-based crimes in newspaper reports. *Electronic Research Journal of Social Sciences and Humanities*, 1(4), 97–113.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Arnold.
- Jeffries, L. (2010). *Critical stylistics: The power of English*. Palgrave Macmillan.
- Lazar, M. M. (2014). Feminist critical discourse analysis: Relevance for current gender and language research. In M. Ehrlich, S. Meyerhoff, & J. Holmes (Eds.), *The handbook of language, gender and sexuality* (2nd ed., pp. 180–200). Wiley-Blackwell.
- Muchnik-Rozanov, Y., & Tsybulsky, D. (2021). Linguistic analysis of science teachers' narratives using AntConc software. *Technion – Israel Institute of Technology*.

-
- Oludayo, T., & Udechukwu, C. (2020). Characterizing rapists and their victims in select Nigerian newspapers. In *Societal and historical dimensions of sexual violence* (pp. 153–176). IntechOpen. <https://doi.org/10.5772/intechopen.91705>
- Osisanwo, A., & Ojo, R. (2022). "They are sexual objects": Construction of female rape victims in selected Nigerian newspapers. *Mediální studia*, 16(1), 92–117.
- Simonsen, S. (2024). Discursive blame attribution strategies in migration news frames: How blame for perceived migration-related problems is mediated in journalistic framing. *Discourse, Context & Media*, 57, 1–8.
- Tabbert, U. (2013). *Crime through a corpus: The linguistic construction of offenders, victims and crimes in the German and UK press* [Doctoral dissertation]. University of Huddersfield.
- Turkewitz, R. R. (2010). *All the news that's fit to print? A content analysis of newspapers' portrayal of rape and sexual assault* [Doctoral dissertation]. [Institution].
- Ujevic, D. (2015). *(Re)presentations of sexual violence against women: An analysis of media reports of rape* [Doctoral dissertation]. [Institution].
- van Hulst, M., Metze, T., Dewulf, A., de Vries, J., van Bommel, S., & van Ostaijen, M. (2024). Discourse, framing and narrative: Three ways of doing critical, interpretive policy analysis. *Critical Policy Studies*, 1–20. <https://doi.org/10.1080/19460171.2024.2326936>

Publisher's Note/Disclaimer: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of NJH. NJH disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.