

<https://doi.org/10.5281/zenodo.20555570>

Corpus-Based Analysis of Power Dynamics in V for Vendetta



Iqra Imran	University of Management and Technology, Sialkot iqraimrann1@gmail.com
Aamna Waheed	aamnawaheed45@gmail.com
Maria Naseem	University of Management and Technology UMT, Sialkot maria.naseem@skt.umat.edu.pk

Abstract: *The current research used the corpus linguistics approach to analyze the discourse of power in Alan Moore's graphic novel V for Vendetta. Although the novel has been studied in depth in qualitative scholarship, no previous study has engaged the entire verbal text of the novel in systematic and data-driven corpus analysis. A single-text corpus consisting of the full verbal text of the novel was created in AntConc (v4.3.1) for the purpose of producing word frequency distributions, lines of text that show the focus word or words in context (KWIC), collocation profiles, and n-gram cluster frequencies. Word frequency analysis indicated that the most common content words were character names (V: f=518, Evey: f=340, Finch: f=253, Leader: f=216). The prevalence of characters who resist over the institutional power indicated by the Leader who authored the novel allows for the suggestion that discursiveness in a novel such as this favored resistance over institutional power. A KWIC concordance of the three terms revealed three different registers of power: ideological proclamation, coercive enforcement, and moral-epistemic critique, respectively. The narrative's implicit depiction of the instability of institutional authority was made apparent with the change of Leader's concordance profile from imperative commands to supplicatory. Denial emerged as the four word pattern most frequently mentioned (I didn't do, f=5), as well as the clusters of directives (I want you to, f=3) and the definitional sequences (a man who, f=4). The results indicated that power in V for Vendetta was not only multi-layered but also contestatory. The study set up a repeatable corpus methodology for the verbal analysis of graphic narratives and provided an empirical complement to the existing qualitative interpretation of the text.*

Keywords: *corpus linguistics, V for Vendetta, frequency analysis, concordance, KWIC, n-grams, lexical clusters*

1. Introduction

Language is not a neutral means of communication. It is used in literary and political discourse as a tool to construct social hierarchies, to justify authority, and to facilitate or hinder resistance (Partington, 1998). Although the language/power dynamic has been a significant focus of discourse research, the

empirical and quantitative analysis of literary texts is relatively less studied than their qualitative close reading traditions (Crosthwaite et al., 2023).

V for Vendetta, a graphic novel by Alan Moore and David Lloyd, is a text that is extremely political, where characters use language as a means to govern, surveil, and rebel against a

dystopian Britain in the future. The novel has a rich set of data for corpus-based linguistic investigation because of its verbal dimension, which includes character speech and narrative description. Although the novel is well studied in the political sphere, no previous work has analyzed the verbal text systematically in the context of a corpus.

This study directly tackles that question. The research adopts a manually transcribed corpus of the novel's full verbal text and employs the AntConc software, n-gram cluster extraction and frequency analysis to delineate the lexical patterns in which the power is realised in the text. It is not an interpretative goal in the first place, but a goal of description, that is, to determine what information the quantitative data on word distribution, concordance patterning, and cluster frequencies contains about the discursive structure of power in the novel.

The article is organized as follows. The relevant literature from the field of corpus linguistics, literary corpus stylistics, and even from preexisting research on *V for Vendetta* is covered in section 2. In Section 3, the research gap is identified and the single research question and objective for the study is presented. The theoretical framework for the analytical design is presented in Section 4. Section 5 describes the methodology used, such as corpus construction, processing and analysis. The results are organized by tool and output type in Section 6. Section 7 discusses the implications of the corpus evidence. The study is summarized and recommendations for future research are presented in Section 8.

2. Literature Review

2.1 Corpus Linguistics in Literary Studies

Since the 1990s, there has been a significant increase in the use of corpus methods for literary corpora, as computational tools became available and as corpus stylistics as a sub-discipline emerged (Leech & Short, 2007; Mahlberg, 2013). Frequency, collocation, and distributional patterning are the features of Corpus stylistics that treat them as evidence of literary and rhetorical structure, as a step to getting away from impressionistic reading to

reproducible, data-driven results (Biber et al., 2002). One of the basic claims of corpus stylistics is that patterns that cannot be seen in individual texts can be detected when the samples are quantified and aggregated.

Crosthwaite, Ningrum, and Schweinberger (2023) record the approximate rate of growth of corpus-assisted literary studies since 2010 of about 40%, with special reference to the use of tools like AntConc for identifying shifts in lines of power and representation in language. The authors highlight the methodological productivity of distant, computational reading, and their claim is that this combination of distant reading and close reading brings empirical rigour to the interpretive frameworks.

The corpus stylistic research into Dickens' style by Mahlberg (2013) illustrates the function and ideological positioning of lexical patterns and cluster frequencies within the Dickensian canon. Likewise, Wan (2018) used AntConc to examine the lexical manifestation of trauma and power in modern fiction in his corpus-based study of a book called *A Thousand Splendid Suns*, paving the way for applying corpus techniques to politically oriented literary corpora. This position is further supported by Partington's (1998) study of patterns and meanings in corpus data, which elaborates on the theoretical framework whereby high-frequency co-occurrence and clusters can be seen as semantically relevant structures.

2.2 Corpus Tools for Discursive Analysis

Laurence Anthony of Waseda University created AntConc, which is one of the most popular corpus analysis programs used in linguistic and literary research. The main applications it can be used for are: generating word lists according to frequency, displaying concordance lines for specific search terms, sorting concordance lines by the searched word, extracting n-grams and clusters, and collocation profiling. Baker (2006) provides a theoretical foundation and methodology for the use of these tools for discourse analysis, stating that repetitive language is not random but rather reflects the ideological orientations of a text or text set.

The article "Emerging corpus-based approaches in literary research" (Lingvospektr 2024) shows how the corpus analysis of *V for Vendetta* and Orwell's 1984 shows common patterns in the use of lexical items, suggesting deliberate intertextual recontextualization of dystopian vocabulary. In that study, collocation analysis confirmed that there were semantic shifts which were related to narrative context and character ideology, which was useful for literary thematic research using quantitative profiling.

Using a corpus-based intersectional approach, Sharma and Haider (2024) explore the intersectional nature of the power relations in modern Anglophone fiction works through the use of repeated lexical metaphors of confinement that encode racial and gendered hierarchies. Literary texts, they say, are linguistic battlegrounds where marginalized voices question the dominant ideology in recognizable stylistic and lexical patterns; a point that is directly relevant to the discourse architecture of *V for Vendetta*.

2.3 V for Vendetta: Existing Scholarship

Scholarly interest in *V for Vendetta* is broad, but technique is restricted to qualitative methods. As a political performance text, the novel is explored by Di Liddo (2009) who looks at Moore's narrative techniques to see how layers of ideology are added. Lewandowski (2020) treats the novel as a polyphonic political manifesto, and makes the point that the novel's relationship with *The Count of Monte Cristo* and Orwell's 1984 serve as discursive resources in critiquing the nature of authoritarianism.

In the novel, Munywoki (2021) theorizes on the fragmentation of postmodern narratives and the supremacy of images, suggesting that the novel is a reflection of the disorientation of the subjects in totalitarian regimes, for such timelines are fragmented and overlapping, while the dialogue is fragmented. MacFarlane (2019) examines gender, power and language, and she shows how the novel's dialogue patterns capture *V*'s ambiguity and Evey's evolution from vulnerability to agency. Zahoor et al (2020) carry out a corpus-assisted study of the discourse patterns of gendered power in *V for Vendetta*

focusing on the aspect of interruption, silencing and assertion in the speech of characters.

The literature lacks a comprehensive, word frequency-based, full KWIC concordance and n-gram cluster extraction study of the complete verbal text of *V for Vendetta*, with an explicit focus on the quantitative structure of power discourse. The present study attempts to fill this gap in the empirical evidence.

3. Research Gap, Research Question, and Research Objective

3.1 Research Gap

Although there has been a plethora of work on *V for Vendetta* and an increasing amount of work on using corpus techniques in literary studies, there is no earlier study that has analysed the full verbal text of Moore's novel through a systematic corpus-determined word list frequency, KWIC concordance and n-gram cluster extraction. Current corpus-based research on the novel is restricted in its focus and scope, with only selective features or small data sets. Baker and Egbert (2016) have spoken about the possibilities of discovering patterns in the language that are not apparent through the conventional reading, but they have not been fully realised in the context of *V for Vendetta*. In the current study, this is addressed by the first complete corpus-based study of the novel's verbal text, which gives an empirical linguistic description of the novel's discourse of power.

3.2 Research Question

What do the word frequency distributions, KWIC concordance patterns and n-gram cluster frequencies show about the lexical and structural organization of power discourse in *V for Vendetta*?

3.3 Research Objective

To find, understand and explore concordance and cluster patterns of occurrences of power-related language for *V for Vendetta*, to provide an empirical, data-driven characterization of the encoding of discursive power within verbal text.

4. Theoretical Framework

The analytical design used in this study is corpus linguistic, based on the discourse-theoretic

theory that language is a communicative vehicle that is not only a structure but also a mechanism by which power relations are constructed, normalized and challenged. This premise, which has been linked to Foucauldian discourse theory, and which is operationalised in the corpus stylistic tradition, creates the interpretative scaffolding in which quantitative corpus findings are interpreted.

A basic theory is that discourse—defined as a system of language use with specific knowledge claims and institutional practices—is both a medium and an expression of power (Foucault, 1995). Within this structure, language is not only a reflection of power relations, but also an active representation of power relations. Power is linguistically produced and reproduced in the contexts of command, repeated authority-asserting phrases, and the lexical environments of authority figures.

Critically, it is not applied to this theoretical framework in the form of qualitative thematic exegesis but rather as a guideline for the most analytically relevant corpus measures. The theoretical reason for frequency analysis is that the relative prominence of specific words and phrases in a text corresponds to the discursive priorities of the text. The justification for conducting concordance analysis is based on the fact that the common linguistic contexts of a word, its collocational profile, are considered evidence of its ideological charge. The use of n-gram and cluster analysis is justified because it is argued that recurrent multi-word sequences are more useful in capturing the structure of ideology than word frequencies are (Mahlberg, 2013; Baker, 2006).

Since this framework is not a predictive one, the selection of corpus measures and their interpretation is methodological in nature; the results, however, are based on quantitative distributional evidence, not on interpretive inference.

5. Methodology

5.1 Corpus Description

The corpus is the full verbal text of the graphic novel *V for Vendetta*, by Alan Moore and David

Lloyd (DC Comics, 1990; Vertigo, 2005). The verbal dimension (all written dialogue and narrative prose description) of the novel was manually extracted and transcribed, as the original publication format contained both written and visual components. The dataset excluded non-linguistic patterning aspects, such as visual elements like panel illustrations, layouts, and purely symbolic content, because the focus of the study is on linguistic patterning, and AntConc is designed to process plain text.

The transcription has been completed and consistent, with every character's speech, internal thoughts and story descriptions recorded. The text has been transcribed and presented in a plain text (.txt) file, encoded in UTF-8. The resulting single text corpus is the entire verbal aspect of the novel and is the basis of all the corpus measures reported in this study.

Unlike other multi-text corpora, *V for Vendetta* is a single-text corpus and does not allow traditional cross-corpus keyword analysis with the help of a reference corpus comparison. Accordingly, all results of frequency based studies are presented in the text in absolute and relative numbers, as is customary in single text corpus stylistic analysis (Mahlberg, 2013; Baker, 2006).

5.2 Data Collection Procedures

The primary data collection consisted of manually transcribing all verbal content of *V for Vendetta*. This procedure was selected because no machine-readable plain text version of the verbal material in the novel was found for direct input to the corpus software. The transcription was conducted systematically: the text of each panel was transcribed in reading order, speaker identification was transposed in brackets (for ease in concordance interpretation) but not in the plain-text file (to avoid overcounting words for the speaker identification markers).

A sample of the pages was re-read with the transcribed text to make sure the transcription was accurate and errors were corrected, a process known as consistency check. This review was then saved as a .txt file and subsequently loaded into AntConc version 4.3.1 for the rest of the corpus operations.

5.3 Corpus Processing Procedures

The corpus was processed in AntConc with four main tools. First the Word List tool was used to create a ranked list of all the tokens in the corpus from which the most frequently occurring 20 tokens were selected and tabulated. Second, three chosen keywords (Leader, Creedy and Finch) were used in the Concordance (KWIC) tool. The terms were chosen because of their high usage in the word list and the significance of their structural meaning as authority-related character names. The complete concordance lines were checked and representative instances were chosen for each keyword to show the most common co-occurrence patterns.

Third, three word and four word Clusters/N-Grams were extracted from the corpus using the minimum frequency threshold of 3 to ensure that the reported patterns were not singular collocations. A cluster list was formed and then sorted by frequency. Fourth, the collocation profiling method was analyzed in terms of prominent keywords, and the semantically significant co-occurring terms within a five-word window on either side of each node word were identified.

5.4 Analytical Framework

The analytical process involved three steps. In the first step the frequency data obtained from the Word List were tabulated and analyzed to determine what content words of structural importance among the high frequency forms. The second stage involved analyzing the KWIC concordance lines of selected keywords to determine the common linguistic context in which authority-related terms occur. The third stage involved analyzing the n-gram and cluster data to identify multi-word sequences of analytical significance in relation to the study's research question. Results obtained from both (and in the third stage all three) were brought together in the Discussion section. All analysis was done in AntConc 4.3.1 and no other software was used for corpus manipulation.

6. Discussion

The corpus evidence found in Section 6 reveals the common sense of the structure of power

discourse within the context of quantitative linguistic patterning. There are three major analytical claims that emerge from the data.

First, the frequency distribution of character names in the word list confirms that resistance has been the organizing principle in the novel's verbal text, if not more so than authority. V (f=518) is more than twice as common as Leader (f=216), while the total for resistance-associated character names (V: 518, Evey: 340, Finch: 253) far outweighs the total for authority-associated character names (Leader: 216). This asymmetry is not a minor detail, however: A text that institutionalizes its topic as a system of power has a primary discursive attitude of resistance over compliance, which is reflected in the scarcity or abundance of the name of the protagonist, the resistant. By this criterion, the V for Vendetta corpus signifies a discursive perspective that foregrounds the resistance (Baker 2006).

Second, by comparing the Leader, Creedy, and Finch concordance, it can be seen that there is a stratified discourse architecture whereby different registers of power (declarative ideology, coercive enforcement, and internal critique) are deployed across distinct environments of the Leader, Creedy, and Finch's concordance. This is not the same systematic process for close reading: traditional close reading may allow for the identification of individual quotation passages, but the collocational profile for each keyword, generated across all concordance lines, allows for a more holistic characterization of the function in which each character is used. The change of the Leader's concordance profile from commanding to supplicating is especially telling: it shows distributional evidence of the power narrative of the whole corpus, from assertion to collapse, recorded by one word.

Third, the n-gram and cluster data show that the most common multi-word sequence in the corpus is a denial pattern, I didn't do, f=5; the next most common four-word clusters are directive sequences, that's why you, I want you to, and locative-prepositional frames, in front of the. The power idea is almost absent from the novel (the dominant four-word cluster is denial),

which implies that discourse in the novel is less focused on the idea of power being asserted than the idea of power being resisted: characters refuse to acknowledge actions, responsibilities, and place. This is typical of a text where authority is frequently validated and challenged, and guilt and complicity are significant discursive themes.

Complementary is added with the three-word cluster data. The frequency of 'a man who' and 'be rid of' points to a discourse that is preoccupied with defining and categorizing persons, and gesturing towards rejection and exclusion. These patterns reveal that the verbal text of V for Vendetta has a logic of discursivity, in which power is acted and contested through the language of identity attribution and disavowal.

The findings support and add empirical evidence to the claims made in previous qualitative literature on V for Vendetta. The quantitative evidence of multiple stratified discourse registers in the character concordance profiles, as detailed by Lewandowski (2020), backs up Lewandowski's argument that the novel reads as a polyphonic political manifesto. Munywoki's (2021) argument that fragmentation and disorientation are textually encoded is consistent with the cluster evidence of denial and epistemic uncertainty. The present study complements the study of gendered power patterns of Zahoor et al. (2020) by providing a more comprehensive word list and cluster analysis, creating a wider distributional frame of reference in which gendered patterns are embedded.

7. Results and Findings

7.1 Frequency Analysis: Word List

The Word List tool in AntConc produced a ranked list of all the tokens in the V for Vendetta

corpus. The top 20 most frequent words are shown in Table 1. As typical in English language corpora, high frequency function words (the, a, of, to, and, is, in, that), generally associated with the structure of English grammar, appear in the upper ranks of the list, not the text-specific content.

The content words in the top 20 are of greater analytical value. The novel's verbal text is strongly dialogic, as I (538), along with you (538), are tied for rank 4. These second and first person pronouns occur frequently, indicating a discourse with direct address, interpersonal confrontation, and relational negotiation, all of which are related to a text organized around the relationship between people and institutions and a focus on power relations.

The named entity V is the most common in the corpus, ranking at position 6 (f=518). Evey appears at rank 12 (f=340), Finch at rank 17 (f=253), and Leader at rank 19 (f=216). These four character names suggest a structural organisation of the novel's power discourse directly. V has a higher frequency than Leader (518:216) which is notable, especially since Leader is the most institutional figure in the fictional world, but the novel's discursive centre of gravity is not with the insistence of orders; it's with the insistence of resistance.

The function word he (f=315, rank 13) and the possessive/pronoun her (f=284, rank 15) are also among the top 20, reflecting continued effort to use third-person pronouns, to mark the characteristics of characters across the story, and to track character action and states throughout the narrative. The contracted verb forms (is, has) which are separated from the preceding words, such as the rank 10 (f=371) token 's', are artifacts of transcription.

Table 1: Top 20 High-Frequency Words in the V for Vendetta Corpus

Rank	Word	Frequency
1	the	1463
2	a	634
3	i	538
4	you	538
5	of	530
6	v	518
7	to	493
8	and	448
9	is	378
10	s	371
11	it	347
12	evey	340
13	he	315
14	in	298
15	her	284
16	that	282
17	finch	253
18	his	237
19	leader	216
20	she	172

7.2 Concordance and KWIC Analysis

7.2.1 KWIC Analysis: Leader

The keyword Leader is consistently co-occurring with forms of language that are commanding, declarative and coercive in a KWIC concordance. The word Leader is used across the concordance lines in contexts where there are modal imperatives, performative declarations, and direct orders. Other common phrases heard representatives say are ‘*Do I make myself clear?*’, ‘*Shot on sight!*’, and ‘We must remind them that this is a country of values, of order, of strength. These patterns create a

discursive node of authority assertion and institutional control in the name of Leader.

Perhaps most importantly, the concordance lines also record instances in which the Leader's linguistic register changes into supplication and despair. The phrase ‘*Don't abandon me!*’ also occurs in a concordance line near the end of the corpus, which is a clear contrast to the imperative-dominant patterns of previous occurrences of the Leader. The distribution of such a change – from commanding to pleading – in the concordance data is corpus-level evidence of the narrative's representation of institutional

power as unstable and contingent.

In the collocational environment of the node word Leader, there is a group of semantically related words which include the words from the semantic field of national authority and collective identity such as nation, order, values, strength, unity, and obedience, that appear in a five-word window around the node word. The structure and language found in this collocational profile are in line with the discourse of authoritarian nationalism and the identification of Leader's speech as a register of "ideological prescription".

7.2.2 KWIC Analysis: Creedy

Concordance lines for Creedy show a clear, but complementary, pattern. The co-occurring forms in Creedy's concordance profile are more often imperative and action-directive than declarative and ideological as in Leader's. The phrase 'Don't just stand there! The modal imperatives and urgent commands that occur in Creedy's linguistic environment are represented by the words 'Follow them!'.

Terms associated with enforcement, compliance and institutional hierarchy are also concentrated, as in the Creedy concordance, follow, orders, stand, move, and there occur with greater frequency. This results in Creedy's language being seen as an operational node of coercive-power, whereas Leader's language is seen as an ideological and declarative one. The two character keywords are linked together to create a double discourse of power: legitimation on the top level (Leader); coercion on the operational level (Creedy).

7.2.3 KWIC Analysis: Finch

The concordance profile of Finch is the most analytically distinct of the three keywords studied. Finch's KWIC lines are characterized by introspective, epistemic, and morally evaluative language, whereas the other two, Leader and Creedy, are command-oriented or declaration-oriented. The co-occurrence of these words, such as feel, think, know, something, wrong, and nervous, gives rise to a collocational profile that is characterised as uncertain, interior, and ethical-reflective.

Two of the more important lines of concordance display this. Firstly, Finch's role in 'Government-directed terrorism never sows healthy ideas into the public' aligns him with internal critique of the regime message and positions him as an epistemological space of questioning the regime message, rather than making a claim to it. Second, *'I feel their eyes. Even in this silence'* is the internalisation of surveillance as a psychological state and Finch's language codes the subject experience of being watched. Finch's use of his first-person epistemic verbs (feel, think, know) is noticeably more frequent than in the Leader or Creedy profiles, suggesting he is a focus of interiority and conscience in the novel's discourse system.

7.3 N-Gram and Cluster Analysis.

For this, the Clusters/N-Grams tool in AntConc was used to generate three-word and four-word combinations that appeared at least three times in the corpus. Two of the most common clusters are shown in Table 2, in each of these categories.

Table 2: Most Frequent 3-Word and 4-Word Clusters in the V for Vendetta Corpus

Three-Word Cluster	Freq.	Four-Word Cluster	Freq.
a man who	4	i didn't do	5
a violent carson	3	that's why you	4
be rid of	3	in front of the	4
a series of	3	evey i don't know	4
across the city	3	i want you to	3

Among the three-word clusters, 'a man who' (f=4) is the highest-frequency nominal cluster. It serves mainly introductory and characterizing purposes, usually introducing attributive clauses, which specify characters by action or value or by identity traits. Subsequent constructions contribute to a discourse that sees people defined and classified, a theme that is explored throughout the novel.

The cluster 'be rid of' (f=3) is rejection and the desire for elimination. Recurrence as a cluster indicates discursive distancing and opposition, where characters or institutions try to get rid of or exclude threatening elements. The locative cluster 'across the city' (f=3) creates surveillance-inflected space awareness, with its repetition the creation of a 'watched, constrained' city space in the spatial discourse of the text. The frame 'a series of' (f=3) functions as a quantifying frame, usually preceding descriptions of events, actions or violations, and creates an institutional documentation and reportage register within the novel.

The four-word cluster most common in the cluster list overall is 'I didn't do' (f=5). It is analytically significant because it is a very frequent four-word pattern in the corpus, that is, the discursive logic of the text is organized around acts of denial, self-distancing from responsibility, and defensive assertion. This cluster can be found in emotionally charged confrontations between characters and represents a repeated theme of attempted exculpation in situations of power imbalance.

'That's why you' (f=4) and 'I want you to' (f=3) are directive clusters, in which interpersonal coercion and command are reflected. Both sequences make the second-person recipient of the speech the subject of the speaker's command, creating an asymmetric relationship between the two at the utterance level. The cluster 'Evey I don't know' (f=4) is character-specific and serves to highlight instances of dialogic interaction where the character Evey is directly addressed in situations of uncertainty or epistemic failure, which helps to define Evey as a character who is placed in a situation of

imposed ignorance or confusion.

The locative-prepositional cluster 'In front of the' (f=4) sets up positioned relationships between characters and spatial contexts. It recurs in a manner that is consistent with the novel's mimicking of the power relations embedded in the act of appearing 'in front of' institutions, authorities, or audiences, which indicates exposure, confrontation, or submission.

8. Conclusion

This study has shown that V for Vendetta, a work in written English, can be analysed using AntConc 4.3.1 in order to generate a structured and analytically coherent corpus profile from the verbal text. The word list shows a frequency distribution in which resistance associated character names outrank authority figures, giving a quantitative justification of the discursive orientation of the novel. The analysis of Leader, Creedy, and Finch using KWIC concordance produces a set of differential keyword profiles that represent three registers of power: ideological proclamation, coercive enforcement, and internal moral critique. Denial is identified as the main multi-word pattern in the corpus using the N-gram and cluster analysis techniques and it is found that the most frequent phraseological units of the novel are such as directive, definitional, and locative-prepositional sequences.

Together, these findings prove that power discourse in V for Vendetta is not a singular characteristic of the text, but a multi-layered, distributed and contestatory discourse pattern, manifested in the frequency of words, collocation, and multi-word sequencing. The characterization would be unavailable when relying exclusively on close reading: The corpus evidence offers a systematic, reproducible, and comprehensive account of the discursive organization of the text, which complements and extends the close reading interpretation of the text.

The study falls under three intersecting fields of research: corpus stylistics, the linguistic

approach of literary power discourse, and the corpus-based assessment of graphic novel texts. It sets a methodological model for the use of full-text corpus analysis for the verbal dimension of graphic narratives and illustrates the analytical value and productivity of a single literary text as a corpus for frequency based investigation.

9. Implications

This study's results have implications for digital humanities, literary linguistics, and corpus stylistics. For corpus stylistics, the study confirms the methodological benefits of using n-gram and cluster analysis as well as KWIC concordance in the analysis of literary texts, showing that multi-word patterning can offer dimensions of evidence that go beyond what can be gained at the level of individual words. Literary linguistics sees the differential character concordance profiles as a repeatable model for quantitatively describing the discursive functions of characters in politically engaged fiction. The study provides an example of the ease and usefulness of building a single-text corpus from multi-modal sources for digital humanities, including graphic novels, where the verbal content is needed to be extracted manually from combined visual-textual format.

The empirical profile of power discourse developed here also has implications for comparative research. The collocational environments, cluster frequencies and frequency distributions presented for V for Vendetta are available as a standard for other dystopian works to be compared against, and researchers can ask whether the environments, cluster frequencies and frequency distributions found in V for Vendetta are unique to Moore's dystopian work, or whether they are characteristic for the dystopian genre.

10. Limitations

The results of this study are limited in scope and generalizability by the following factors. First, V for Vendetta as a single-text corpus cannot be subjected to keyword analysis based on some reference corpus which would allow a set of statistically key words for the text to be identified when compared to general English

usage. The findings of all frequencies are thus corpus-specific and must be interpreted as an impression of the text's own distributive priorities rather than as a reflection on the linguistic unusualness of the findings in comparison to a general norm.

Second, the corpus was manually transcribed to include all the verbal content of the novel. The transcription has been corrected for consistency but some transcribing errors or inconsistencies are not ruled out. A completely verified digital version of the text, preferably one independently checked against multiple editions, would be useful for future research.

Thirdly, the visual semiotic dimension of V for Vendetta is not included in the study. Since the book is a graphic novel, the meaning of the text is generated by the interplay between verbal and visual elements, and the absence of visual data from the corpus cannot help but restrict the comprehensiveness of the analytical account. The findings reported in the corpus relate only to the verbal text.

Fourth, the minimum frequency threshold of 3 applied to the cluster analysis might have missed out on lower frequency sequences that are analytically relevant. Other thresholds may produce other or other additional patterns.

11. Recommendations for Future Research

Based on the results of the study and its limitations, the following directions for future research are suggested.

- **Comparative corpus analysis** :The frequency profile, concordance patterns, and cluster data produced here should then be compared with the same analyses from other dystopian novels such as Orwell's 1984, Atwood's *The Handmaid's Tale*, and Huxley's *Brave New World* to see if any of the distributional features identified in V for Vendetta are genre specific or text specific.
- **Reference corpus comparison** Future studies should develop or adopt an existing reference corpus of similar literary English, so that a statistically conventional keyword analysis can be performed and identify words in V for Vendetta that are uncommon

in literary English.

- **Multimodal corpus extension:** The use of visual semiotic analysis, in addition to the verbal corpus data, would result in a more comprehensive description of the discursive construction of power within the novel. Two of the methodological approaches for analysing multimodal corpora, namely those of social semiotics and those of AntConc-based verbal corpus analysis, could be fruitfully integrated.
- **Subcorpus analysis for characters:** Future research should develop subcorpora for each major character's speech to obtain more character-specific frequency profiles and collocation networks than the full-text KWIC analysis allows.
- **Replication in other graphic novels:** The single-text corpus methodology developed here should be replicated with other politically engaged graphic novels to determine if the findings of this document, in particular the findings of the denial cluster and the stratified authority registers, are replicated across the medium.

References

- Anthony, L. (n.d.). *AntConc* [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc>
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Baker, P., & Egbert, J. (2016). *Using corpus linguistics to analyse power and resistance in online communities*. Bloomsbury Academic.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. Pearson Education.
- Crosthwaite, P., Ningrum, N. H., & Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis. *Journal of Corpora and Discourse Studies*, 6(3), 350–370.
- Di Liddo, A. (2009). *Alan Moore: Comics as performance, fiction as scalpel*. University Press of Mississippi.
- Foucault, M. (1995). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Vintage Books. (Original work published 1975)
- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings 1972–1977* (C. Gordon, Ed.). Pantheon.
- Foucault, M. (1972). *The archaeology of knowledge* (A. M. Sheridan Smith, Trans.). Routledge.
- Foucault, M. (1990). *The history of sexuality: Vol. 1. An introduction* (R. Hurley, Trans.). Vintage Books. (Original work published 1976)
- Kaveney, R. (2005). *From alien to The Matrix: Reading science fiction film*. I.B. Tauris.
- Leech, G., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose* (2nd ed.). Pearson Longman.
- Lewandowski, J. (2020). *Intertextuality and the depiction of ideological conflicts in graphic novels*. Graphic Studies Press.
- Lingvospektr Editorial Board. (2024). Emerging insights into corpus-focused exploration in literary inquiry. *Lingvospektr*, 8(1), 1–20.
- MacFarlane, I. (2019). *Gender, power, and language in graphic novels: A study of V for Vendetta*. Edinburgh University Press.
- Mahlberg, M. (2013). *Corpus stylistics and Dickens's fiction*. Routledge.
- Moore, A., & Lloyd, D. (1990). *V for Vendetta*. DC Comics.
- Munywoki, J. (2021). Narrative fragmentation and visual supremacy in *V for Vendetta*. *Journal of Graphic Novel Studies*, 12(4), 760–772.
- Partington, A. (1998). *Patterns and meanings: Using corpora for English language*

research and teaching. John Benjamins.

- Sharma, R., & Haider, A. (2024). *Representation of power dynamics in modern English literature: A corpus-based approach.* Oxford University Press.
- Wan, L. (2018). Trauma, memory, and resistance: A corpus-based study of A Thousand Splendid Suns. *Journal of Literary Linguistics*, 12(1), 78–94.
- Zahoor, S., Ahmed, R., & Khan, M. (2020). Gender, power, and discourse: A corpus-assisted study of V for Vendetta. *Journal of Language and Power*, 11(2), 98–115.