

Assignment 1: Institutional Language and Narrative Power in Harry Potter Canon and Fan Fiction

Date: 24/2/2026

Introduction

In what ways does power move between fan-made and canonical narratives? Do fan fiction works replicate and reinterpret institutional authority, or do they undermine it? Prior to starting this study, we believed that fan fiction would place more focus on human agency, interpersonal relationships, and decision-making, whereas official Harry Potter novels would prioritize institutional authority—Hogwarts, academics, and the Ministry of Magic. Our working hypothesis was that fan fiction would have a greater proportionate frequency of language connected to agency, whereas canon would be dominated by linguistic indications of institutional authority.

We created a five-text corpus of three fan fiction pieces (Twin Sister, Adoptive Kaiju, and Resurrection) and two canonical novels (Harry Potter and the Sorcerer's Stone and Harry Potter and the Deathly Hallows) in order to verify this hypothesis. First, using Voyant Tools for exploratory remote reading and RMarkdown in posit. We examined the distribution and relative frequency of words connected to institutions and agencies throughout the corpus using cloud for structured lexical comparison.

The anticipated conflict between canon and fandom is complicated by what unfolded. Even in transformational fan tales, institutional authority continues to play a major role in vocabulary. Fan fiction seems to restructure institutional authority inside modified narrative frameworks rather than substituting simply individualistic autonomy for formal power frameworks.

Corpus and Methodological Design

This project combines exploratory and confirmatory computational methods. Voyant Tools enabled rapid visualization of word frequency (Cirrus), term trends across documents, and keyword distribution (Bubblelines). These tools function as exploratory instruments, surfacing patterns that may not be immediately visible through linear reading.

RMarkdown was then used to tokenize the corpus, remove stopwords, normalize word frequencies, and group selected terms into conceptual categories. Following the explanation in *Hermeneutica*, Chapter 2 (“How Computers Measure Words”), computational text analysis operates by abstracting literature into tokens. Words become measurable units divorced from syntactic and narrative context. This transformation makes cross-text comparison possible, but it also risks oversimplifying meaning.

To examine how authority circulates, we constructed two lexical categories:

- **Institutional terms:** dumbledore, professor, ministry, headmaster, authority
- **Agency terms:** choose, decide, control, plan, trust, love

These categories allowed structured comparison across texts while remaining aware that lexical grouping is itself an interpretive act.

Voyant Insights: Lexical Prominence and Structural Authority

Figure 1: Cirrus Word Cloud

The Cirrus word cloud reveals that “dumbledore” appears prominently across the corpus. Institutional markers such as “professor” and “ministry” are also visually dominant, particularly in canonical texts. Interestingly, in *Twin Sister*, character-specific names rise to prominence, signaling a shift toward identity-centered reinterpretation.

The word cloud confirms that authority figures remain lexically central across both canon and fan fiction. However, prominence alone does not indicate whether authority is affirmed, questioned, or reconfigured.

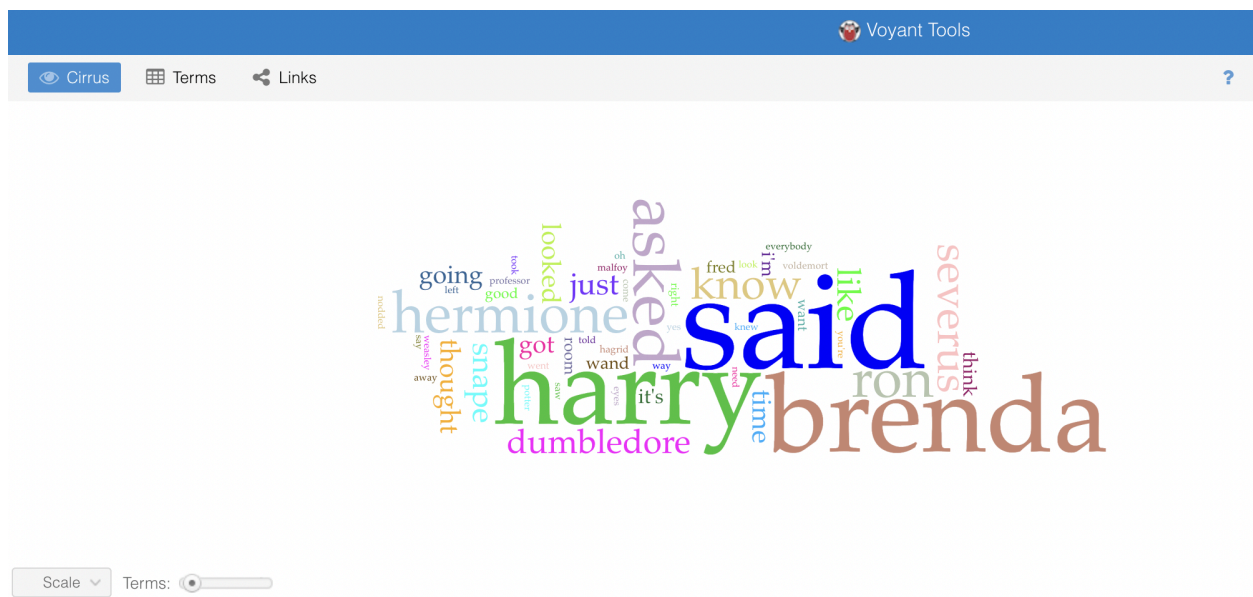


Figure 1. Most frequent terms across the five-text corpus (Voyant Cirrus).

Figure 2: Term Trends

The Trends graph compares selected keywords across documents. Several patterns emerge:

- “Dumbledore” peaks dramatically in *Deathly Hallows*.
- “Professor” is most concentrated in *Sorcerer’s Stone*.
- “Family” spikes significantly in *Twin Sister*.
- “Plan” remains relatively moderate across texts.

These patterns suggest that canonical texts foreground institutional structures, while fan fiction introduces relational language. However, the persistence of institutional terms in fan works challenges the assumption that fan fiction dismantles authority outright.

The Trends visualization reveals something that close reading might not quantify easily: structural continuity across textual transformations.

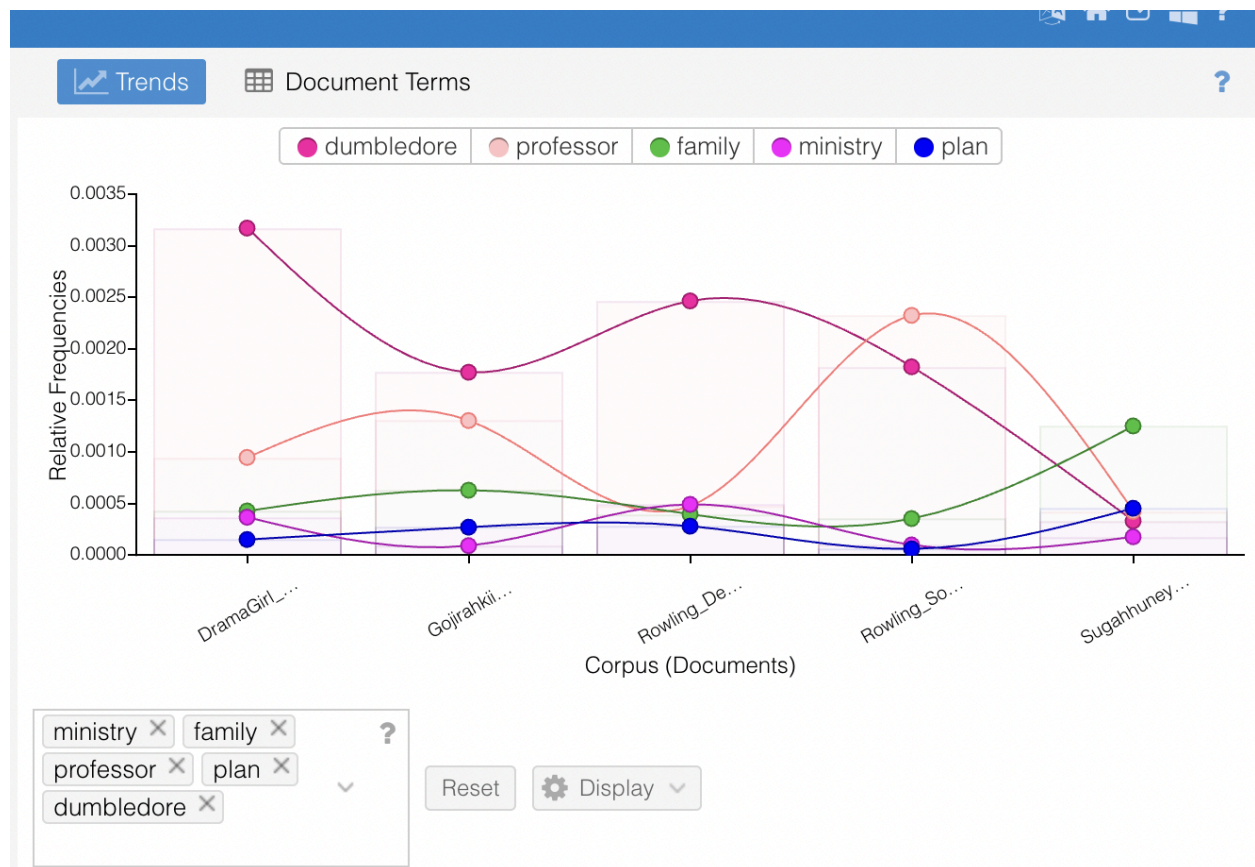


Figure 2. Relative frequency of selected keywords across canonical and fan fiction texts.

Figure 3: Bubblelines (Dumbledore Distribution)

The Bubblelines visualization shows how frequently and how consistently “dumbledore” appears across documents. In *Deathly Hallows*, references cluster densely throughout the narrative. In *Twin Sister* and *Adoptive Kaiju*, the term appears with notable consistency rather than marginal presence.

This is a crucial finding. If fan fiction sought to displace institutional authority, we might expect diminished lexical centrality of key authority figures. Instead, authority persists lexically.

Yet frequency does not capture tone. As Kestemont and Herman argue in “Can Computers Read (Literature)?”, computational tools measure textual features but cannot interpret irony, skepticism, or critique. A frequent name may signal reverence, subversion, or narrative dependency. Distant reading identifies structural presence, not evaluative stance.



Figure 3. Distribution of the term “dumbledore” across the corpus.

Add LIVE WIDGET

HTML Snippet

<!-- Exported from Voyant Tools (voyant-tools.org). -->

<iframe style='width: 100%; height: 800px;' src='https://voyant-tools.org/?panels=cirrus%2Creader%2Ctrends%2Csummary%2Ccontexts&corpus=5f1008ebba61f03f42e70984179b2c82'></iframe>

R Analysis: Institutional vs. Agency Vocabulary

To move beyond exploratory visualization, we implemented structured comparison in R.

After tokenization and stopword removal, we calculated word counts by text and normalized frequencies per 10,000 words. This normalization accounts for variation in text length, preventing longer works from skewing raw counts.

Figure 4: Institutional vs. Agency Word Counts

The grouped bar chart shows that institutional vocabulary dominates across canonical texts, particularly *Deathly Hallows*. Surprisingly, *Twin Sister* also exhibits strong institutional density. Rather than rejecting authority, this fan text intensifies engagement with it.

Agency markers increase modestly in fan fiction, but not dramatically. The difference is present yet nuanced.

This challenges my initial hypothesis. Fan fiction does not simply replace institutions with individual choice; it repositions authority within altered narrative frameworks.

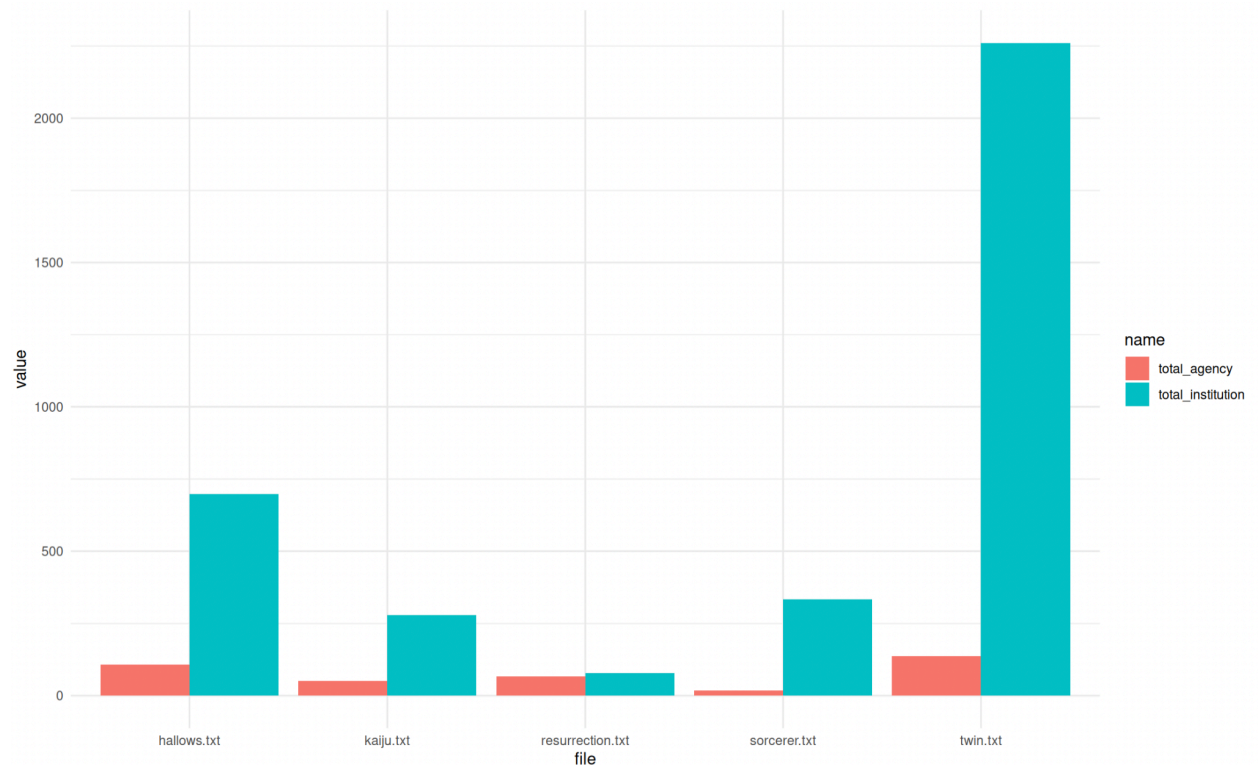


Figure 4. Comparison of institutional and agency word counts by text.

Figure 5: Heatmap (Normalized Frequency per 10,000 Words)

The heatmap refines this analysis by normalizing frequencies.

Key observations:

- “Dumbledore” frequency remains high even in fan fiction.
- “Professor” peaks in *Sorcerer’s Stone*.
- “Family” increases in *Resurrection*.
- Agency markers remain comparatively low overall.

Normalization confirms structural continuity across canon and fandom. Authority is not erased; it is reinterpreted.

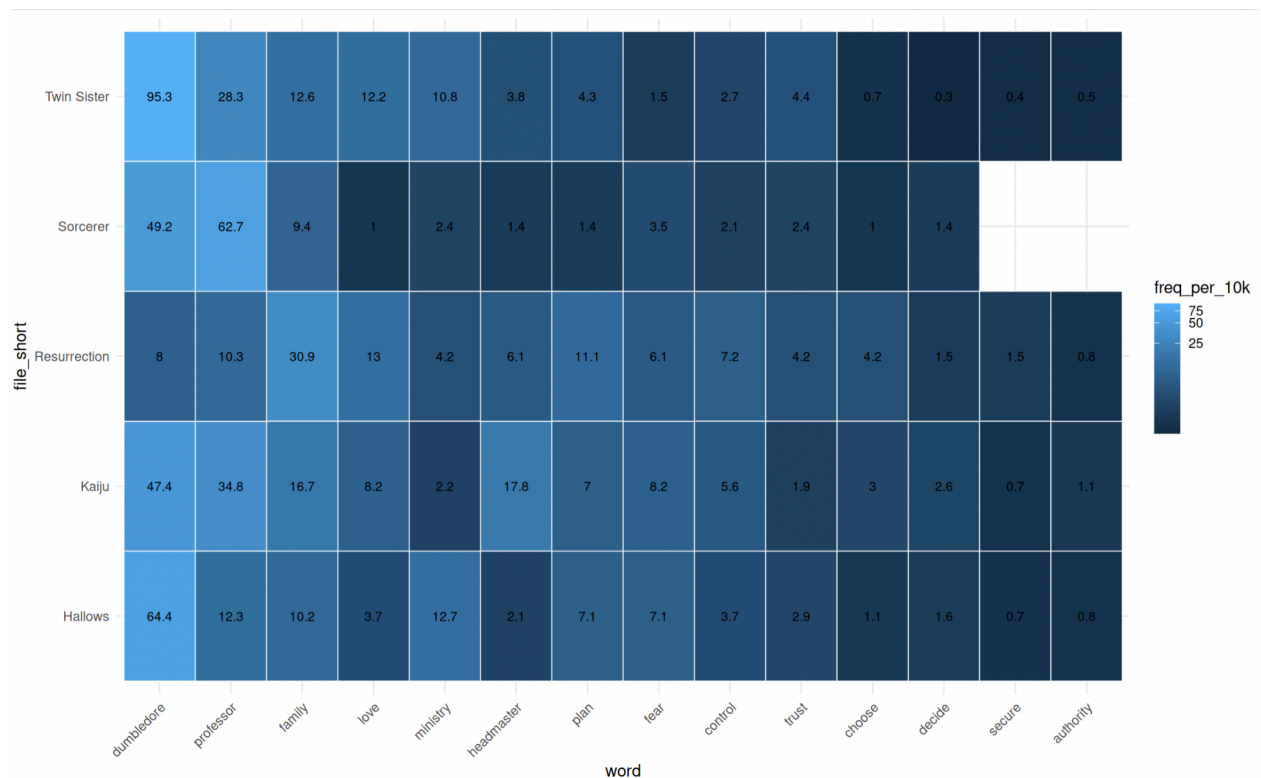


Figure 5. Normalized keyword frequency (per 10,000 words) across texts.

What Computational Analysis Reveals

Computational analysis reveals patterns that sequential reading might obscure. It would be difficult to intuitively compare lexical density across five long texts without statistical aggregation. Distant reading exposes proportional emphasis and structural prominence.

However, as Ted Underwood cautions in “The Risks of Distant Reading,” frequency patterns risk overinterpretation. Quantitative prominence does not equal narrative importance. A word may appear rarely yet shape central conflict.

Similarly, corpus composition shapes outcomes. As discussed in class regarding the pitfalls of Google Ngram, shifts in dataset structure influence apparent trends. Our corpus includes both canonical novels and fan fiction, genres that operate under different narrative conventions. Observed differences may reflect genre norms as much as ideological transformation.

Computational tools reveal tendencies, not truths.

Reading, Fast and Slow

Class discussions of close versus distant reading further contextualize this project. Speed reading and LLM summarization prioritize efficiency; distant reading prioritizes scale. Yet both sacrifice nuance.

Our workflow occupies an intermediate position. Voyant enables rapid pattern recognition, while R enforces deliberate categorization. Unlike large language models, which synthesize narrative probabilistically, this analysis remains transparent: word lists, grouping logic, and normalization methods are explicit.

As “People Don’t Read But LLMs Do” suggests, machine reading identifies statistical regularities rather than interpretive meaning. This project demonstrates both the power and limitation of that approach.

Scope and Scale

Focused comparison is made possible by restricting the corpus to five works, although generalizability is limited. There may be more noticeable changes in agency language in a bigger corpus of fan fiction. On the other hand, a smaller corpus might mask fandom variations.

Understanding is shaped by scale. Cross-text comparison is made possible by distant reading, but sophistication is restored by close reading. The most effective strategy alternates between them.

Conclusion

This analysis complicates the assumption that fan fiction displaces institutional authority in favor of personal agency. While relational language increases modestly in fan works, institutional markers—particularly references to Dumbledore—remain structurally central.

Distant reading reveals continuity where rupture was expected.

Yet computational analysis alone cannot determine whether authority is affirmed or subverted. As Underwood reminds us, distant reading identifies patterns but does not interpret them. Interpretation remains human work.

By pairing Voyant exploration, R-based modeling, and theoretical reflection, this project demonstrates both the utility and the limits of reading like a computer.

Works Referenced

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. Chapter “The Risks of Distant Reading.”

Kestemont, Mike, and Luc Herman. “Can Computers Read (Literature)?”

Hermeneutica. Chapter 2, “How Computers Measure Words.”

“The Pitfalls of Using Google Ngram.”

“People Don’t Read But LLMs Do.”

“READY FOR GRADING!”