

# Rhetorical invocation: a four-layer computational analysis of discourse vocabulary in institutional art writing

Haram Choi

`hachoi@hfk-bremen.de`

University of the Arts Bremen

---

## Article

**Keywords:** institutional discourse, computational discourse analysis, register analysis, International Art English, LLM evaluation, rhetorical invocation

**Posted Date:** October 16th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9188505/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Rhetorical invocation: a four-layer computational analysis of discourse vocabulary in institutional art writing

## Abstract

In fields where output quality cannot be assessed through objective metrics, institutional authority is produced partly through language. Institutions deploy specialised vocabulary to signal analytical sophistication and field membership. Whether this language performs the argumentative operations it formally resembles has not been measured at scale.

This study develops a four-layer computational framework to quantify how institutional discourse vocabulary diverges structurally from academic writing. The framework is applied to approximately 62,400 texts from 13 major art institutions, compared against 1,673 peer-reviewed articles across eight discourse categories.

Four independent metrics converge. First, institutional texts achieve syntactic convergence with academic writing when document length is controlled (adjusted  $d = -0.12$ ), because they adopt the same post-nominal modification structures that characterise academic prose. Second, within individual documents, International Art English (IAE) suffix adjectives form no repeated adjective-noun pairs (mean nMCE = 0.972;  $d = +2.00$  vs. academic baseline). This pattern holds across all 13 institutions. Third, discourse keywords appear without explicit causal or contrastive connectives in 60.1% of institutional documents versus 20.0% in academic texts (OR = 5.48 [4.86, 6.18]). Fourth, two independent large language models assign Analytical Tendency scores 3.5 times below the academic baseline.

Within the institutional corpus itself, lower modifier uniformity predicts higher analytical tendency by a factor of 7.1, a gradient that operates independently of domain-specific conventions. These findings define *rhetorical invocation*: the deployment of formal markers of analytical rigour in contexts structurally decoupled from the argumentative operations those markers support in academic writing. The measurement framework combines information-theoretic, dependency-syntactic, discourse-connective, and log-probability metrics. It requires no domain-specific assumptions and is transferable to any institutional corpus where language performs symbolic authority.

**Keywords:** institutional discourse, computational discourse analysis, register analysis, International Art English, LLM evaluation, rhetorical invocation

# 1. Introduction

In fields where performance cannot be assessed through objective criteria, career outcomes depend on institutional network position and accumulated symbolic capital rather than on demonstrable quality. Fraiberger et al. (2018) established this empirically for the contemporary art world: early access to high-prestige institutions predicts lifetime career trajectories more strongly than any quality-based measure. In such fields, prestige is not discovered. It is produced and reproduced through the practices of the field itself.

Language is one of those practices. Soro (2021) documents how online art magazines construct symbolic value through discursive performance. They deploy specialised vocabulary as a marker of field membership and intellectual authority. Soro (2023) extends this analysis to show how art organisations strategically incorporate discourses from the political and economic fields, a process of interdiscursivity through which they negotiate their positioning and legitimacy.

These findings jointly establish that the contemporary art world is a well-documented case of a structural condition common to prestige-driven institutional fields: language that must perform analytical authority in the absence of verifiable quality criteria. What they cannot establish, because they rely on qualitative analysis of small samples, is the structural signature of that performance at scale. How are the formal markers of analytical rigour distributed across the full textual output of 13 major institutions? Do the syntactic, lexical, argumentative, and semantic dimensions of this performance point in the same direction? Can the structural gap between performing analytical authority and executing it be measured with sufficient precision to serve as a transferable instrument?

This study answers these questions. It develops a four-layer computational framework and applies it to approximately 62,400 institutional texts compared against 1,673 peer-reviewed academic articles. Each layer targets an independent structural dimension: syntactic complexity (NPC), within-document adjective-noun collocation patterns (nMCE), argumentative connectivity (DMI), and semantic analytical tendency (AT). The four research questions ask whether institutional texts approximate academic syntax, whether they repeat adjective-noun pairings within documents, whether discourse keywords are anchored by explicit logical connectives, and whether probability mass concentrates at analytical or contextual depth.

The four layers and their findings are summarised below.

| Layer | Structural Dimension                 | Metric   | Finding vs. Academic Baseline        |
|-------|--------------------------------------|----------|--------------------------------------|
| I     | Syntactic scaffolding                | NPC-Post | Adopted (adjusted $d = -0.12$ )      |
| II    | Within-document collocation patterns | nMCE     | No repeated pairings ( $d = +2.00$ ) |

| Layer | Structural Dimension       | Metric | Finding vs. Academic Baseline    |
|-------|----------------------------|--------|----------------------------------|
| III   | Argumentative connectivity | DMI    | Absent (60.1% vs. 20.0%)         |
| IV    | Semantic analytical depth  | AT     | Contextual (3.5x below baseline) |

Layers I and II establish that the formal apparatus of academic discourse, its grammatical scaffolding and specialised vocabulary, is structurally present. Layers III and IV establish that the argumentative and semantic operations that generate rigour in academic writing are structurally sparse. The convergence of all four independent measurements defines the pattern termed *rhetorical invocation* in the Discussion.

Full pipeline documentation, scoring prompts, and disaggregated figures are available at [repository anonymised for review].

## 2. Background

### 2.1 Institutional Discourse and Symbolic Capital

Bourdieu (1984, 1996) provides the theoretical framework: agents in cultural fields accumulate symbolic capital, the recognised authority to define legitimate taste and value, through practices that are legible to others within the field. Texts are among those practices. They do not merely convey information; they perform institutional position. Fraiberger et al. (2018) confirm this quantitatively for the art world: network position and institutional affiliation determine career trajectories more strongly than any property of the work itself.

The vocabulary through which this symbolic function operates has a distinct identity. Rule and Levine (2012) identify International Art English (IAE) as a coherent register, a set of lexical and grammatical features that circulate internationally across institutional art communication. IAE signals institutional belonging. Its adoption identifies the user as a participant in a shared discursive community. Soro (2021) documents how this vocabulary constructs symbolic value through discursive performance in online art magazines. Soro (2023) extends the analysis to show that art organisations strategically incorporate discourses from the political and economic fields. This interdiscursivity, the absorption of external discursive practices, functions as a mechanism through which institutions negotiate their autonomy and legitimacy.

DiMaggio and Powell (1983) provide a structural explanation for why such practices converge across organisations within the same field. Their account of mimetic isomorphism predicts that organisations facing uncertainty adopt shared formal

structures as legitimacy signals, driven by field-level pressures rather than by independent design decisions. This prediction is tested directly in the invariance analysis in Section 5.

## 2.2 Computational Framework

Computational register analysis requires measuring multiple structural dimensions simultaneously, because no single feature captures the full complexity of a stylistic distinction (Biber 1988). Following the text-as-data framework for computational social science (Grimmer et al. 2022), this study treats institutional texts as observational data from which structural properties of language use can be measured at scale.

Evaluating semantic depth computationally introduces a specific challenge. Large language models reproduce human evaluative biases, including a tendency to prioritise surface stylistic features over substantive analytical depth (Feuer et al. 2024; Zheng et al. 2024). This halo effect makes generative LLM evaluation unreliable for comparing registers that share surface features but differ in semantic substance. The methodological solution is to extract log-probabilities over a constrained ordinal rubric rather than to prompt the model to generate a qualitative evaluation (Kadavath et al. 2022; Liu et al. 2023). The model’s probability mass, not its prose, constitutes the measurement.

## 3. Data and Methods

### 3.1 Institutional Art Corpus

The study constructs a corpus from 13 major contemporary art institutions, covering the period 2010 to 2025. The institutions were selected to represent a range of formats: critical magazines, artist-run platforms, museum publishing programmes, and international media art centres. The corpus captures the English-language, digitally native nodes of international institutional art communication. Commercial gallery texts, non-English-language output, and small-scale artist-run spaces are structurally absent. These absences share a directionality: the corpus over-represents institutions with established English-language digital publishing infrastructure.

The extraction pipeline preserves the Document Object Model (DOM) of each page, maintaining hierarchical content structure. A two-phase mid-size language model classification pipeline (Qwen2.5-14B-AWQ) first separates curatorial content from navigation, metadata, and administrative blocks, then assigns each page a depth category (A1: analytical, A2: contextual, A3: lists, A4: non-curatorial). Only A1 and A2 documents are retained. Including A2 documents is a deliberate choice: this study examines how institutions deploy discourse vocabulary across their full textual output, not only in analytically rigorous publications.

| Institution | <i>N</i> (docs) | Type              |
|-------------|-----------------|-------------------|
| artforum    | 17,344          | Critical magazine |

| Institution      | N (docs)       | Type                        |
|------------------|----------------|-----------------------------|
| e-flux           | 10,815         | Artist platform / journal   |
| mousse magazine  | 10,771         | Critical magazine           |
| neuralit         | 9,640          | Media art centre            |
| creative_app_net | 3,691          | Digital art platform        |
| ars_electronica  | 3,439          | Media art festival / centre |
| tate             | 2,048          | Museum (UK)                 |
| v2lab            | 967            | Media art institute         |
| transmediale     | 796            | Media art festival          |
| spikeart         | 394            | Critical magazine           |
| afterall         | 392            | Academic journal / platform |
| zkm              | 388            | Media art centre            |
| stedelijk        | 322            | Museum (NL)                 |
| <b>Total</b>     | <b>~62,400</b> |                             |

*LLM depth scoring was executed on 60,777 documents after deduplication. Median word count: 617.*

## 3.2 DOAJ Academic Baseline

The baseline consists of 1,673 peer-reviewed articles from the Directory of Open Access Journals (DOAJ), sampled across eight discourse categories (ecology, gender, postcolonial, race, capitalism, technology, identity, power) using targeted Boolean API queries. Year-balanced sampling across 2010-2025 restricted each journal to a maximum of three articles. The final collection spans 835 distinct journals with a median word count of 4,928. Because DOAJ articles undergo peer review, they require no language model based depth assignment.

The two corpora differ substantially in median document length (617 vs. 4,928 words). This length disparity is a structural feature of the two registers, not incidental noise. All four analytical layers apply explicit length controls.

## 3.3 Lexical Instruments

Three lexical instruments standardise the vocabulary under analysis across all four layers.

**Instrument 1: Discourse Keywords.** A reference vocabulary of 530 keywords distributed across eight discourse categories, manually curated to represent discourse vocabulary shared by both institutional art writing and academic discourse. Proper nouns are excluded.

**Instrument 2: IAE Suffix Filter.** Seven Latin and Greek suffix morphemes (-ic, -al, -ive, -ary, -ist, -ous, -ian), standardly associated with learned vocabulary in English lexicography

(Coxhead 2000), isolate qualifying adjectives for Layer II. A 48-item stop-list excludes non-specialised high-frequency forms (*national, digital, historical*).

**Instrument 3: Connectives Lexicon.** A structured list of explicit causal and contrastive connectives, adhering to the Penn Discourse TreeBank framework (Prasad et al. 2008). The Conservative metric restricts measurement to unambiguous markers (*because, therefore, thus, however, although*) and excludes polysemous forms (*since, as*). The Liberal metric includes all connective occurrences. Conservative DMI is the primary measure throughout.

| Instrument                 | Purpose                     | Items                                              |
|----------------------------|-----------------------------|----------------------------------------------------|
| Discourse Keywords         | Target shared vocabulary    | 530 keywords, 8 discourses                         |
| IAE Suffix Filter          | Isolate learned adjectives  | 7 suffixes, 48 stop-words                          |
| Connectives (Conservative) | Unambiguous logical markers | <i>because, therefore, thus, however, although</i> |
| Connectives (Liberal)      | All logical markers         | Includes polysemous forms                          |

Full pipeline documentation, prompts, and lexical instrument specifications: [repository anonymised for review].

## 4. Layer I: Syntactic Convergence

Does institutional art writing adopt the grammatical architecture of academic prose? Academic writing develops along a specific syntactic trajectory: it moves from simple clausal structures toward dense noun phrases compressed by post-nominal modification (Parkinson and Musgrave 2014). Post-nominal complexity, not pre-nominal modification, is the defining structural marker of advanced academic prose (Crossley and McNamara 2014). Measuring the ratio of post-modification to pre-modification across the two corpora provides the first structural diagnostic.

Noun Phrase Complexity is calculated in two dimensions, both normalised against total nouns per document:

$$\text{NPC-Pre}(d) = \frac{\text{amod}(d) + \text{advmod}(d)}{N_{\text{nouns}}(d)}$$

$$\text{NPC-Post}(d) = \frac{\text{prep}(d) + \text{relcl}(d) + \text{appos}(d) + \text{acl}(d)}{N_{\text{nouns}}(d)}$$

| Metric   | Art ( $N = 60,800$ ) | DOAJ ( $N = 1,660$ ) | Raw $d$ | Adjusted $d$ [95% CI]         |
|----------|----------------------|----------------------|---------|-------------------------------|
| NPC-Pre  | 0.2220               | 0.2544               | -0.38   | <b>-0.60</b> [-0.649, -0.551] |
| NPC-Post | 0.3246               | 0.3387               | -0.33   | <b>-0.12</b> [-0.169, -       |

| Metric | Art (N = 60,800) | DOAJ (N = 1,660) | Raw <i>d</i> | Adjusted <i>d</i> [95% CI] |
|--------|------------------|------------------|--------------|----------------------------|
|        |                  |                  |              | 0.071]                     |

*Raw d: Welch's t-test,  $p < 0.001$  for both. Adjusted d: OLS regression controlling for  $\log(\text{total nouns})$ ,  $p < 0.001$  for both coefficients.*

Controlling for length collapses the NPC-Post gap to  $d = -0.12$ . Art texts and academic texts deploy post-nominal modification at comparable density when document length is equalised. Conversely, the NPC-Pre gap widens to  $d = -0.60$ , indicating that art texts restrict pre-nominal attributive adjectives relative to academic prose. The structural picture is specific: institutional art writing converges with academic writing in post-nominal complexity while maintaining lower pre-nominal density.

The corpus is not uniform. Within e-flux, the journal category ( $N = 131$ ) reaches NPC-Post of 0.4516 ( $d = +1.39$  vs. DOAJ), substantially exceeding the baseline. Within artforum, news ( $N = 11,637$ ) records 0.2588 ( $d = -1.27$ ), well below it. Twelve of 13 institutions fall within  $\pm 15\%$  of the DOAJ NPC-Post baseline.

The cross-layer Spearman correlation between NPC-Post and Liberal DMI (Layer III) is  $\rho = -0.009$  ( $p = 0.092$ , ns;  $N = 37,470$ ). Post-nominal syntactic complexity and the density of explicit logical connectives are statistically independent. A text can deploy elaborate syntactic scaffolding without activating any causal reasoning. This independence is the central contribution of Layer I: the structural differences identified in Layers II through IV are not downstream effects of syntactic complexity.

## 5. Layer II: Within-Document Collocation Patterns

Layer I established that institutional art texts adopt post-nominal syntactic structures at rates comparable to academic writing. This chapter asks a more precise question: when both registers deploy discourse keywords, how do they distribute the adjectives that modify those keywords within individual documents?

Normalised Modifier Collocation Entropy (nMCE) measures the Shannon entropy (Shannon 1948) of the adjective distribution across all IAE-qualifying adjective-noun pairs within a single document, normalised against the theoretical maximum.

$$\text{nMCE}(d) = \frac{H(\text{adj-noun distribution})_d}{\log_2 K}$$

An nMCE of 1.0 means each adjective-noun pairing occurs exactly once: no pairing is repeated. A value approaching 0.0 means a small set of pairings recurs throughout the text. Documents with fewer than two unique qualifying adjectives are excluded.

Critically, nMCE is computed per document, not across the corpus. It measures whether, within a single text, adjective-noun pairings recur or remain one-off events. An academic

paper that uses *social capital* eight times and *social movement* five times gives *social* a concentrated, low-entropy distribution over nouns. An art text where *social* appears once each with *medium*, *practice*, and *justice* produces a maximally uniform distribution. The nMCE gap therefore reflects the presence or absence of within-document terminological repetition: whether authors build adjective-noun pairs into recurring conceptual units within a single text.

| Corpus | <i>N</i> | Mean nMCE | SD     | Cohen's <i>d</i> [95% CI] |
|--------|----------|-----------|--------|---------------------------|
| Art    | 49,791   | 0.9723    | 0.0387 | +2.00 [1.950, 2.050]      |
| DOAJ   | 1,669    | 0.8660    | 0.0665 | (reference)               |

Welch's *t*-test,  $p < 0.001$ .

Within each institutional art text, IAE-suffix adjectives (*social*, *cultural*, *political*, *critical*) form no repeated adjective-noun pairs. Each pairing is a one-off occurrence. In DOAJ texts, the same adjectives recur with specific noun partners (*social capital*, *critical theory*, *natural resource*), building terminological consistency. The vocabulary is present in both registers; the within-document terminological repetition that constitutes concept-building is not.

**Document-level versus corpus-level patterns.** The nMCE metric operates at the document level. At the corpus level, collocation patterns are mixed. Table 5.0 illustrates: some adjectives show higher concentration in DOAJ (*natural resource*: 14.2%), while others concentrate in the art corpus (*contemporary art*: 38.3%). The nMCE finding does not rest on corpus-level concentration but on within-document behaviour.

| Adjective    | DOAJ top-1 collocation      | Share | Art top-1 collocation    | Share |
|--------------|-----------------------------|-------|--------------------------|-------|
| natural      | <i>natural resource</i>     | 14.2% | <i>natural world</i>     | 9.4%  |
| critical     | <i>critical theory</i>      | 6.6%  | <i>critical thinking</i> | 3.4%  |
| political    | <i>political party</i>      | 4.8%  | <i>political issue</i>   | 1.6%  |
| social       | <i>social medium</i>        | 4.3%  | <i>social medium</i>     | 5.3%  |
| contemporary | <i>contemporary society</i> | 5.0%  | <i>contemporary art</i>  | 38.3% |

Table 5.0: Corpus-level collocation concentration. Top-1 share = frequency of the most common noun partner as a percentage of all pairings for that adjective.

**Filter robustness.** Re-running nMCE under three alternative lexical conditions (no filter: all adjectives; AWL adjectives only; top-100 most discriminating adjectives) produces  $d > +1.20$  in all cases. The primary finding does not depend on the specific morphological boundaries selected.

**Length controls.** An OLS regression controlling for log(word count) reduces the effect to  $d = +1.375$  [1.326, 1.424] (OLS coefficient  $p < 0.001$ ). Length explains approximately 20% of the raw gap. The remaining effect is large.

**Invariance.** All 13 institutions record mean nMCE above the DOAJ baseline, with effect sizes ranging from  $d = +0.80$  (stedelijk) to  $d = +3.42$  (neuralit). The lowest category-level  $d$  observed is +1.07 (tate/research). This contrasts with Layer I, where individual categories reached  $d = -0.25$ . No institutional format approaches the concentrated adjective-noun pairings observed in academic writing. Within-document collocational non-repetition is a property of the institutional register at every level, not a byproduct of promotional content.

**Triangulation.** The cross-layer Spearman correlation between nMCE and Analytical Tendency (Layer IV) is  $\rho = -0.211$  ( $p < 0.001$ ;  $N = 34,609$ ). A quartile analysis confirms the gradient:

| nMCE Quartile | <i>N</i> | Mean nMCE | Mean AT |
|---------------|----------|-----------|---------|
| Q1 (lowest)   | 3,930    | 0.894     | 0.114   |
| Q2            | 3,968    | 0.965     | 0.053   |
| Q3            | 3,925    | 0.987     | 0.041   |
| Q4 (highest)  | 3,973    | 1.000     | 0.016   |

From Q1 to Q4, mean AT falls by a factor of 7.1. Documents where adjectives concentrate on specific noun targets achieve analytical tendency scores 7.1 times higher than documents where adjectives distribute uniformly. Because Q1 lies within the art corpus (mean nMCE = 0.894), this gradient operates within the institutional register, not merely as a contrast between two corpora. The correlation is not large enough to suggest redundancy between nMCE and AT; the two metrics capture distinct properties of the text.

## 6. Layer III: Argumentative Structure

Layers I and II established that institutional art texts adopt complex post-nominal syntax and deploy discourse vocabulary without forming repeated adjective-noun pairings within individual documents. A third question follows: are the discourse keywords in these texts connected through explicit causal and contrastive reasoning?

Academic writing systematically deploys explicit connectives that guide readers through an argument, because peer-reviewed scholarship is structured around verifiable claims and logical transitions (Hyland 2005). Discourse Marker Interaction (DMI) measures the ratio of explicit logical connectives found within a specified syntactic radius of discourse keywords to the total number of keyword occurrences in the document.

$$\text{DMI}(d) = \frac{\sum \text{valid connectives within dependency radius}}{\text{total keyword matches}}$$

The dependency parser identifies connectives structurally linked to each keyword within four dependency steps, a cognitively grounded limit on syntactic distance (Liu 2008). The Conservative metric excludes all polysemous markers and restricts measurement to unambiguous logical connectives (*because, therefore, thus, however, although*).

| Corpus | <i>N</i> | Conservative DMI | Liberal DMI | Zero-Rate    |
|--------|----------|------------------|-------------|--------------|
| Art    | 62,655   | 0.0285           | 0.0915      | <b>60.1%</b> |
| DOAJ   | 1,673    | 0.0515           | 0.1267      | <b>20.0%</b> |

In 60 out of 100 institutional art documents, every discourse keyword lacks an explicit logical connective within its syntactic vicinity. In the academic baseline, this is true for only 20 out of 100. A chi-square test on the binary outcome (DMI = 0 vs. DMI > 0) confirms the disparity:  $\chi^2(1) = 870.8$ ,  $p < 0.001$ , OR = 5.48 [95% CI: 4.86, 6.18]. Art texts are 5.48 times more likely to deploy discourse keywords without any explicit causal or contrastive connective nearby.

**Polysemy control.** Removing polysemous markers (Conservative metric) widens the gap ( $d = -0.25$ ) relative to the Liberal metric ( $d = -0.21$ ). The polysemy objection is therefore eliminated: accounting for ambiguous markers strengthens the finding.

**Composition control.** Restricting analysis to the analytically oriented subset of the art corpus ( $N = 27,071$ ), the Conservative DMI is 0.025, still less than half the DOAJ rate of 0.051. The structural absence of explicit argumentation around theoretical keywords persists in the portion of the institutional corpus where analytical engagement is most expected.

**E-flux anomaly.** E-flux/journal, a format explicitly oriented toward long-form theoretical writing, records a Conservative DMI of 0.014. This is lower than many promotional formats from other institutions. Theoretically dense content does not automatically generate explicit logical connectives around discourse keywords.

**Triangulation.** The cross-layer Spearman correlation between Conservative DMI and Analytical Tendency (Layer IV) is  $\rho = +0.298$  ( $p < 0.001$ ;  $N = 37,474$ ). As explicit connective density increases, analytical depth increases systematically. Texts that deploy discourse keywords without explicit connectives are not merely choosing a different rhetorical style; they are statistically associated with lower analytical depth.

# 7. Layer IV: Semantic Depth and Multi-Dimensional Triangulation

## 7.1 Method

Layers I through III describe formal properties of the text. The final question goes deeper: do institutional texts engage with discourse keywords analytically, or do they position them as contextual references?

Free-form LLM evaluation introduces a systematic bias. Models trained on human writing reproduce a tendency to prioritise surface stylistic features over substantive depth (Feuer et al. 2024). This halo effect makes generative evaluation unreliable for comparing registers that share surface features but differ in semantic substance. This study eliminates the generative bias by extracting log-probabilities over a constrained six-level ordinal rubric, rather than asking the model to generate an evaluation.

| Level | Label         | Description                                                        |
|-------|---------------|--------------------------------------------------------------------|
| L0    | No Engagement | Keyword absent or text non-substantive                             |
| L1    | Mention       | Keyword listed without elaboration                                 |
| L2    | Decorative    | Keyword used for rhetorical effect                                 |
| L3    | Contextual    | Keyword appears in relevant context without analytical development |
| L4    | Analytical    | Keyword explained; causes, effects, or relationships articulated   |
| L5    | Theoretical   | Keyword integrated into sustained theoretical argument             |

Analytical Tendency (AT) captures the probability mass at the two highest levels:

$$AT(doc, discourse) = P(depth = 4) + P(depth = 5)$$

The full corpus was evaluated independently using two architecturally distinct models: Llama-3.3-70B and Qwen-2.5-72B. Both produce independent AT estimates for every document-discourse pair.

## 7.2 Aggregate Results

| Corpus | Model | <i>N</i> | L3 (%) | L4 (%) | AT |
|--------|-------|----------|--------|--------|----|
|--------|-------|----------|--------|--------|----|

| Corpus | Model | N      | L3 (%) | L4 (%) | AT            |
|--------|-------|--------|--------|--------|---------------|
| Art    | Llama | 60,777 | 81.2   | 10.3   | 0.0523        |
| Art    | Qwen  | 60,777 | 90.2   | 7.9    | 0.0295        |
| DOAJ   | Llama | 1,673  | 52.9   | 46.1   | <b>0.1851</b> |
| DOAJ   | Qwen  | 1,673  | 63.2   | 36.5   | <b>0.1094</b> |

The DOAJ baseline exceeds the art aggregate by a factor of 3.54 (Llama) and 3.71 (Qwen). Both models rank DOAJ above the art corpus by the same margin. The absolute AT values differ between models, but the relative ordering is fully reproduced.

Art texts concentrate probability mass at L3 (Contextual): the vocabulary describes what a work is about without explaining why or how. DOAJ texts concentrate mass at L4 (Analytical): the vocabulary serves as the instrument of an argument. Both models reproduce this pattern.

### 7.3 Institutional Spectrum

Three content formats across three institutions exceed the DOAJ baseline: spikeart/essay (AT = 0.3406, 184% of DOAJ;  $N = 32$ ), stedelijk/Journal Article (0.2634, 142%;  $N = 90$ ), and afterall/essay (0.2164, 117%;  $N = 228$ ). These formats demonstrate that DOAJ-level analytical depth is achievable within the institutional register. However, they collectively represent fewer than 700 documents out of approximately 60,000 scored. A composite of all analytically oriented institutional formats yields AT = 0.0908, which is 49% of the DOAJ baseline. Even the most analytically capable institutional formats, taken together, fall below the academic reference. The segment-level ranking is preserved across both models (Spearman rho = +0.407 between per-segment AT values;  $p < 0.001$ ), confirming that the analytical depth hierarchy reflects properties of the texts rather than model-specific calibration artifacts.

### 7.4 Multi-Dimensional Triangulation

The cross-layer Spearman rank correlation matrix confirms that all four layers measure distinct structural dimensions.

| Metric Pair              | Spearman rho | 95% CI           | $p$        |
|--------------------------|--------------|------------------|------------|
| NPC-Post x DMI (Liberal) | -0.009       | [-0.019, +0.001] | 0.092 (ns) |
| NPC-Pre x AT             | +0.164       | [+0.154, +0.174] | < 0.001    |
| NPC-Post x AT            | +0.092       | [+0.082, +0.102] | < 0.001    |
| nMCE x AT                | -0.211       | [-0.221, -0.201] | < 0.001    |
| DMI (Conservative) x AT  | +0.298       | [+0.289, +0.307] | < 0.001    |

*N = 34,607-37,474 documents with valid scores. 95% CIs via Fisher z-transformation. All tests two-tailed.*

No pair exceeds  $|\rho| = 0.30$ . The four metrics are correlated enough to converge on a shared empirical object, but non-redundant enough that none can substitute for the others. The most diagnostic entry is NPC-Post x Liberal DMI:  $\rho = -0.009$ , statistically non-significant. Complex post-nominal syntax is completely independent of argumentative connectivity. A text can adopt the full syntactic scaffolding of academic prose and still contain no argumentative structure around the vocabulary it deploys.

## 7.5 Human Validation

An independent annotator coded 203 (document, discourse) pairs from the art corpus on the same L0-L5 scale.

| Pair            | kappa (quadratic) | 95% CI         |
|-----------------|-------------------|----------------|
| Human vs. Llama | 0.324             | [0.221, 0.429] |
| Human vs. Qwen  | 0.421             | [0.276, 0.559] |
| Llama vs. Qwen  | 0.141             | [0.079, 0.209] |

*N = 203 pairs. 95% CIs via 10,000-iteration bootstrap. All kappa values significantly above zero (lower CI bound > 0).*

Kappa values of 0.32-0.42 indicate fair-to-moderate agreement, typical for fine-grained ordinal annotation. Spearman correlations between human labels and continuous AT are  $\rho = +0.319$  (Llama) and  $\rho = +0.439$  (Qwen), both  $p < 0.001$ . Inspection of the confusion matrix reveals that most disagreements occur when the human assigns L0 and the model assigns L1-L3. The model over-estimates engagement at the low end of the scale. Because art texts have a higher rate of human-L0 cases, this inflation affects art AT more than DOAJ AT. Therefore, the reported Art-DOAJ gap is a conservative estimate: the true gap is likely larger.

## 8. Discussion

### 8.1 Sociological Interpretation

The four layers converge on a structural description: institutional art texts adopt the syntactic architecture and specialised vocabulary of academic prose while the argumentative and semantic operations that generate analytical rigour in academic writing are largely absent. This pattern defines *rhetorical invocation*.

The institutional corpus is not uniform. It spans a continuous spectrum from long-form critical formats that match or exceed the academic baseline to brief logistical formats that concentrate near zero analytical depth. The analytically oriented stratum generates the discourse vocabulary and discursive authority that the institution requires. These texts establish discursive capital (Bourdieu 1996). The logistically oriented stratum then circulates that vocabulary across a much larger surface area of institutional

communication. Because the vocabulary circulates through the same terms that the analytical stratum developed, it projects the register of analytical authority without replicating the analytical operations. The result is a production system where discursive capital, generated by a small volume of analytically deep texts, is leveraged by a large volume of contextually oriented ones.

The invariance of nMCE across all 13 institutions, with a minimum effect size of  $d = +1.07$  even at the most academically oriented content category (tate/research), is not explained by any single institution's editorial policy. DiMaggio and Powell's (1983) account of mimetic isomorphism provides a parsimonious explanation: organisations operating within the same field converge on shared formal structures as legitimacy signals, driven by field-level pressures rather than by independent design decisions (see also Meyer and Rowan 1977). International Art English (Rule and Levine 2012) is the product of this convergence at the level of vocabulary. The nMCE invariance documents it at the level of within-document collocation structure.

## 8.2 Domain Generality

The domain-generality of the measurement framework is supported by evidence internal to the corpus rather than by theoretical assertion. The nMCE-AT gradient (Section 5) is computed within the institutional art corpus: documents in the lowest uniformity quartile (Q1, mean nMCE = 0.894) achieve analytical tendency scores 7.1 times higher than documents in the highest uniformity quartile (Q4, mean nMCE = 1.000). This relationship holds within a single domain's register, not only as a contrast between two different corpora. It indicates that the metrics capture a structural property of how language organises claims, not a genre convention specific to art writing.

The four metrics require no art-specific assumptions. NPC measures post-nominal modification density via dependency parse. nMCE measures within-document adjective-noun collocation entropy. DMI counts unambiguous connectives in the syntactic vicinity of a predefined keyword set. AT extracts log-probability mass over an ordinal depth rubric. Each metric can be applied to any corpus for which a relevant keyword vocabulary can be defined. Whether this structural pattern characterises institutional language in other prestige-driven fields, including policy institutions, academic disciplines, and cultural organisations, is an open empirical question with implications for how symbolic authority is reproduced through language across social fields.

## 8.3 The Genre Defence

A standing objection is that art writing pursues epistemological goals incompatible with academic argumentation: opacity, polysemy, and the productive indeterminacy of aesthetic experience (Glissant 1997). Under this objection, the structural differences reflect deliberate genre divergence rather than any deficit. This study takes no position on the epistemic value of poetic discourse. Three observations bound the genre defence empirically. First, three content formats exceed the DOAJ AT baseline (Section 7.3). If

analytical absence were an inherent property of the genre, these formats could not exist. Second, the nMCE invariance holds across all formats, including the most analytically oriented ones. The genre defence does not explain why even analytically ambitious institutional texts fail to develop repeated terminological pairings. Third, if accepted without qualification, the genre defence would make any measured difference unfalsifiable by definition. The spectrum data demonstrate that it is not universally applicable.

## 8.4 Limitations

The corpus covers 2010-2025 and cannot support diachronic analysis. Content format identification relies on URL-based heuristics, an operational proxy that classifies documents probabilistically. The four-layer analysis identifies co-occurring structural properties; it does not supply evidence about the intentions of the authors who produced those texts. The analysis is limited to the 530 discourse keywords in the predefined vocabulary.

## 8.5 Conclusion

This study constructs a four-layer computational analysis of how institutional art texts and peer-reviewed academic articles deploy a shared discourse vocabulary. Institutional texts adopt the syntactic architecture of academic writing (adjusted  $d = -0.12$ ). Within this structure, IAE-suffix adjectives form no repeated adjective-noun pairings within individual documents ( $d = +2.00$ ), invariant across all institutions. Discourse keywords appear without explicit logical connectives in 60.1% of documents (OR = 5.48). Two independent language models concentrate probability mass at contextual rather than analytical depth (3.54x gap). These four findings, individually defensible and collectively convergent, identify *rhetorical invocation*: a structural pattern in which the formal markers of analytical rigour are present while the argumentative and semantic operations that generate rigour are largely absent. The measurement framework is domain-general. Applying it to political institutional discourse, such as parliamentary reports, party policy documents, and expert advisory statements, would test whether this structural pattern characterises prestige-driven institutional language broadly.

## References

- Biber D (1988) Variation across speech and writing. Cambridge University Press, Cambridge
- Bourdieu P (1984) Distinction: a social critique of the judgement of taste (trans: Nice R). Harvard University Press, Cambridge, MA
- Bourdieu P (1996) The rules of art: genesis and structure of the literary field (trans: Emanuel S). Stanford University Press, Stanford

- Coxhead A (2000) A new academic word list. *TESOL Q* 34(2):213–238
- Crossley SA, McNamara DS (2014) Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *J Second Lang Writ* 26:66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- DiMaggio PJ, Powell WW (1983) The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *Am Sociol Rev* 48(2):147–160. <https://doi.org/10.2307/2095101>
- Feuer B et al. (2024) Style outweighs substance: failure modes of LLM judges in alignment benchmarking. *arXiv preprint arXiv:2409.15268*. <https://doi.org/10.48550/arXiv.2409.15268>
- Fraiberger SP et al. (2018) Quantifying reputation and success in art. *Science* 362(6416):825–829. <https://doi.org/10.1126/science.aau7224>
- Glissant E (1997) *Poetics of relation* (trans: Wing B). University of Michigan Press, Ann Arbor
- Grimmer J, Roberts ME, Stewart BM (2022) *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press, Princeton
- Hyland K (2005) *Metadiscourse: exploring interaction in writing*. Continuum, London
- Kadavath S et al. (2022) Language models (mostly) know what they know. *arXiv preprint arXiv:2207.05221*
- Liu H (2008) Dependency distance as a metric of language comprehension difficulty. *J Cogn Sci* 9(2):159–191
- Liu Y et al. (2023) G-Eval: NLG evaluation using GPT-4 with better human alignment. In: *Proceedings of the 2023 conference on empirical methods in natural language processing*. Association for Computational Linguistics, Singapore, pp 2511–2522. <https://doi.org/10.48550/arXiv.2303.16634>
- Meyer JW, Rowan B (1977) Institutionalized organizations: formal structure as myth and ceremony. *Am J Sociol* 83(2):340–363. <https://doi.org/10.1086/226550>
- Parkinson J, Musgrave J (2014) Development of noun phrase complexity in the writing of English for academic purposes students. *J Engl Acad Purp* 14:48–59
- Prasad R et al. (2008) The Penn discourse treebank 2.0. In: *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*. European Language Resources Association, Marrakech, pp 2961–2968
- Rule A, Levine D (2012) *International art English*. Triple Canopy 16
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423

Soro T (2021) How symbolic value is constructed through utterances: discursive power and performances of reputation in the field of online art magazines. *Poetics* 89:101572

Soro T (2023) How online discourse networks fields of practice: the discursive negotiation of autonomy on art organisation about pages. *Humanit Soc Sci Commun* 10:298. <https://doi.org/10.1057/s41599-023-01748-1>

Zheng L et al. (2024) Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Adv Neural Inf Process Syst* 36:46595–46623

## Data availability

The institutional art corpus and DOAJ baseline were constructed from publicly accessible sources. Derived datasets (document-level structural features, logprob-based depth scores) and the discourse keyword vocabulary are available at [repository anonymised for review]. Raw corpus texts remain copyright of the respective publishers and are not distributed.

## Code availability

All analysis scripts for reproducing the four-layer analysis from the provided data are available at [repository anonymised for review].

## Author contributions

[Anonymised for double-blind review.]

## Competing interests

The author declares no competing interests.

## Ethical approval

This study analyses publicly available institutional text archives and open-access academic journal articles. No human participants were recruited, no personal data were collected, and no biological material was used. The research was conducted entirely on publicly accessible textual data. Ethical approval was therefore not required.

## Informed consent

This study does not involve human participants. All analysed texts are publicly accessible institutional publications and open-access academic journal articles. No individual could be identified from the analysed data. Informed consent was therefore not applicable.

## AI usage disclosure

This study uses large language models as research instruments. Llama-3.3-70B-Instruct and Qwen-2.5-72B-Instruct were used for depth scoring (Section 7). Qwen2.5-14B-AWQ was used for corpus classification (Section 3). All model outputs were used as quantitative measurements, not as editorial content. AI-assisted editing tools were used during manuscript preparation. The author reviewed and takes full responsibility for all content.