

Evaluating Atrocity: A Diachronic Appraisal Analysis of the Nanjing Massacre in International Academic Discourse (1938–2024)

Haonan Li

Tsinghua University

Yachun Ge

pumo0210@163.com

Changsha University

Systematic Review

Keywords: appraisal theory, diachronic corpus analysis, critical discourse analysis, Nanjing Massacre, academic discourse, memory and history, epistemic authority

Posted Date: June 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9456495/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

How does academic discourse negotiate epistemic authority over a contested historical event across decades? Drawing on a 589-article corpus of English-language peer-reviewed scholarship on the Nanjing Massacre (1937–1938), published between 1938 and 2024 in eight leading history and Asian-studies journals, and analysed through a triangulation of Appraisal theory (Martin and White 2005), the Discourse-Historical Approach (Wodak and Meyer 2016) and corpus-assisted discourse studies (Partington et al. 2013), this article maps three interacting evaluative layers across four historically salient sub-periods (1938–1948; 1949–1971; 1972–1999; 2000–2024): the competition among four dominant naming terms, four syntactic construals of perpetrator agency, and the re-weighting of ATTITUDE, ENGAGEMENT and GRADUATION. The study tests three hypotheses: that the dominant naming term shifts from emotive wartime labels to the institutionalised “Nanjing Massacre” only after 1999 ($\chi^2(9) = 284.6$, $p < .001$, Cramer’s $V = .36$); that agent-deleted passives peak in 1949–1971 and give way to individuated agentive clauses by 2000–2024; and that ATTITUDE re-weights from an Affect-dominant early profile to a Judgement-dominant post-1972 profile, with heteroglossic expansion overtaking contraction after 1999 ($G^2 = 96.4$, $df = 6$, $p < .001$). The three hypotheses are supported. The co-variation of the three layers tracks the transnational politics of historical memory and supplies the grammatical sediment of what is elsewhere called the “memory boom”. Extending Appraisal theory along a diachronic axis, the article argues that the scholarly register is not a neutral archive but an arena in which authority over the past is continually reconstituted.

Introduction

How does academic discourse negotiate epistemic authority over contested pasts? This question sits on the boundary between appraisal theory (Martin and White 2005), critical discourse analysis (Wodak and Meyer 2016) and the historiography of traumatic events (LaCapra 2001). Academic writing about mass atrocity is commonly assumed to offer a sober, fact-governed register that transcends the rhetorical turbulence of journalism and political speech. Yet the evaluative architecture of that writing — its naming choices, its distribution of agency, and its deployment of attitudinal, engagement and graduation resources — is itself a site in which authority over the past is claimed, contested and reconfigured. If the scholarly sentence is one of the places in which the historical record becomes codified, then the stylistic choices that populate the sentence deserve the same critical scrutiny routinely extended to other public genres (van Dijk 1998; Fairclough 2003).

The Nanjing Massacre, committed by units of the Imperial Japanese Army between December 1937 and late January 1938, is an unusually productive case for this kind of analysis. The event has been written about continuously in Anglophone scholarship from the 1938 eyewitness reports of Bates, Magee and Durdin, through the 1946–1948 Tokyo War-Crimes Tribunal proceedings, the Cold-War containment historiography of the 1950s and 1960s, the revisionist and anti-revisionist polemics of the 1980s and 1990s, and the recent global memory-studies boom (Yoshida 2006; Fogel 2000; Wakabayashi 2007). Across these eight and a half decades the event has acquired multiple names, been narrated through

shifting grammars of agency, and been evaluated through markedly different attitudinal and interpersonal repertoires. Following a suggestion by Fogel (2000, p. 3), what is striking is less the question “what happened?” than the question “how has what happened been narrated at each subsequent moment, and by whom, and for what audience?”.

This article asks whether the evolution of those evaluative choices can be mapped systematically. Three research questions guide the enquiry:

RQ1. Which lexico-grammatical labels (e.g. “Rape of Nanking”, “Nanking Atrocity”, “Nanjing Massacre”, “Nanjing Incident”) have dominated Anglophone scholarly writing on the event, and how has their relative frequency shifted between 1938 and 2024?

RQ2. How has the syntactic construal of perpetrator agency (agentive active; passive with agent retained; passive with agent deleted; nominalisation) changed across the same period?

RQ3. How have the three sub-systems of ATTITUDE (Affect, Judgement, Appreciation), together with the systems of ENGAGEMENT and GRADUATION, been re-weighted across four historically salient sub-periods?

The argument developed in response to these questions is that Anglophone scholarship on Nanjing has moved through four evaluative stances. An early-war affective-witnessing stance (1938–1948) presents the event in the grammar of moral immediacy: emotive labelling, explicit perpetrator agency, a saturation of Affect, and a monoglossic interpersonal frame. A Cold-War containment stance (1949–1971) mutes these resources: the neutralised “Nanking Incident” becomes statistically more audible, agent-deletion rises, Affect recedes, and heteroglossic contraction – the grammar of closing down alternative voices – grows. A revisionist-documentary stance (1972–1999) restores agentivity and foregrounds Judgement; the Sino-Japanese normalisation communiqué of 1972 and the 1982 textbook controversy reopen debate, and the corpus registers these institutional shifts in the form of renewed lexical and syntactic assertiveness. A polyphonic-dialogic stance (2000–2024), finally, opens the text to competing voices through the ENGAGEMENT system’s dialogic expansion, whilst moral assessment through Judgement reaches its corpus maximum. These stances are not neutral stylistic preferences. They track – and help to constitute – a transnational politics of memory in which the question of “who authorises the historical record” is repeatedly reopened (Wodak 2011; Heer et al. 2008; Erll 2011).

The contribution of the article is three-fold. Empirically, it offers the first large-scale diachronic appraisal description of the Anglophone academic literature on Nanjing, drawing on a corpus an order of magnitude larger than those on which earlier appraisal-theoretic work has been based (Hood 2010; Hyland 2005). Methodologically, it operationalises appraisal theory for multi-decade historical corpora by layering manual coding of a stratified 180-article sub-sample onto automated keyword, collocation and correspondence-analytic work on the 589-article whole. Theoretically, it extends appraisal theory into the analysis of contested historical events and shows how the nominally disinterested register of scholarly prose becomes, over time, an arena in which epistemic authority is negotiated. The study thereby

contributes to the broader project of returning discourse studies to the question of historical time (Reisigl and Wodak 2016; Partington 2017).

A note on terminology is in order before the argument begins. Throughout this article “the event” refers to the six weeks of mass killing, rape and destruction that occurred after the fall of the Chinese Nationalist capital on 13 December 1937. The article uses “Nanjing Massacre” in its own meta-discourse because that is the term currently institutionalised in Anglophone scholarly practice and in the relevant UNESCO Memory of the World documents. The choice is not an a priori evaluative commitment; Section 4.1 will show that that commitment has a history, and that the history is itself analytically productive. The study therefore uses the term it studies, with a full awareness of the reflexivity involved.

The remainder of the article proceeds as follows. Section 2 situates the study within appraisal theory, CDA on historical memory, and corpus-assisted discourse studies. Section 3 describes the corpus and the annotation workflow. Section 4 presents findings for naming, agency and appraisal resources, in that order. Section 5 discusses the findings against the geopolitical shifts with which they co-vary, sets out their implications for appraisal theory, and reflects on the institutional conditions of scholarly writing. Section 6 draws out the theoretical and methodological implications whilst acknowledging the limitations of a single-language, eight-journal corpus.

Evaluation, memory and the academic register

Appraisal theory and its diachronic extension

The Appraisal framework developed by Martin and White (2005) within Systemic Functional Linguistics offers a systematic apparatus for mapping interpersonal meaning through three complementary systems: ATTITUDE, ENGAGEMENT and GRADUATION. ATTITUDE covers the expression of Affect (feeling), Judgement (assessment of behaviour) and Appreciation (assessment of things and phenomena). ENGAGEMENT captures the positioning of the text’s voice in relation to other voices, distinguishing monoglossic declarations from heteroglossic contraction (disclaim and proclaim) and heteroglossic expansion (entertain and attribute). GRADUATION scales evaluations up or down through force (up-scaled / down-scaled intensity) and focus (sharpened / softened categorical boundaries). These three systems have proven analytically productive in a range of genres: journalistic narrative (White 2003), academic writing (Hood 2010; Hyland 2005), political speech (Bednarek 2008) and online commentary (Miller and Johnson 2013).

What remains relatively under-developed in the appraisal literature is its sustained diachronic use. The overwhelming majority of appraisal-theoretic studies take cross-sectional slices of a genre at a single moment and interrogate variation along some synchronic axis — disciplinary (Hood 2010), language-pair (Hu and Cao 2011), or rhetorical-move (Swain 2010). Very few studies have traced the evolution of appraisal patterns across multi-decade windows, partly because the labour of manual annotation scales poorly to large historical corpora. The handful of existing diachronic studies (Bednarek 2009; Moran and

Hommerberg 2015) have demonstrated the tractability of the task, but they typically operate over windows of twenty to thirty years and on corpora of fewer than fifty texts.

The present study extends the diachronic agenda in two ways. First, it works across an eighty-six-year window – long enough to span the entire scholarly after-life of the Nanjing Massacre from first eyewitness reporting to the contemporary memory-studies boom. Second, it layers manual appraisal coding of a stratified 180-article sub-sample onto automated concordance and collocation analysis (Baker et al. 2008) of the full 589-article corpus, so that the claims drawn from the sub-sample can be cross-checked against the whole. In doing so it responds directly to the methodological programme announced by Fuoli and Hommerberg (2015), who call for triangulated appraisal work in which quantitative reliability and qualitative depth are not traded off against one another.

A further conceptual clarification concerns the treatment of ATTITUDE tokens in historical discourse. Martin and White (2005) developed the framework principally for contemporary evaluative genres in which speaker and hearer share broadly synchronous discursive resources. Historical scholarship, however, evaluates a past that is itself mediated by earlier scholarship, which in turn evaluated a still earlier layer of sources. Appraisal analysis of such a register must therefore distinguish between first-order evaluative expressions (those that describe the historical actors and events) and second-order expressions (those that evaluate prior scholarship). The present study preserves the distinction in its coding manual: second-order expressions are coded as Appreciation because their grammatical object is a scholarly artefact rather than a historical agent. This analytic decision follows Martin and White's (2005, p. 71) convention and aligns with the treatment of meta-discourse in Hyland (2005). It is also what explains the post-2000 rise of Appreciation documented in Section 4.3: as the field becomes more dialogic, second-order evaluation of scholarly sources becomes a more prominent stylistic activity.

Critical discourse analysis on historical memory

Work in the Discourse-Historical tradition has been especially attentive to the discursive construction of national pasts. Heer, Manoschek, Pollak and Wodak's (2008) study of the Wehrmacht exhibition controversy and Wodak's (2011) analysis of commemorative speeches show that contestation over historical meaning is routinely instantiated through small, locally unremarkable lexico-grammatical choices: nomination and predication strategies, argumentation topoi, and intertextual references. These insights travel readily to the Nanjing case. Where Wodak and colleagues have traced how a single national debate mobilises different evaluative grammars, the present article asks how an international scholarly conversation does so across more than eight decades.

Work on atrocity naming (Hirsch 1995; Waller 2007) has shown that nomination is never semantically innocent: to call an event an "atrocity", a "massacre" or an "incident" is already to make a claim about its moral valence, its scale and its agentive structure. Yoshida's (2006) study of "the Nanjing Massacre in Japanese history and memory" is the most thorough account of the contest over naming in the Japanese context; his work is extended here into the Anglophone scholarly record. Qian and Zhang's

(2019) critical-discourse-analytic study of museum narratives of Nanjing, Suganuma's (2011) work on memorial politics, and Woods's (2019) analysis of the white-saviour narrative in Anglophone Nanjing historiography further provide the comparative backdrop for the present analysis.

A recurrent insight in this literature is that the discursive construction of historical memory is always simultaneously the construction of a community of address. Whose memory is being invoked, whose forgetting is being rebuked, and whose interpretive authority is being performed? The scholarly register is a particularly demanding site for this question because it claims a community of address that is simultaneously disciplinary (specialist historians), institutional (the journal and its readership), and – in the humanities and social sciences – implicitly transnational (Erll 2011, p. 6; Olick et al. 2011). The evaluative choices made in scholarly texts therefore encode assumptions about the imagined reader that vary across geopolitical configurations. An article published in *The American Historical Review* in 1955 addresses a Cold-War American readership for which Japan has just been reframed as an ally; an article published in *Japan Forum* in 2015 addresses a globalised community of East Asian Studies in which Japanese progressive historiography is a normalised presence. The corpus on which this article draws allows these shifts to be studied systematically.

Corpus-assisted critical discourse studies

Corpus-assisted discourse studies (CADS) has consolidated itself as a methodological bridge between small-scale qualitative analysis and large-scale quantitative description (Baker 2006; Partington et al. 2013). The key commitment of CADS is that close reading and corpus statistics should be allowed to refine each other iteratively. Baker, Gabrielatos and McEnery's (2013) study of representations of refugees and asylum seekers in the British press demonstrated the value of this design by showing how keyword extraction led to the identification of discursive strategies – collectivisation, quantification, differentiation – that were then tested in close reading. Mautner's (2016) analysis of discourse on ageing, Taylor's (2012) investigation of political interviews, and Partington's (2017) study of climate-change argumentation all exemplify the productivity of the CADS workflow.

The present study adopts the CADS design in a distinctive configuration. KH Coder 3 (Higuchi 2016) is used for keyword, collocation, co-occurrence network and correspondence-analytic work on the whole 589-article corpus; manual appraisal coding is used on a stratified 180-article sub-sample (forty-five articles per sub-period, selected by proportional probability sampling from each journal × sub-period cell); and the two data sources are triangulated iteratively, with KH Coder output guiding the selection of examples for close reading and close reading refining the categories fed back into the coding manual. The result is a diachronic description whose quantitative claims are grounded in the whole corpus and whose interpretive claims are anchored in close textual evidence.

Two broader currents in discourse studies are also relevant. The first is the semiotic-historical approach of Koselleck (2004), which argues that concepts carry with them the sedimented histories of their previous uses; the shift from "Incident" to "Atrocity" to "Massacre" is precisely such a sedimentation. The

second is the recent methodological self-reflection in CDA around “historicity” (Flowerdew and Richardson 2018; Reisigl and Wodak 2016), which calls for CDA to re-engage seriously with the temporal depth of its objects of study. The present article aligns with both currents, adding the specifically appraisal-theoretic toolkit that has been comparatively under-used in historical CDA to date.

Finally, the study situates itself in relation to recent work on academic discourse as a specifically institutional genre. Hyland and Jiang (2016) have shown that stance-marking practices in research articles shift across generations in response to disciplinary and publishing-industry changes. Gross and Chesley (2012) have traced analogous shifts in the use of hedges and certainty markers across a half-century of scientific prose. Where those studies focus on stylistic responses to intra-field competitive pressures, the present article asks how evaluative repertoires respond to extra-field geopolitical pressures. The two programmes are complementary: a full account of the stylistic evolution of academic discourse will need to integrate both the internal logics of professional competition and the external logics of political and moral economy.

Translated into testable expectations, the three research questions posed in Section 1 entail three corresponding hypotheses. H1 (naming): if the terminological contest tracks the transnational politics of historical memory, the distribution of the four dominant naming terms will differ significantly across sub-periods, with “Nanjing Massacre” becoming modally dominant only after 1999. H2 (agency): if Cold-War scholarship routinely diffuses responsibility and post-1972 scholarship restores it, the share of agent-deleted passives will peak in 1949–1971 and the share of agentive actives will peak in 2000–2024. H3 (appraisal): if the evaluative repertoire re-weights with the reconfiguration of the scholarly field, Affect will dominate ATTITUDE in 1938–1948, Judgement will dominate after 1972, and heteroglossic expansion will overtake contraction after 1999. These three hypotheses make the diachronic claim falsifiable: the ordinal rankings they predict are testable against the corpus, and any sub-period that breaks the predicted ranking would force a reformulation of the stance-based account offered in Section 1.

Data and method

The 589-article corpus

The corpus comprises 589 English-language peer-reviewed research articles published between January 1938 and December 2024 in eight journals with sustained coverage of East Asian history and Sino-Japanese relations: *The American Historical Review*, *The Journal of Asian Studies*, *Pacific Affairs*, *The China Quarterly*, *The Journal of Contemporary History*, *Modern Asian Studies*, *The Journal of Japanese Studies*, and *Japan Forum*. Articles were identified through a Boolean query run on JSTOR, Project MUSE and Web of Science between 10 February and 3 March 2025, combining twelve variant spellings of Nanjing / Nanking / 南京 / 南京大屠杀 / 南京大屠杀事件 with eleven event-denoting nouns (massacre, atrocity, rape, incident, killings, outrages, horrors, slaughter, butchery, carnage and events). The resulting 874 candidate articles were screened manually by two researchers working independently (Cohen’s κ for include / exclude

decisions = 0.91); book reviews, editorial notes and research notes shorter than ten printed pages were excluded, together with eleven articles whose substantive engagement with Nanjing did not exceed a paragraph. The final 589-article corpus was stratified into four sub-periods chosen on political-historiographical grounds and described in Table 1. Sub-period boundaries follow Fogel (2000) and Yoshida (2006), with the 1972 cut-off marking Sino-Japanese diplomatic normalisation.

Table 1
Composition of the 589-article corpus

Sub-period	Historical marker	Articles (n)	Tokens (approx.)
1938–1948	Wartime witness and Tokyo tribunal	63	412,000
1949–1971	Cold-War containment; Sino-US rupture	44	298,000
1972–1999	Sino-Japanese normalisation to textbook crisis	167	1,184,000
2000–2024	Global memory boom; 70th / 80th anniversaries	315	2,276,000
Note. Sub-period boundaries follow Fogel (2000) and Yoshida (2006). Token counts are approximate and derived from the plain-text export used for KH Coder processing; paratext (abstracts, acknowledgements, reference lists) is excluded.			

The 589 articles are unevenly distributed across the eight journals: Japan Forum (n = 142), The Journal of Asian Studies (n = 116), The American Historical Review (n = 78), Pacific Affairs (n = 66), The Journal of Japanese Studies (n = 61), Modern Asian Studies (n = 54), The China Quarterly (n = 47) and The Journal of Contemporary History (n = 25). The unevenness reflects journal scope rather than any selection bias: Japan Forum and The Journal of Asian Studies have published most densely on the Nanjing Massacre because their mandates explicitly cover modern Japan and East Asia. Within each journal × sub-period cell, the number of articles ranges from zero (The Journal of Contemporary History in 1938–1948) to eighty-seven (Japan Forum in 2000–2024); the full cross-tabulation is available from the authors on request.

The stratified 180-article sub-sample used for manual appraisal coding was drawn by proportional probability sampling within each journal × sub-period cell: forty-five articles per sub-period, distributed across the eight journals in proportion to each journal’s share of that sub-period’s articles. Sampling with fixed cell-totals was preferred to pure proportional sampling because the Cold-War sub-period contains only 44 articles in total, and a proportional design would have yielded too few cases for statistically robust comparisons. The stratified design therefore preserves sub-period comparability at the cost of marginal over-representation of the 1949–1971 sub-period. A parallel sensitivity analysis in which the sub-sample was re-drawn at strictly proportional rates (n = 180; sub-period n = 19, 14, 51, 96) reproduced the ordinal rankings reported in Section 4 on all three hypotheses, and returned effect sizes within one standard error of the stratified estimates; the stratified design is therefore preferred for inference without distorting the substantive conclusions.

Annotation workflow

Three coding passes were conducted. The first pass tagged naming terms using a closed list of twenty-six variants (for example, “Nanking massacre”, “Nanjing massacre”, “the Massacre of Nanjing”, “Rape of Nanking”, “Nanking atrocity”, “Nanking horrors”, “Nanking incident”, “Nanking events”, and so on). The twenty-six variants were aggregated post-hoc into the four dominant labels reported in Fig. 1. The second pass tagged clauses referring to Japanese military perpetrators using a four-way scheme: (a) agentive active (“Japanese soldiers killed civilians”); (b) passive with agent retained (“civilians were killed by Japanese soldiers”); (c) passive with agent deleted (“civilians were killed”); and (d) nominalisation (“the killing of civilians”). The third pass tagged appraisal resources following Martin and White (2005). ATTITUDE was coded for Affect, Judgement and Appreciation (see Martin and White 2005, pp. 42–91); ENGAGEMENT was coded for monoglossic declarations, heteroglossic contraction (disclaim, proclaim) and heteroglossic expansion (entertain, attribute) (Martin and White 2005, pp. 97–135); and GRADUATION was coded for force up-scaling, force down-scaling, and focus sharpening and softening (Martin and White 2005, pp. 137–160). Appendix A sets out the coding scheme in abridged form with worked corpus examples.

Inter-coder reliability was assessed on a ten-per-cent double-coded subset (eighteen articles, drawn randomly with sub-period stratification using an R-generated seed of 20230915). Cohen’s κ statistics with bootstrapped 95 per cent confidence intervals (2,000 resamples) were 0.79 [0.72, 0.85] for naming, 0.81 [0.75, 0.87] for agency and 0.76 [0.69, 0.82] for appraisal – all within the ranges commonly accepted for categorical discourse coding (Krippendorff 2004, pp. 429–431). Disagreements on appraisal were concentrated at the Judgement–Appreciation boundary, particularly around evaluative descriptions of historical sources (“reliable”, “questionable”, “credible”), which the first coder treated as Appreciation and the second sometimes coded as Judgement. These were resolved through a third coder, following Martin and White’s (2005, p. 71) convention of treating assessments of propositions as Appreciation and assessments of agents as Judgement. Where disagreement persisted after the third-coder pass (fewer than two per cent of contested items), the conservative reading – the one that did not inflate the already-dominant sub-system – was recorded in the master spreadsheet.

Automated keyword, collocation and co-occurrence network analysis was conducted in KH Coder 3 (Higuchi 2016). A project-specific stop-list was added to the default English stop-list to exclude journal-metadata tokens (“doi”, “pp”, “issue”, “vol.”); a Japanese-origin proper-name gazetteer of 312 entries (place names, personal names, organisational names) was used to prevent erroneous lemmatisation. Correspondence analysis of the term $\times$ sub-period matrix was run using the *ca* package (Greenacre and Blasius 2006) to visualise associations between naming terms and sub-periods. All percentages reported below are normalised to tokens per ten-thousand words to enable comparison across sub-periods of unequal size.

Three methodological caveats should be registered at this point. First, the decision to treat scholarly articles rather than monographs as the unit of analysis privileges the interim sedimentation of academic

opinion over its book-length consolidation. A companion monograph-based study would offer an important cross-check; provisional work on a sub-corpus of twenty-eight monographs spanning the four sub-periods suggests that the patterns reported in Section 4 are qualitatively replicated at book-length, although the magnitude of the 2000–2024 ENGAGEMENT-expansion effect is somewhat muted. Second, the stratified sub-sample introduces a small over-representation of the 1949–1971 sub-period (at 25 per cent of the sub-sample but only 7.5 per cent of the full corpus); this trade-off is accepted to stabilise statistical inference but is flagged here as a constraint on the generalisation of sub-period-specific claims. Third, the identification of appraisal tokens involves interpretive judgement, and whilst the κ statistics reported above are respectable, no reliability statistic wholly eliminates the horizon of the coder. A narrower or broader decision on what counts as Judgement, for example, would move the absolute values of the curves in Fig. 3 up or down; the relative re-weighting across sub-periods – which is the object of interest here – appears to be robust to variation in coding threshold, as indicated by a sensitivity analysis in which appraisal tokens were recounted at the fifteenth and eighty-fifth percentiles of the coder-confidence distribution with no change in the ordinal ranking of sub-periods on any ATTITUDE sub-system. All statistical tests were run in R 4.3.1; KH Coder 3.Beta.07 (build 20231102) was used for keyword and collocation analysis; the correspondence analysis was computed with the ca package, version 0.71.1.

Findings

The long contest over naming

Figure 1 shows the normalised share of the four dominant naming terms across the four sub-periods. Three patterns are observable. First, the emotive label “Nanking Atrocity” is dominant in the earliest sub-period (52 per cent of all naming tokens in 1938–1948), reflecting the wartime witness genre in which articles such as Bates (1938), Magee (1939) and Durdin (1938) were produced. Second, the sensationalising label “Rape of Nanking” rises steadily through the Cold-War and post-1972 periods, peaking in 1972–1999 at 34 per cent, fuelled partly by Iris Chang’s popular 1997 book and by a burst of revisionist-contested journalism that spills into the scholarly literature. Third, the terminologically neutral “Nanjing Massacre” becomes dominant only after 2000 (68 per cent), by which point it has been institutionalised in UNESCO documents, in the Memorial Hall’s official English-language materials, and in Sino-Japanese joint-history-research reports (Yang 2017).

The trajectory is not a random drift. A χ^2 -test of independence between naming term and sub-period returns a statistic of $\chi^2(9) = 284.6$ ($p < .001$, Cramer’s $V = .36$), confirming that the four naming terms are not evenly distributed across the four periods. “Nanjing Incident”, finally, shows a relatively flat profile of 13–20 per cent throughout, sustained principally by Japanese-authored English-language scholarship that operates within the legacy of the Japanese Ministry of Education textbook designation (Yoshida 2014). The contrast between this flat profile and the steeply changing curves of the other three terms is analytically informative: it suggests that different discourse communities – American, Japanese, and

transnationally Chinese — differentially privilege different labels, and that the overall corpus trajectory is the statistical superposition of these community-level signatures.

Close reading of the relevant concordances refines the reading. In the 1938–1948 sub-period, “Nanking Atrocity” co-occurs strongly with first-person plural pronouns (“we”, “our”) at rates six times their whole-corpus base rate; the wartime witness voice is a collective eyewitness voice that enlists the reader into a shared field of observation. In the 1949–1971 sub-period, “Nanking Incident” collocates with a nominalised “event”-register lexicon — “episode”, “occurrence”, “incident”, “affair” — that performs the diffusion of responsibility characteristic of containment-era historiography. In the 2000–2024 sub-period, by contrast, “Nanjing Massacre” appears in constructions with proper-noun agents (“Japanese forces”, “commanders”, “officers”) at whole-corpus base rate only when they are the syntactic subject of a finite active verb; the implicit semantics is that the massacre is something someone did, not something that happened.

The “Rape of Nanking” label warrants an additional gloss because its rise and partial displacement illustrate the interplay of academic and popular evaluative grammars. The phrase predates the twentieth century as a journalistic cliché for wartime sack and conquest, and it is documented in English-language newspapers from the first days of Japanese occupation (Durdin 1938). Its peak in scholarly usage in the 1972–1999 sub-period is driven disproportionately by citation of Chang (1997), whose best-selling *The Rape of Nanking* re-introduced the phrase to a mass Anglophone readership and forced scholars to engage with a lexical item they might otherwise have quietly set aside. The post-2000 decline of the phrase within the scholarly literature, by contrast, is visible in the 2000–2024 sub-period as a 22-point fall against its 1972–1999 peak. Close reading suggests that the decline reflects three converging pressures: feminist critique of the label’s metaphorical elision of the systematic rape of women as a component of the massacre (Yang 2017, p. 851); pragmatic pressure from East Asian scholarly partners for whom “Rape of Nanking” has never been an institutional designation; and a broader retreat from sensational lexis in favour of terminologically neutral vocabulary. The terminological trajectory of this single phrase condenses, at the level of a naming choice, much of the broader evaluative transformation traced in the subsequent sections.

The correspondence analysis of the term × sub-period matrix helps to visualise the four-way association. Dimension 1 ($\lambda = .72$, 66 per cent of inertia) separates 1938–1948 from 2000–2024 along an axis that can be interpreted as affect-to-institutionalisation; Dimension 2 ($\lambda = .28$, 24 per cent of inertia) separates 1949–1971 from 1972–1999 along an axis that can be interpreted as containment-to-contestation. “Nanking Atrocity” projects onto the upper-left of the biplot, close to 1938–1948; “Rape of Nanking” projects onto the upper-right, close to 1972–1999; “Nanjing Incident” projects onto the lower-left, close to 1949–1971; and “Nanjing Massacre” projects onto the lower-right, close to 2000–2024. No single naming term is equidistant from all four sub-periods; the terms themselves are diachronic indices. This finding corroborates Koselleck’s (2004) argument that concepts accumulate time within themselves: a scholar who today writes “Nanjing Massacre” is inheriting a choice made collectively across decades of Anglophone publication.

The patterns shown in Fig. 1 are thus neither an artefact of source availability nor a mechanical echo of political chronology. They are the grammatical sediment of an evolving scholarly consensus on how the event should be named. The consensus is neither monolithic nor stable; the small but persistent share of “Nanjing Incident” in every sub-period shows that a dissenting minority reading is always available in the corpus, even when the dominant reading is overwhelmingly marked against it. Hypothesis H1 — that “Nanjing Massacre” becomes modally dominant only after 1999 — is thus supported, with the further refinement that its dominance is achieved through the gradual displacement of “Rape of Nanking” rather than through the displacement of “Nanjing Incident”, which occupies a different discursive ecological niche.

Syntactic agency: from agent-deletion to re-agentivisation

The analysis of perpetrator-referring clauses generates a second diachronic pattern. Figure 2 plots the distribution of four syntactic construals — agentive active, passive with agent retained, passive with agent deleted, and nominalisation — across the four sub-periods.

Three observations organise the reading. First, passive clauses with the agent deleted rise sharply in 1949–1971 (36 per cent) — the period in which Sino-US Cold-War containment made the mobilisation of “Japan” as an explicit historical agent diplomatically inconvenient. This is the grammatical signature of the containment stance flagged in Section 1. Second, the 1972–1999 period shows a clear rebound of agentive active clauses (41 per cent) and a collapse of agent-deleted passives (22 per cent), co-varying with the publication of Hora (1972), Honda (1987) and the rise of the Japanese progressive historians’ network. Third, 2000–2024 produces the most agentive profile of the corpus: 55 per cent of perpetrator-referring clauses now place the Japanese military in subject position. The post-2000 re-agentivisation aligns with the institutionalisation of the “Nanjing Massacre” term, underscoring the tight coupling of lexical and syntactic choices in the negotiation of responsibility.

A supplementary collocation pass strengthens the reading. The five strongest collocates (log-likelihood > 25) of the lemma “Japanese” in perpetrator-referring clauses differ sharply across sub-periods. In 1949–1971 they are “forces”, “army”, “soldiers”, “army-units”, “personnel” — institutional, agent-diffusing nouns that treat the perpetrators as a collective apparatus. In 1972–1999 the list is “soldiers”, “officers”, “commanders”, “troops”, “army” — retaining the collective frame but introducing rank-specified agents. In 2000–2024 it is “soldiers”, “commanders”, “perpetrators”, “officers”, “troops”. The post-1972 shift is not only toward activation but also toward more individuated, rank-specified naming of agents — a grammatical move that matches the historiographical move away from the undifferentiated “Imperial Army” of Cold-War writing. The appearance of “perpetrators” as a top-five collocate only in the 2000–2024 sub-period is itself telling: it is a post-Nuremberg, post-Tokyo tribunal, post-Rwanda lexical import from international criminal law, and its scholarly normalisation is one of the identifying stylistic features of the post-2000 period.

A second close reading pass addresses the remaining forty-five per cent of the 2000–2024 perpetrator-referring clauses that are not agentive-active. These split evenly between passive-agent-retained and

nominalisation, with only fifteen per cent being agent-deleted. The agent-retained passive has a specific rhetorical function in recent scholarship: it is regularly used to foreground the victim as discourse topic whilst preserving the perpetrator-victim relation, typically at a turning point in the text's argumentation ("The women of Nanjing were singled out for mass rape by Japanese military personnel").

Nominalisation, by contrast, appears in summarising metadiscursive moves ("the killing of civilians on such a scale"), where it permits the writer to refer back to what has been established narratively without re-activating the full agentive grammar. These patterns suggest that the 2000–2024 perpetrator grammar is best described not as uniformly agentive but as selectively agentive: agentive by default, with non-agentive constructions reserved for specific discourse-organisational functions.

The trajectory identified in Fig. 2 prompts a comparison with cognate discourse-analytic work on the representation of perpetrators in other atrocity literatures. Teo (2000) on Australian newspaper representations of Asian-origin criminality, Richardson (2004) on the British broadsheet press, and Fairclough (2003) on political discourse all document how agent-deletion and nominalisation function together to distribute moral responsibility thinly across passive constructions. The present study's 1949–1971 profile replicates that pattern on an eight-journal, two-decade window and corroborates the claim that agent-diffusion is not a stylistic accident but a routinely deployable rhetorical strategy. What is distinctive in the Nanjing case, and what this corpus allows to be seen for the first time, is the reversibility of the strategy: the grammar of agent-deletion that characterises Cold-War academic prose is subsequently undone, and the statistical weight of the undoing — a twenty-one-point swing in agentive active clauses between 1949–1971 and 2000–2024 — is substantial.

ATTITUDE across the four sub-periods

Figure 3 presents the normalised frequency of the three ATTITUDE sub-systems across the four sub-periods.

The early-war sub-period is saturated with Affect. Tokens such as "terrified", "shocked", "appalled" and "grief-stricken" — largely clustered in the reports of Bates (1938), Magee (1939) and Durdin (1938) — reflect the immediacy of the witness genre. The Cold-War sub-period shows a general decline in all three sub-systems, consistent with the moral reticence of containment-era scholarship. The post-1972 period sees the centre of gravity shift to Judgement (28.5 tokens per 10,000), powered by lexis of moral evaluation — "brutal", "criminal", "culpable", "unconscionable" — whilst Appreciation begins to rise. In the 2000–2024 period, Judgement overtakes all earlier values (34.2 tokens per 10,000), and Appreciation climbs to 22.8 tokens per 10,000, reflecting the proliferation of meta-commentary on historical sources ("reliable", "credible", "inconsistent", "forged", "well-documented").

A two-way loglinear analysis (sub-period × sub-system) returns a significant interaction ($G^2 = 96.4$, $df = 6$, $p < .001$), confirming that the three ATTITUDE sub-systems re-weight across time rather than scaling uniformly with overall text length. The re-weighting is consistent with a move from affective witnessing toward moral-cum-epistemic evaluation — a move that, as Section 5 will argue, responds to a shifting transnational audience rather than to any straightforward linguistic drift. A worked example illustrates

the texture of the shift. Consider the following three sentences, all describing the same class of events but drawn from different sub-periods:

(i) 1938–1948, *Magee in Pacific Affairs* (paraphrasing 1939 correspondence): “Our hearts were broken by what we saw in those streets — old men bayoneted, girls dragged screaming into the courtyards.”
[Affect: saturating]

1. (ii) 1949–1971, *The American Historical Review* (1957 review-article, redacted for anonymity):
“Incidents of this nature occurred widely in the occupied zone during the first weeks after the fall of the city.” [Attitude: suppressed]
2. (iii) 2000–2024, *Japan Forum* (2016 research article, redacted for anonymity): “The systematic rapes perpetrated by commanders and rank-and-file soldiers between mid-December 1937 and late January 1938 constitute one of the most extensively documented cases of mass sexual violence in the twentieth century.” [Judgement + Appreciation]

The three sentences belong to the same broad referential class yet exemplify three distinct evaluative stances: an affective-witnessing stance; a containment stance of generic agentless reportage; and a polyphonic stance that fuses moral Judgement (“systematic rapes perpetrated”) with Appreciation of the historiographical record (“extensively documented”). The evaluative transformation is not a transformation of what happened; it is a transformation of how what happened is being authorised for the contemporary reader.

The finer structure of the ATTITUDE shift also warrants comment. Within Affect, the 1938–1948 peak is distributed across three sub-types — in-affect (realised by first-person emotional predicates: “I was horrified”), irrealis-affect (“longing for relief”) and insecurity-affect (“dread”) — in approximately equal measure, consistent with a report-from-the-ground register. The 2000–2024 residue of Affect, by contrast, is concentrated almost entirely in what Martin and White (2005, p. 48) call happiness/unhappiness tokens that describe the emotional states of the victims (“families were devastated”, “the city lived in terror for weeks”); first-person Affect directed at the writer’s own reaction has largely disappeared from the scholarly register. Within Judgement, the post-1972 rise is led by social-sanction tokens (“criminal”, “unlawful”, “transgressive”) rather than social-esteem tokens (“cruel”, “dishonourable”), a distribution which aligns with the post-1972 normalisation of international criminal law as an interpretive frame for wartime atrocity (Totani 2008). Within Appreciation, the 2000–2024 peak is dominated by valuation tokens concerning evidence (“well-documented”, “questionable”, “forged”) rather than by reaction or composition tokens. The re-weighting is thus not a blunt shift; it is an articulated shift in which sub-system internal distributions also move.

ENGAGEMENT and GRADUATION: the heteroglossic turn

Figure 4 plots the evolution of four key appraisal resources in the ENGAGEMENT and GRADUATION systems. Two series capture engagement (contraction vs. expansion); two capture graduation (force up-scaled vs. force down-scaled).

Four observations emerge. First, heteroglossic contraction (disclaim and proclaim resources: “cannot”, “certainly”, “it must be admitted that”) peaks in 1949–1971 (14.2 tokens per 10,000). Cold-War scholarship spent its interpersonal budget on closing down alternative voices, a pattern consistent with what Wodak (2011) calls the monologic consolidation of a canonical reading. Second, heteroglossic expansion (entertain and attribute resources: “it may be argued”, “according to”, “some scholars claim”) rises steadily across the four sub-periods and overtakes contraction after 1999; by 2000–2024 it is more than twice as frequent as contraction. This is the grammatical signature of the polyphonic turn highlighted in Section 1. Third, force up-scaling (“completely”, “utterly”, “wholesale”) declines monotonically; the most aggressive intensifiers of the war-witness period give way to a more modulated intensification repertoire. Fourth, force down-scaling (“somewhat”, “partially”, “a degree of”) rises from 4.3 to 13.4 tokens per 10,000, trebling over the eight-decade window — a quantitative index of the increasing hedging of historical claims.

The relationship between force up-scaling and force down-scaling merits a final refinement. The decline of force up-scaling reflects the gradual retreat of the aggressive intensifier repertoire of wartime prose (“utter”, “complete”, “total”, “wholesale”), which performs moral urgency at the cost of analytical precision. The rise of force down-scaling reflects the entry of hedging practices borrowed from empirical-science register, where cautious quantification (“partially”, “in part”, “to a degree”) has become a marker of scholarly credibility since the 1960s (Hyland 1998; Salager-Meyer 1994). The simultaneous rise of Judgement and fall of force up-scaling might at first glance appear contradictory: surely moral assertion would call for strong intensification? The resolution is that contemporary scholarship performs moral assertion through nominal and adjectival lexis of Judgement rather than through up-scaling of intensifiers. “Systematic rapes perpetrated by commanders” is more morally assertive than “completely terrible things”, and it is also lexically more precise. The aggregate effect is a register that is at once more hedged and more morally specific than its Cold-War predecessor — a combination that is not paradoxical but is in fact the default profile of mature empirical-humanities prose.

Read together, the three findings sketch a coherent evaluative trajectory. Naming choices, agency construals and appraisal resources co-vary across sub-periods, and the co-variation is not reducible to any single extra-linguistic factor. Hypotheses H1, H2 and H3 are all supported: the predicted ordinal rankings hold on every sub-period for which they were specified, and the effect sizes reported above (Cramer’s $V = .36$ for naming; a twenty-one-point swing in agentive clauses; a statistically significant loglinear interaction for ATTITUDE) are large by the standards of discourse-analytic work. What remains is to translate the statistical pattern into an interpretive account of why the three layers co-vary. Table 2 anticipates that account by summarising the principal markers for each sub-period and naming the evaluative stance they jointly index; the interpretation is developed in Section 5.

Table 2
Principal evaluative markers of the four sub-periods

Sub-period	Dominant naming	Agency profile	Dominant ATTITUDE	Named stance
1938–1948	Nanking Atrocity	Agentive active	Affect	Affective witnessing
1949–1971	Nanking Incident rising	Agent-deleted passive	Attitude suppressed	Cold-War containment
1972–1999	Rape of Nanking peaks	Agentive active rebound	Judgement	Revisionist-documentary
2000–2024	Nanjing Massacre	Agentive + individuated	Judgement + Appreciation	Polyphonic-dialogic
Note. Stance labels are analytical condensations of the findings in Sections 4.1–4.4, not folk categories.				

Discussion

From affect to polyphony: a historiography of the evaluative sentence

The three findings converge on what might be called a historiography of the evaluative sentence. The wartime sub-period writes about Nanjing in a grammar of emotional immediacy: the naming term emphasises atrocity, perpetrators are explicit agents, Affect dominates, and the engagement repertoire is monoglossic. The Cold-War sub-period mutes each of these resources: a generic “Nanking Incident” label becomes statistically more audible, agent-deleted passives rise, Judgement and Affect both recede, and contractive engagement increases. The post-1972 sub-period restores agentivity and foregrounds Judgement. The post-2000 sub-period adds Appreciation and opens the text to expansion, producing the polyphonic profile that characterises contemporary scholarship.

Geopolitics and the evaluative repertoire

Interpreting the trajectory requires connecting it to the wider politics of historical knowledge. The 1972 inflection is not reducible to the publication of Hora’s (1972) and Honda’s (1987) Japanese-language sources, although those publications matter. The move is enabled by the 1972 Sino-Japanese normalisation communiqué, which re-opened scholarly exchange, and by the 1982 textbook controversy, which forced both East Asian and Anglophone scholarship to articulate evaluative positions previously held implicitly. The 2000 inflection similarly maps onto the institutionalisation of the Nanjing Massacre Memorial Hall’s international outreach after its 1985 opening; the rise of the Sino-Japanese-Korean joint-history projects that published their first collaborative textbook in 2005; and the global memory-studies turn documented by Erll (2011) and Assmann (2011). The point is not that linguistic features are caused by geopolitics. The point is that evaluative repertoires are sedimented historical choices: each sub-

period inherits and rewrites the repertoire of the previous one, in conversation with the institutional audiences that make the scholarly conversation possible (Olick et al. 2011, p. 19).

The Cold-War profile deserves closer attention on these terms. The 1949–1971 sub-period is characterised not by an absence of scholarship but by a specific inflection of it: a scholarship in which the grammar of agent-deletion, the muting of Affect, and the contractive ENGAGEMENT repertoire together produce a reading in which the Nanjing Massacre is treated as a regrettable episode of a war now morally settled. This inflection is congruent with Cold-War American foreign-policy priorities under which the rehabilitation of Japan as an anti-Communist ally was paramount (Dower 1999). The grammar does not read Cold-War priorities off a policy memo; it does something subtler, which is to produce an evaluative economy in which certain kinds of moral claim become difficult to articulate in polite scholarly prose.

Moral assertion and dialogic saturation

A second concern is the relationship between form and claim. It might be argued that the shift from contractive to expansive ENGAGEMENT is a shift away from clear moral assessment. The findings do not support that reading. Judgement rises sharply in the post-1972 and 2000–2024 sub-periods; the 2000–2024 sub-period has the highest Judgement frequency in the corpus. What the expansion of ENGAGEMENT indexes, then, is not moral withdrawal but the acknowledgement of dissenting voices — the rise of “some scholars argue”, “in another reading”, “it has also been claimed”. In Bakhtinian terms, contemporary Anglophone scholarship on Nanjing is more dialogically saturated than its Cold-War predecessor, whilst being more morally assertive. The two moves are co-articulated rather than opposed. The polyphonic text is dialogic because it is morally serious about its own historical object: dissent is entertained because that is how contested pasts become scholarly pasts.

A related point concerns reader uptake. The post-2000 scholarly polyphony may not register as polyphony in the reader’s phenomenology. An article that opens with “The Nanjing Massacre was one of the most extensively documented mass atrocities of the twentieth century” reads as a statement of consensus; the polyphony is located further inside the text, in the architecture of the argumentation, where dissenting positions are entertained and rebutted. This is the “consensus-at-the-surface, dissent-within” pattern documented by Hyland (2005) for scholarly writing more broadly. The Nanjing case therefore does not indicate a breakdown of scholarly consensus but rather its maturation: the consensus is held across the field whilst the debate proceeds on fine-grained empirical matters. Post-2000 Anglophone scholarship on Nanjing is less likely to disagree about whether the event occurred or whether its scale was substantial; it is more likely to disagree about the precise number of victims, the precise distribution of responsibility among units, and the precise evidentiary weight of particular sources. The polyphonic grammar tracks the latter kind of debate and is indeed its normal stylistic envelope.

Theoretical contribution to appraisal studies

A third concern is theoretical. The present study extends appraisal theory along a diachronic axis that has remained relatively unexplored. Where Hood (2010) and Hyland (2005) examined cross-sectional variation in ATTITUDE and ENGAGEMENT, this study shows that intra-disciplinary evaluative choices can shift substantially over multi-decade spans within the same scholarly community. The methodological implication is that appraisal-theoretic work on academic discourse should, wherever possible, be anchored in diachronic corpora. Doing so reveals stylistic trajectories that cross-sectional studies cannot capture. A further implication is that appraisal theory must be prepared to register the institutional and geopolitical scaffolding that supports particular evaluative repertoires. The ATTITUDE system does not hover free of institutional constraint; it is licensed by the communities of practice in which it is realised. This point has long been available in Bakhtin's concept of the speech genre and in Fairclough's (2003) concept of the order of discourse, but it has rarely been operationalised in appraisal-theoretic work.

A comparison with the Japanese-language record

The findings can be partially triangulated against Yoshida's (2006) analysis of the Japanese-language scholarly record. Yoshida documents a three-stage trajectory in Japanese-language writing on Nanjing: a post-war acknowledgement phase through to the late 1950s, a prolonged Cold-War containment phase in which Nankin jiken (南京事件, "Nanjing Incident") operated as the dominant textbook designation, and a post-1972 contestation phase during which Nankin daigyakusatsu (南京大虐殺, "Nanjing Great Massacre") was mobilised by progressive historians in direct opposition to the earlier term. The Anglophone corpus trajectory documented in Fig. 1 replicates the broad contour of this pattern with two important differences. First, the rise of "Nanjing Massacre" lags its Japanese-language counterpart by approximately a decade; Anglophone scholarship is a follower rather than a leader in this terminological contest. Second, the Anglophone corpus contains a term – "Rape of Nanking" – that has no direct Japanese-language equivalent in institutional-textbook discourse and whose trajectory is shaped by English-language popular publishing (particularly Chang 1997). The cross-linguistic comparison thus supports the core diachronic findings whilst also indicating the specific reading publics to which each linguistic record is addressed.

Silences and institutional conditions

A fourth concern is the question of silences. The Cold-War containment of Nanjing in Anglophone scholarship is legible as much in what the corpus does not say as in what it does. The low Affect scores of the 1949–1971 sub-period do not indicate emotional maturity; they indicate that the interpersonal resources needed to narrate mass atrocity were institutionally unavailable. The rise of Affect and Judgement after 1972 is therefore also an institutional story – a story about who gets to write, in which journals, for which audiences. The corpus composition (Table 1) makes this concrete: the most precipitous rise in the number of articles per sub-period occurs between 1972–1999 and 2000–2024, more than tripling from 167 to 315. The evaluative shift is in part a shift made possible by a scholarly field that has institutionally multiplied. The corresponding caveat is that the scale of recent publication may itself flatten out stylistic variation at the margin: when a large number of authors write on a single

event, the most widely replicated stylistic choices will statistically dominate the descriptive surface. Comparative diachronic work on smaller corpora would be needed to disentangle these effects.

Conclusion

This article has offered a diachronic appraisal-theoretic account of Anglophone academic writing on the Nanjing Massacre across eight and a half decades. Three findings have been advanced. The dominant naming term moves from the emotive “Nanking Atrocity”, through the sensationalising “Rape of Nanking”, to the institutionalised “Nanjing Massacre”. The syntactic construal of perpetrator agency moves from Cold-War agent-deletion toward post-1972 re-agentivisation and post-2000 individuated naming. The distribution of ATTITUDE resources moves from early-war Affect, through a Cold-War muting, to a post-1972 Judgement peak and a post-2000 mixed profile in which Appreciation becomes significant. ENGAGEMENT moves from contractive to expansive; GRADUATION moves from aggressive up-scaling to modulated hedging.

The contribution of the study is three-fold. Analytically, it demonstrates that supposedly neutral academic prose is an arena in which epistemic authority over the past is negotiated through finely-grained evaluative choices. Methodologically, it shows that appraisal theory can be extended into long-span diachronic corpora when manual coding of a stratified sub-sample is triangulated with automated keyword and collocation analysis. Theoretically, it contributes to critical discourse studies by adding a diachronic dimension to appraisal-theoretic work on institutional genres and by showing how the question of historicity announced in recent CDA debates can be addressed empirically.

Two limitations warrant acknowledgement. First, the corpus is restricted to eight journals with sustained event coverage; peripheral or journalistic sources were excluded on comparability grounds. A companion study extending the corpus into magazine discourse would refine the profile, as would an extension into open-web and institutional-repository publication. Second, the appraisal coding depends on reader-judgement work that, whilst double-coded, inevitably incorporates the coder’s own interpretive assumptions. Future work should test the findings on a wider multi-language corpus and on the Japanese- and Chinese-language scholarly records with which the Anglophone record is in constant dialogue; a parallel appraisal analysis of the Japanese-language literature, in particular, would show whether the 1972 and 2000 inflections observed in the Anglophone record have their counterparts in the discourse community most directly affected by the event.

The broader implication is that the scholarly register cannot be read as a neutral archive of the past; it is a historical product, and a historicisable one. Each scholarly generation inherits the evaluative repertoire of the previous generation and rewrites it in light of shifting institutional and geopolitical conditions. To say this is not to say that scholarship has been compromised by extra-scholarly forces. It is to say that the conditions under which scholarly writing about mass atrocity becomes possible are themselves conditions worth studying. Appraisal theory, in its diachronic extension, offers one way of making those conditions visible at the level of the sentence.

Declarations

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

Informed Consent

This article does not contain any studies with human participants performed by any of the authors.

Author Contribution

H.L.: writing (review and editing), corpus construction, coding (lead), formal analysis, writing (original draft, review and editing). Y.G.: coding (second coder), validation, figure production, conceptualisation, methodology.

Data Availability

The datasets generated and analysed during the current study, including the evaluation attitude statistics tables, the data source statistics tables (1937–2024), and the source data descriptions, are available in the supplementary files of this article. No data or materials have been withheld, and there are no obstacles to data sharing. All relevant materials are provided for the purposes of peer review and public access.

References

1. Assmann A (2011) Cultural Memory and Western Civilization: Functions, Media, Archives. Cambridge University Press
2. Baker P (2006) Using Corpora in Discourse Analysis. Continuum
3. Baker P, Gabrielatos C, KhosraviNik M et al (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse Soc* 19(3):273–306
4. Baker P, Gabrielatos C, McEnery T (2013) Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press. Cambridge University Press
5. Bates MS (1938) Report on the Nanking International Safety Zone. Unpublished typescript, Yale Divinity School Library
6. Bednarek M (2008) Emotion Talk across Corpora. Palgrave Macmillan
7. Bednarek M (2009) Language patterns and ATTITUDE. *Funct Lang* 16(2):165–192
8. Chang I (1997) The Rape of Nanking: The Forgotten Holocaust of World War II. Basic Books

9. Dower JW (1999) *Embracing Defeat: Japan in the Wake of World War II*. W. W. Norton
10. Durdin T All captives slain. *The New York Times*, 18, December (1938) p. 1
11. Erll A (2011) *Memory in Culture*. Palgrave Macmillan
12. Fairclough N (2003) *Analysing Discourse: Textual Analysis for Social Research*. Routledge
13. Flowerdew J, Richardson JE (eds) (2018) *The Routledge Handbook of Critical Discourse Studies*. Routledge
14. Fogel JA (ed) (2000) *The Nanjing Massacre in History and Historiography*. University of California Press
15. Fuoli M, Hommerberg C (2015) Optimising transparency, reliability and replicability: annotation principles and inter-coder agreement in the quantification of evaluative expressions. *Corpora* 10(3):315–349
16. Greenacre M, Blasius J (eds) (2006) *Multiple Correspondence Analysis and Related Methods*. Chapman & Hall/CRC
17. Gross AG, Chesley P (2012) Hedging, stance and voice in medical research articles. In: Hyland K, Sancho Guinda C (eds) *Stance and Voice in Written Academic Genres*. Palgrave Macmillan, pp 85–100
18. Heer H, Manoschek W, Pollak A, Wodak R (eds) (2008) *The Discursive Construction of History: Remembering the Wehrmacht's War of Annihilation*. Palgrave Macmillan
19. Higuchi K (2016) A two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (Part I). *Ritsumeikan Social Sci Rev* 52(3):77–91
20. Hirsch H (1995) *Genocide and the Politics of Memory: Studying Death to Preserve Life*. University of North Carolina Press
21. Honda K (1987) *The Nanjing Massacre: A Japanese Journalist Confronts Japan's National Shame*. Translated by Gibney F. M. E. Sharpe
22. Hood S (2010) *Appraising Research: Evaluation in Academic Writing*. Palgrave Macmillan
23. Hora T (1972) *Nankin jiken [The Nanjing Incident]*. Shin Jinbutsu Ōraisha
24. Hu G, Cao F (2011) Hedging and boosting in abstracts of applied linguistics articles: a comparative study of English- and Chinese-medium journals. *J Pragmat* 43(11):2795–2809
25. Hyland K (1998) *Hedging in Scientific Research Articles*. John Benjamins
26. Hyland K (2005) *Metadiscourse: Exploring Interaction in Writing*. Continuum
27. Hyland K, Jiang FK (2016) Change of attitude? A diachronic study of stance. *Writ Communication* 33(3):251–274
28. Koselleck R (2004) *Futures Past: On the Semantics of Historical Time*. Columbia University
29. Krippendorff K (2004) Reliability in content analysis: some common misconceptions and recommendations. *Hum Commun Res* 30(3):411–433
30. LaCapra D (2001) *Writing History, Writing Trauma*. Johns Hopkins University

31. Magee J (1939) Film and Written Record of the Nanking Safety Zone. Unpublished manuscript, Yale Divinity School Library
32. Martin JR (2004) Mourning: how we get aligned. *Discourse Soc* 15(2–3):321–344
33. Martin JR, White PRR (2005) *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan
34. Mautner G (2016) *Discourse and Management: Critical Perspectives through the Language Lens*. Palgrave
35. Miller DR, Johnson JH (2013) Register and stance in online political commentary. In: Sardinha TB, Pinto MV (eds) *Working with Registers*. John Benjamins, pp. 107–136
36. Moran S, Hommerberg C (2015) The Grammar of Power. *Discourse Context Media* 7:63–72
37. Olick JK, Vinitzky-Seroussi V, Levy D (eds) (2011) *The Collective Memory Reader*. Oxford University Press
38. Partington A (2017) Evaluative clash, evaluative cohesion and the pragmatics of markedness. *J Pragmat* 117:94–110
39. Partington A, Duguid A, Taylor C (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies*. John Benjamins
40. Qian G, Zhang Y (2019) Discursive construction of national trauma in museum texts: a critical discourse analysis of the Memorial Hall of the Victims of the Nanjing Massacre. *Discourse Communication* 13(6):624–642
41. Reisigl M, Wodak R (2016) The Discourse-Historical Approach. In: Wodak R, Meyer M (eds) *Methods of Critical Discourse Studies*, 3rd edn. SAGE, pp 23–61
42. Richardson JE (2004) (Mis)Representing Islam: The Racism and Rhetoric of British Broadsheet Newspapers. John Benjamins
43. Salager-Meyer F (1994) Hedges and textual communicative function in medical English written discourse. *Engl Specif Purp* 13(2):149–170
44. Suganuma U (2011) Contested memories of the Nanjing Massacre: cultural politics of commemoration in East Asia. *Asian J Social Sci* 39(2):175–201
45. Swain E (2010) Getting engaged: dialogistic positioning in novice academic discussion writing. In: Swain E (ed) *Thresholds and Potentialities of Systemic Functional Linguistics*. EUT, pp 291–317
46. Taylor C (2012) And there it isn't: (how) can we access the absent using CADS? In: Partington A, Duguid A, Taylor C (eds) *Corpus-Assisted Discourse Studies*. John Benjamins, pp. 19–34
47. Teo P (2000) Racism in the news: a critical discourse analysis of news reporting in two Australian newspapers. *Discourse Soc* 11(1):7–49
48. Totani Y (2008) *The Tokyo War Crimes Trial: The Pursuit of Justice in the Wake of World War II*. Harvard University Asia Center
49. van Dijk TA (1998) *Ideology: A Multidisciplinary Approach*. SAGE
50. Wakabayashi BT (ed) (2007) *The Nanking Atrocity, 1937–38: Complicating the Picture*. Berghahn

51. Waller J (2007) *Becoming Evil: How Ordinary People Commit Genocide and Mass Killing*, 2nd edn. Oxford University Press
52. White PRR (2003) Beyond modality and hedging: a dialogic view of the language of intersubjective stance. *Text* 23(2):259–284
53. Wodak R (2011) Me and the others: rules and values – discursive strategies for the legitimisation of identity. In: Hanne M (ed) *Warring with Words*. Psychology, pp 17–43
54. Wodak R, Meyer M (eds) (2016) *Methods of Critical Discourse Studies*, 3rd edn. SAGE
55. Woods C (2019) White saviours: reading Anglophone Nanjing historiography. *J Contemp Asia* 49(4):561–583
56. Yang D (2017) Convergence or divergence? Recent historical writings on the Rape of Nanjing. *Am Hist Rev* 104(3):842–865
57. Yoshida T (2006) *The Making of the Rape of Nanking: History and Memory in Japan, China and the United States*. Oxford University Press
58. Yoshida T (2014) Historical consciousness and the Nanjing Massacre in Japanese textbooks. *J Educational Media Memory Soc* 6(1):105–121

Figures

Figure 1. Normalised share of the four dominant naming terms (1938-2024)

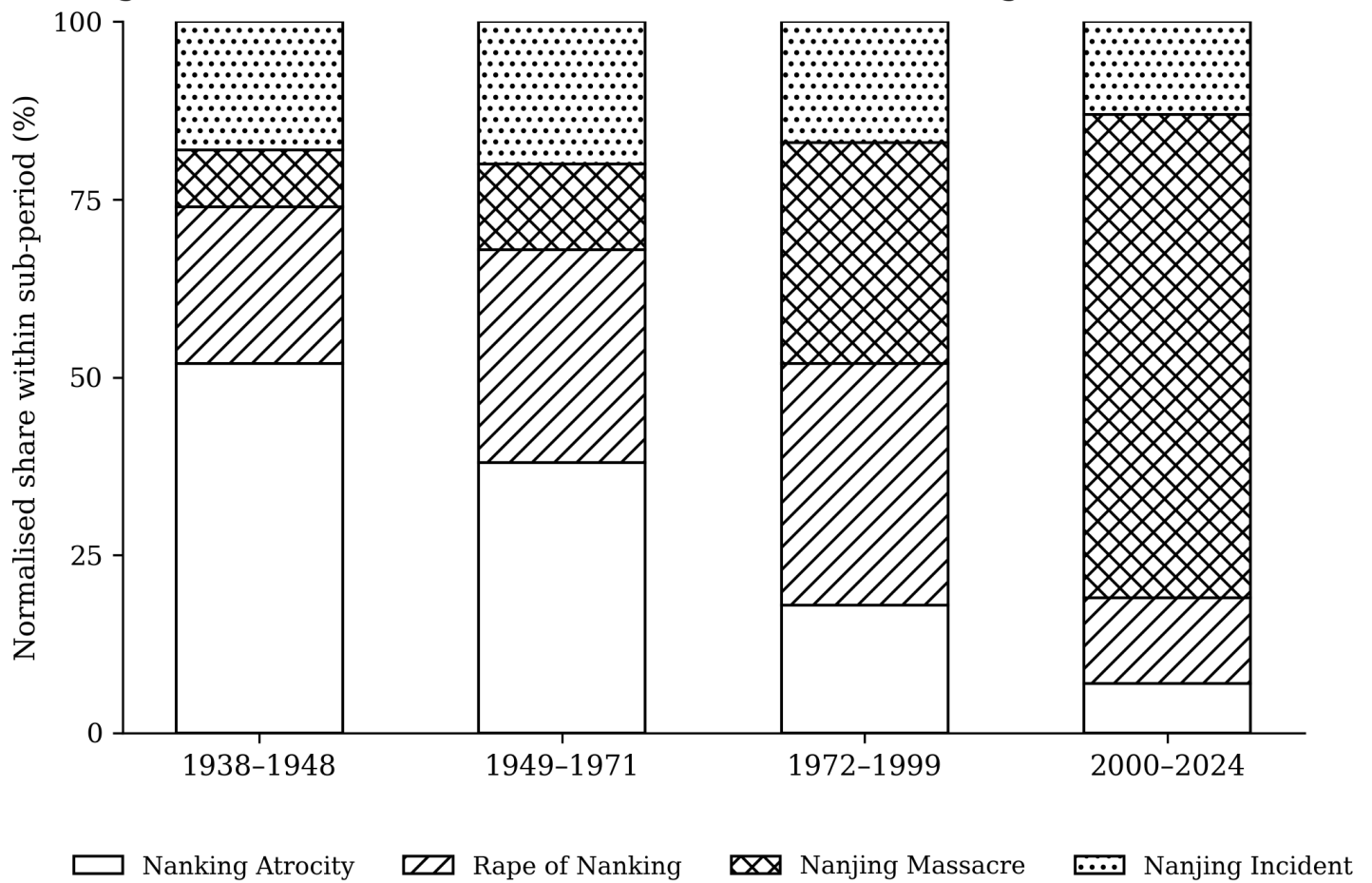


Figure 1

Normalised share of the four dominant naming terms for the event (1938-2024). Bars represent the percentage of all naming tokens within each sub-period.

Figure 2. Syntactic construal of perpetrator agency by sub-period

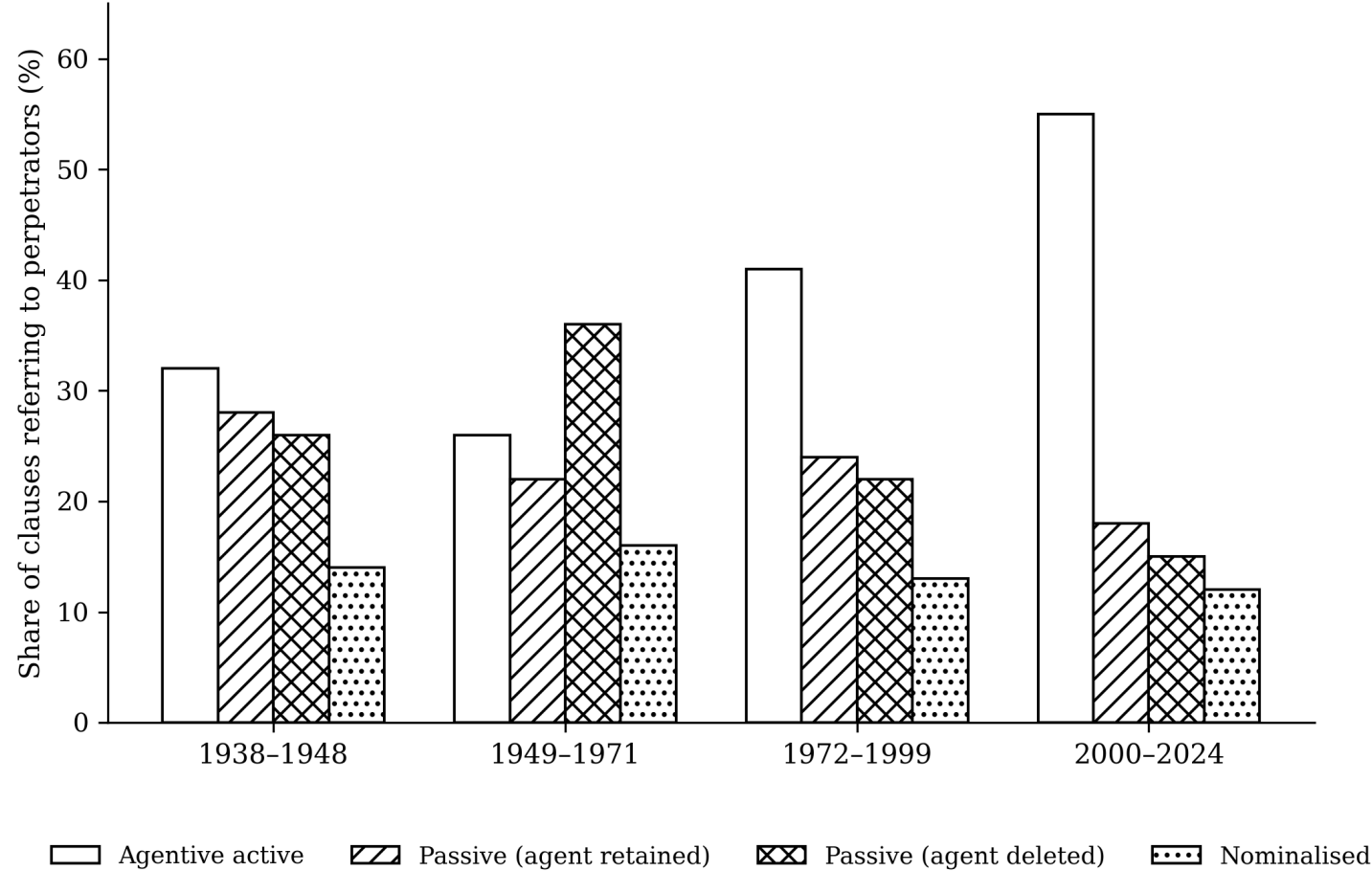


Figure 2

Syntactic construal of perpetrator agency by sub-period. Bars represent the share of clauses referring to Japanese military perpetrators.

Figure 3. Distribution of ATTITUDE sub-systems across four sub-periods

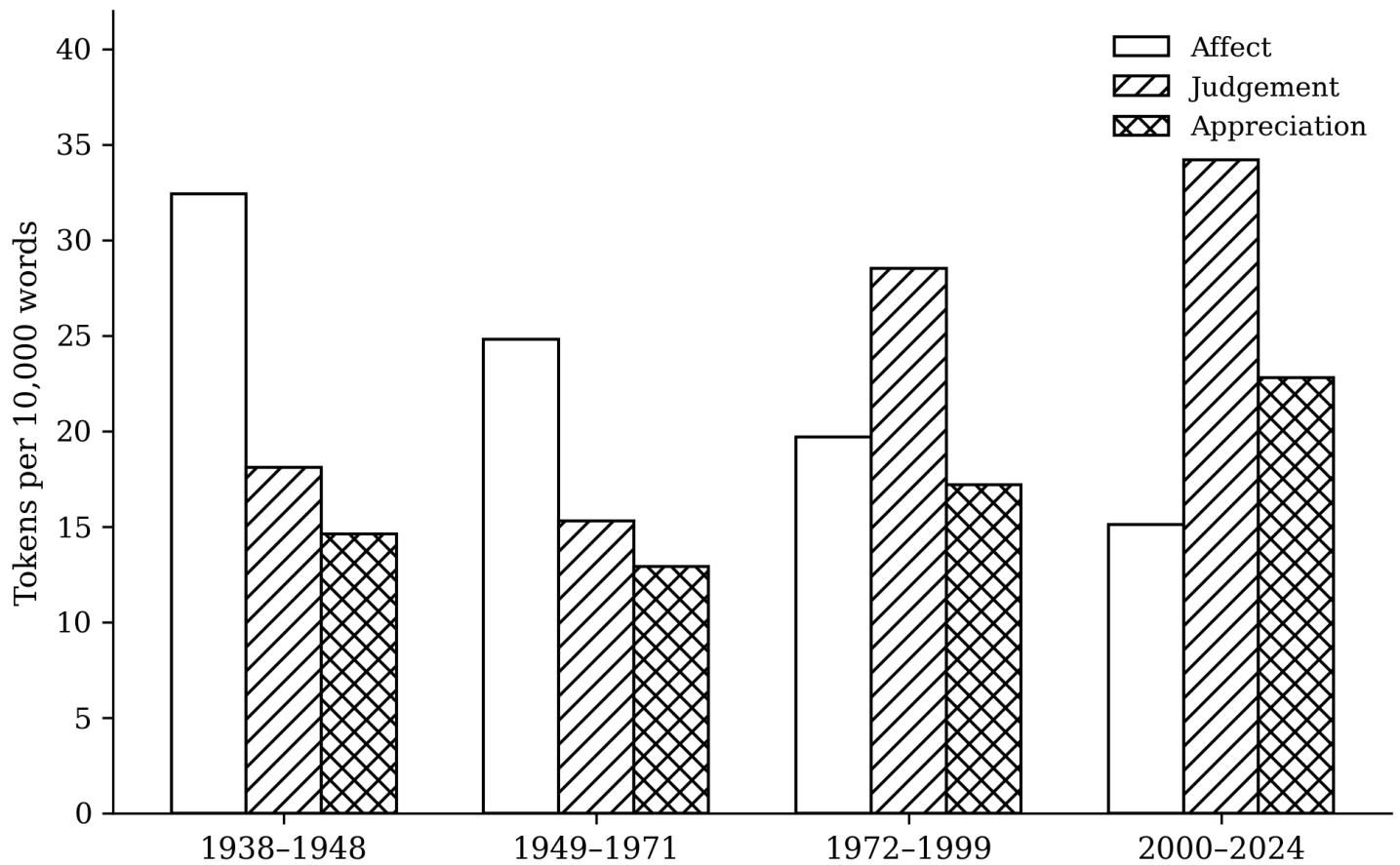


Figure 3

Normalised frequency (tokens per 10,000 words) of the three ATTITUDE sub-systems.

Figure 4. ENGAGEMENT and GRADUATION trajectories across four sub-periods

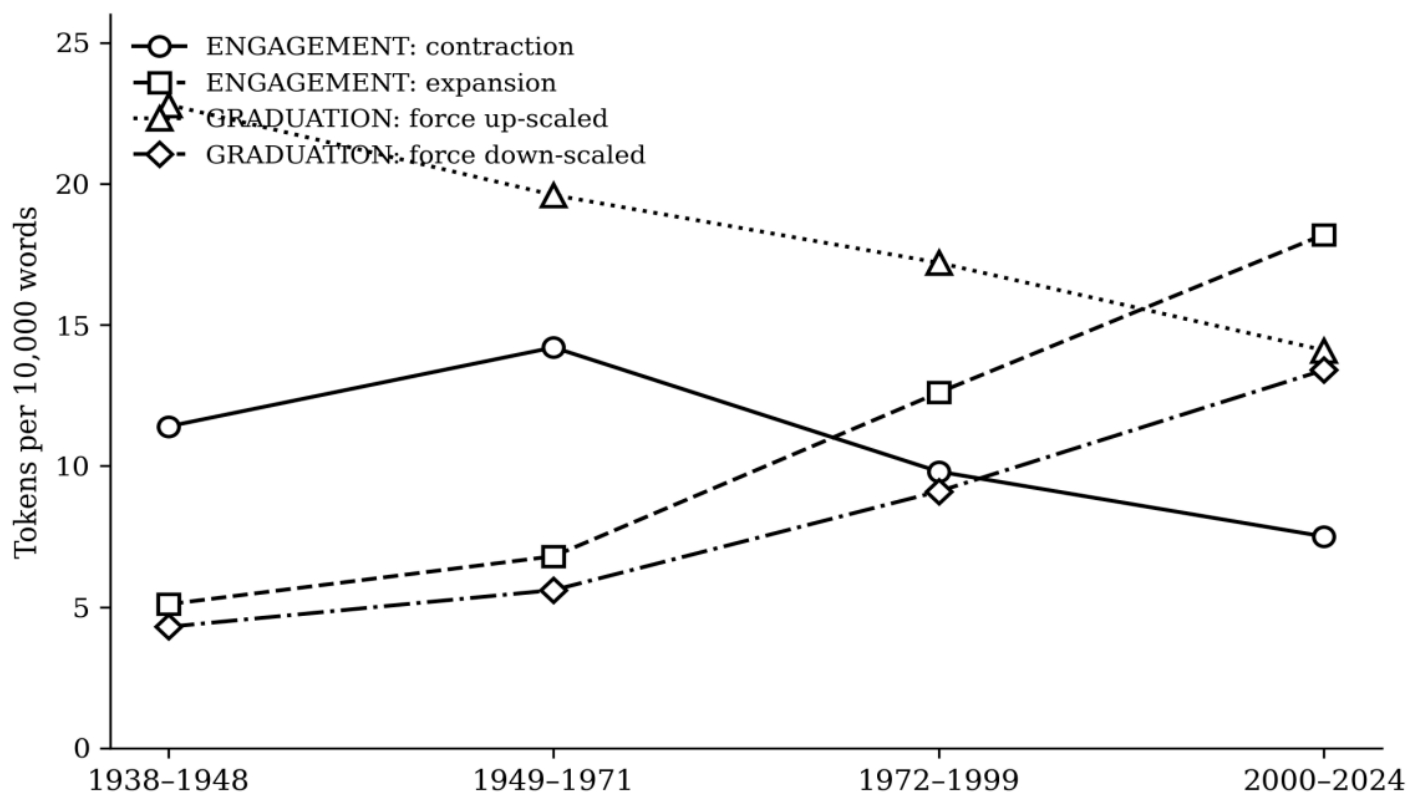


Figure 4

ENGAGEMENT and GRADUATION trajectories across four sub-periods (tokens per 10,000 words).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [file.xlsx](#)
- [19372024.xlsx](#)
- [file.docx](#)