

Open
AccessCheck for
updates

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

From workshop talk to workplace narratives: A corpus-assisted framework for labour, technology, and technical discourse in contemporary English literature

Abdul-Wasea Abdul-Ghani Saif Al-Mekhlafi¹ & Mohammad Osman Abdul Wahab²¹Faculty of Business Administration, Prince Sattam bin Abdulaziz University, Saudi Arabia²Department of English, Faculty of Languages and Translation, King Khalid University, Saudi Arabia*Correspondence: a.almekhlafi@psau.edu.sa

ABSTRACT

Contemporary English literature increasingly imagines work through screens, automated systems, logistics networks, algorithmic supervision, precarious service economies, and hybrid human-machine environments. Yet literary criticism often approaches these representations through broad thematic labels—technology, alienation, labour, surveillance, or posthumanism—without systematically examining the language through which fictional workplaces are made audible and ethically legible. This article develops a corpus-assisted research framework for analysing technical and workplace discourse in contemporary English fiction. It is presented transparently as a research-design study because the revised project does not claim a completed literary corpus or empirical findings. The proposed Contemporary English Workplace Literature Corpus (CEWLC) is a stratified collection of novels, short fiction, and dramatic texts published in English from 1990 to 2025 and selected for sustained representation of industrial, digital, service, logistical, scientific, or automated work. Corpus stylistics is integrated with narratology, speech and thought presentation, critical stylistics, labour-oriented literary criticism, and close reading. The framework specifies corpus construction, copyright-sensitive text preparation, metadata, narrative-zone segmentation, annotation of occupational roles and communicative events, keyword and collocation analysis, agency and modality, metaphor, characterisation, speech presentation, and inter-coder reliability. It also distinguishes fictional workplace discourse from authentic occupational communication: literary dialogue is stylised, narratively distributed, and shaped by genre, focalisation, and aesthetic design. The study's contribution is methodological and interpretive. It offers a reproducible way to examine how contemporary English literature represents authority, risk, expertise, automation, responsibility, and resistance, while providing a basis for literature curricula that connect close reading with the cultural politics of technological work.

KEYWORDS: contemporary English literature, corpus stylistics, workplace narratives, labour, technology, Industry 4.0, narratology, digital humanities

Research Journal in Advanced Humanities

Volume 7, Issue 3, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 21 June 2026

Accepted: 11 July 2026

Published: 23 August 2026

HOW TO CITE

Saif Al-Mekhlafi, A.-W. A.-G., & Wahab, M. O. A. (2026). From workshop talk to workplace narratives: A corpus-assisted framework for labour, technology, and technical discourse in contemporary English literature. *Research Journal in Advanced Humanities*, 7(3). <https://doi.org/10.58256/tbrk4703>



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2026 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Work is one of literature's most persistent social environments, yet it is not always treated as a linguistic environment. Factories, laboratories, offices, call centres, warehouses, kitchens, hospitals, platforms, and automated control rooms are commonly read as settings that signify class, modernity, alienation, discipline, or technological change. Such readings are indispensable, but they can leave the discourse of work analytically underdescribed. Fictional workers do not merely occupy workplaces. They issue and resist instructions, translate specialised knowledge, report failure, conceal uncertainty, negotiate hierarchy, narrate risk, interpret data, and decide who can speak with authority. The verbal texture of these scenes helps literary texts convert economic structures into interpersonal encounters and abstract systems into morally charged events.

This issue has become more urgent in literature concerned with what policy and management discourse often calls Industry 4.0: automation, cyber-physical systems, algorithmic coordination, sensor-based production, platform labour, remote monitoring, and human-machine collaboration. The expression is technologically expansive and ideologically unstable. It can designate a genuine reorganisation of work, a managerial narrative of inevitable progress, or a marketing vocabulary that makes older forms of labour control appear unprecedented. Literature is especially valuable because it does not have to accept the vocabulary of innovation on its own terms. Fiction can expose the hidden labour beneath automation, dramatise the unequal distribution of risk, and show how technical systems alter attention, time, memory, embodiment, and interpersonal trust.

The source manuscript from which the present revision develops was centred on communication in Saudi technical education and proposed a corpus of workshop interaction, safety briefings, technical writing, and employer interviews. Its methodological insistence on observable language, stratified evidence, and non-fabrication remains valuable. However, changing the disciplinary focus to English literature cannot be achieved by inserting the word "literature" into the title. The object of study, corpus architecture, research questions, theoretical framework, units of analysis, ethical concerns, and interpretive claims must all change. Authentic workshop communication and fictional workplace discourse are not interchangeable. Fictional dialogue is selected, compressed, stylised, framed by narration, and assigned aesthetic and ideological functions. A warning in a novel may perform a technical action within the story while simultaneously foreshadowing disaster, characterising a supervisor, or exposing an institution's moral failure.

This article therefore reconceptualises the project as a corpus-assisted study of workplace representation in contemporary English literature. It proposes the Contemporary English Workplace Literature Corpus (CEWLC), a balanced literary corpus designed to capture sustained representations of industrial, technological, scientific, service, logistical, and automated labour. The study combines corpus stylistics with narratology, critical stylistics, speech and thought presentation, and literary criticism of labour and technology. Corpus procedures identify recurrent patterns across texts; close reading explains how those patterns function within scenes, plots, character systems, and genres. This movement between recurrence and singularity is central to corpus stylistics, which supplements rather than replaces literary interpretation (Mahlberg & McIntyre, 2011; McIntyre, 2015; Stockwell & Mahlberg, 2015).

The revised study asks six questions. First, what workplace settings, occupational roles, communicative events, and technological objects recur across the selected literary corpus? Second, what lexical, grammatical, metaphorical, and phraseological patterns construct work, expertise, efficiency, risk, and automation? Third, how do narrative voice, focalisation, and speech or thought presentation distribute authority and vulnerability among workers, managers, experts, machines, and institutions? Fourth, how are obligation, permission, uncertainty, responsibility, and blame linguistically represented in scenes of instruction, troubleshooting, reporting, and failure? Fifth, how do different literary genres and workplace domains vary in their representation of technological labour? Sixth, how can the corpus support English-literature teaching in Saudi universities without reducing literary texts to vocational communication manuals?

The central argument is that workplace language in fiction is not decorative realism. It is a formal resource through which literature organises agency, makes systems perceptible, and tests the ethical claims of technological modernity. A corpus-assisted design can make recurrent features visible, but interpretation must remain sensitive to narrative mediation, genre, historical context, and the asymmetry between literary representation and empirical workplace practice.

2. From Industrial Setting to Literary Workplace

Industrial literature has never been limited to descriptions of machinery. From nineteenth-century factory fiction to contemporary narratives of code, finance, extraction, care, and logistics, literature has linked work to questions of time, bodily discipline, social visibility, and moral responsibility. The contemporary workplace differs from earlier industrial settings not because manual labour has disappeared, but because material and informational labour are increasingly interdependent. A warehouse picker may be directed by an interface; a care worker may be evaluated through metrics; a programmer may maintain invisible infrastructure; a technician may rely on automated diagnostics; and a manager may exercise authority through dashboards rather than direct commands. Fiction can render these distributed relations by moving among dialogue, interiority, description, documents, and technologically mediated messages.

The literary workplace should therefore be treated as a communicative ecology. It includes direct speech among characters, represented emails and reports, manuals and alerts, narrated routines, internal monologue, sensory description, and the silent authority of technical artefacts. A machine can become narratively agential without literally speaking: alarms interrupt action, interfaces restrict choices, and automated decisions reorganise human relationships. The analysis must ask not only what characters say but how narration attributes intention, certainty, causality, and responsibility.

Industry 4.0 narratives often privilege novelty, speed, integration, and efficiency. Contemporary fiction may repeat this vocabulary, parody it, or place it in tension with fatigue, redundancy, insecurity, maintenance, and error. The literary significance of technical language lies partly in this tension. Specialist terminology may establish realism and expertise, but it can also exclude readers or characters, dramatise institutional opacity, and conceal moral decisions behind procedural phrasing. Passive constructions, nominalisations, and impersonal formulations can make harmful outcomes appear system-generated rather than humanly chosen. Conversely, direct naming of agents and consequences can transform a technical failure into an ethical accusation.

A literature-centred study must resist two simplifications. The first is technological determinism: the assumption that machinery or algorithms automatically produce particular social effects. Literary texts often show that technology works through institutions, hierarchies, habits, and unequal access. The second is documentary literalism: the assumption that fictional workplace scenes should be judged primarily by factual accuracy. Technical plausibility matters, especially when a text claims realism, but fiction also uses compression, symbolism, irony, and estrangement. An unrealistic machine may still organise a powerful critique of automation, while an accurate procedural description may serve an ideological rather than informational purpose.

The corpus design consequently separates the represented world from the representational method. Metadata describe workplace domain, period, genre, and technological intensity; annotation identifies communicative events and occupational roles; stylistic analysis examines how those events are narrated. This double attention allows the study to ask both what kinds of work appear and how literature makes them meaningful.

3. Literature Review

3.1 Corpus stylistics and literary interpretation

Corpus stylistics emerged from the recognition that computationally assisted pattern detection can strengthen literary analysis when it remains connected to textual function. Quantitative evidence is especially useful for identifying repeated clusters, distinctive collocations, patterns of speech presentation, and distributed characterisation that are difficult to perceive through unaided reading. Studies of Dickens have shown how recurrent lexical patterns, suspended quotations, body-language clusters, and character-centred concordances contribute to narrative design (Mahlberg, 2007, 2012; Mahlberg et al., 2014; Stockwell & Mahlberg, 2015). Work on Conrad, Joyce, Shakespeare, children's fiction, and colour symbolism similarly demonstrates that corpus procedures can test critical intuitions and reveal patterns that invite new interpretation (O'Halloran, 2007; Stubbs, 2005; Čermáková & Mahlberg, 2021; Wijitsopon, 2022).

The strength of corpus stylistics lies neither in scale alone nor in the claim that frequency determines significance. Literary effects often depend on local deviation, strategic rarity, or a pattern's distribution across narratorial and character zones. A phrase that appears only once may be structurally decisive; a frequent word may be stylistically unremarkable unless its collocational behaviour or narrative position is distinctive. For this reason, corpus stylistics requires movement between statistical salience and close contextual analysis (Mahlberg

& McIntyre, 2011; McIntyre, 2015). Dispersion is particularly important because a high frequency concentrated in one novel or one character should not be represented as a corpus-wide feature (Gries, 2008).

The proposed study extends this tradition to workplace representation. It does not treat fiction as a source of authentic occupational language. Instead, it examines how literary texts select and organise occupational discourse. Keyword analysis can identify the relative salience of risk, performance, failure, efficiency, and control; collocation can reveal the semantic associations of worker, system, machine, body, error, and data; speech-presentation annotation can show whose technical knowledge is quoted directly and whose is summarised or suppressed. These patterns become literary evidence only when interpreted through narrative context.

3.2 Speech, thought, and professional authority

Speech and thought presentation are central to literary workplace scenes because they determine whose understanding is available to the reader. Semino and Short's corpus-based account of speech, writing, and thought presentation provides a methodological basis for distinguishing direct, indirect, free indirect, and narratively reported forms (Semino & Short, 2004). In a technical scene, the choice among these forms has consequences. Direct speech can dramatise urgency, conflict, or expertise; indirect speech can subordinate a worker's voice to narratorial summary; free indirect discourse can merge institutional language with a character's internalised anxiety; represented documents can create an appearance of procedural objectivity.

Authority is also constructed through grammatical and pragmatic choices. Imperatives, deontic modals, conditionals, evidential markers, and technical nominalisations encode what must be done, who may decide, and how certainty is distributed. A supervisor's command may be formally direct but narratively undermined by focalised evidence of incompetence. A junior worker's hesitation may appear linguistically weak while proving ethically or technically correct. Literary analysis must therefore distinguish discourse status within the fictional institution from epistemic reliability within the narrative.

Characterisation research in corpus stylistics is useful here because workplace identity often accumulates through repeated linguistic cues rather than explicit description. Lexical bundles, reporting verbs, forms of address, interruptions, and recurrent metaphors can establish a character as competent, evasive, authoritarian, precarious, or resistant. Mind-modelling approaches further show how readers infer motives and dispositions from patterned textual traces (Stockwell & Mahlberg, 2015). The proposed corpus will examine whether technical experts are granted interiority, whether managers speak in abstract metrics, and whether machines are described through agentive verbs that displace human responsibility.

3.3 Labour, technology, and narrative form

Literary representations of technological work frequently oscillate between fascination and critique. Automation can be narrated as liberation from drudgery, as intensified surveillance, as an ecological burden, or as a restructuring of class and expertise. The language of inevitability is especially important. Expressions such as *must adapt*, *cannot stop*, *the system requires*, or *there is no alternative* can naturalise managerial choices. Fiction can reproduce this rhetoric or expose its contingency by juxtaposing institutional claims with embodied consequences.

The concept of labour must remain broad enough to include maintenance, care, communication, interpretation, and emotional regulation. Digital systems do not eliminate human work; they often redistribute it into less visible forms. A literary corpus that includes only factories and engineers would reproduce the narrowness it seeks to critique. Service work, domestic labour, logistics, medical and scientific work, platform labour, and clerical or managerial activity should therefore be represented alongside industrial production. The corpus should also include speculative texts in which automated systems intensify questions of human agency, but speculative fiction must not dominate the sample so completely that ordinary contemporary labour disappears.

Narrative form matters because different genres organise work differently. Realist fiction may foreground routine, institutional detail, and gradual erosion; satire may exaggerate managerial language; dystopia may turn efficiency into total control; drama may concentrate conflict in dialogue; short fiction may isolate a decisive failure or ethical choice. Genre comparison can therefore reveal not only different topics but different distributions of voice, pace, technical explanation, and causality.

3.4 Digital humanities, scale, and interpretive caution

Large-scale literary analysis has expanded the range of questions that can be asked about genre, cultural change, and textual influence. Digitised collections make it possible to compare extensive corpora, but their apparent abundance can conceal serious biases in publication, survival, copyright, language, and metadata. Research on Google Books has shown that massive scale does not remove sampling limitations (Michel et al., 2011; Pechenick et al., 2015). Canon-building corpora similarly require explicit inclusion principles because a corpus may reproduce inherited hierarchies under the appearance of computational neutrality (Green, 2017).

The CEWLC will therefore be purpose-built rather than assembled through unrestricted keyword harvesting. Texts must contain sustained workplace representation, not merely isolated references to jobs or machines. Copyright status, edition, genre, author background, publication market, and regional variety of English will be documented. The corpus will not claim to represent all contemporary English literature. Its purpose is analytical comparison across carefully defined subcorpora.

Computational methods also require interpretive discipline. Keyword lists and topic models can identify tendencies, but they cannot determine irony, focalisation, ethical evaluation, or aesthetic function without contextual reading. Machine-learning approaches in computational literary studies can support classification and exploratory patterning, yet their outputs depend on training data, feature selection, and operational definitions (Hatzel et al., 2023). The present design therefore privileges transparent corpus procedures and manual validation over opaque prediction.

3.5 English literature pedagogy and the Saudi context

A literary focus does not require abandoning the original project's concern with Saudi education. The pedagogical application changes, however, from technical communication training to English-literature teaching. Saudi English departments operate within a national context that places high value on technological transformation, economic diversification, and graduate employability. Literature courses can respond to this context without becoming instrumental language courses. Their distinctive contribution is to examine how narratives imagine work, authority, technological change, and human value.

Corpus-assisted literary pedagogy can support this task by making interpretive evidence visible. Students can compare concordances of machine, worker, error, efficient, safe, replace, or human across subcorpora; trace how a character's speech differs from narratorial description; examine whether passive grammar obscures responsibility; and test whether a perceived theme is distributed across a text or concentrated in a single scene. Such activities combine close reading with methodological transparency.

The pedagogical risk is reductionism. A novel should not be used merely as a vocabulary source for employment. Literary study concerns ambiguity, form, history, ideology, affect, and ethical imagination. The corpus framework is therefore designed to deepen interpretation rather than convert fiction into a simulation of workplace communication. Its value for Saudi curricula lies in connecting English literary analysis to contemporary questions of labour and technology while retaining literature's critical autonomy.

3.6 Research gap

Existing research offers strong models for corpus stylistics, characterisation, speech presentation, and large-scale literary analysis. Studies of technology and labour provide rich thematic criticism, while digital humanities expands comparative scale. What remains underdeveloped is an integrated framework for analysing workplace discourse as a literary phenomenon across a purpose-built contemporary corpus. Literary criticism often identifies automation, precarity, or surveillance at the thematic level without systematically examining the linguistic construction of instruction, error, risk, expertise, and responsibility. Workplace discourse research, by contrast, analyses authentic institutional communication but does not account for narrative mediation and aesthetic form.

The gap is therefore interdisciplinary but precise. A useful study must preserve the distinction between fictional and authentic discourse, connect corpus findings to narratological categories, treat technological language as both referential and stylistic, and examine how recurring workplace scenes participate in characterisation and plot. The CEWLC is designed to meet these requirements without presenting anticipated patterns as completed findings.

4. Conceptual Framework

The framework integrates five analytical traditions. Corpus stylistics identifies recurrent lexical, grammatical, and phraseological patterns and provides procedures for comparison across texts and subcorpora. Narratology explains how those patterns are distributed through narration, focalisation, character speech, represented writing, and temporal structure. Speech and thought presentation analysis classifies the forms through which workplace knowledge and consciousness become available. Critical stylistics examines agency, modality, naming, negation, presupposition, and hypothetical worlds as resources of ideological representation. Labour-oriented literary criticism situates textual patterns within broader questions of class, precarity, automation, expertise, and institutional power.

These traditions are assigned distinct tasks. Corpus analysis answers questions of recurrence and distribution. Narratology asks who perceives, knows, speaks, and interprets. Critical stylistics asks how linguistic choices naturalise or contest social arrangements. Labour criticism connects textual form to the historical and political meanings of work. Close reading integrates the evidence within specific scenes and whole-text structures.

The key concept is literary workplace discourse, defined as the ensemble of represented speech, thought, writing, narration, and material signs through which a literary text constructs occupational action, hierarchy, expertise, risk, and technological mediation. This definition includes dialogue and documents but also descriptions of tools, interfaces, bodies, routines, and institutional procedures. It differs from authentic workplace discourse because it is artistically selected, narratively framed, and oriented toward readers as well as fictional participants.

Five distinctions guide the analysis. Frequency concerns how often a form or event occurs. Distribution concerns whether it appears across authors, genres, and workplace domains or is concentrated in one text. Narrative salience concerns its location within plot, focalisation, and character development. Ideological force concerns how it supports, questions, or complicates a representation of work. Interpretive consequence concerns the extent to which the pattern changes the reader's understanding of agency, responsibility, or technological power. These dimensions may diverge. A rare warning can foreshadow catastrophe; a frequent efficiency term may become satirical through repetition; an absent worker voice may be more significant than a visible lexical pattern.

The framework also distinguishes machine agency from grammatical agency. A sentence may assign an active verb to a system—"the algorithm rejected," "the platform decided," "the machine demanded"—but this does not establish literal autonomy. Such grammar may personify technology, compress institutional responsibility, or reflect a character's limited understanding. Analysis must examine co-text, narrative stance, and alternative agents before interpreting technological personification.

5. Methodology

5.1 Research design and status

The study adopts a corpus-assisted comparative literary design combining quantitative corpus description with qualitative stylistic and narratological analysis. It is a research protocol rather than a report of completed findings. No corpus counts, literary quotations, statistical results, or interpretive conclusions are presented as empirical outcomes. The purpose of pre-specification is to make later analysis reproducible and to prevent a revised literary focus from resting on invented evidence.

The primary unit of inquiry is the literary work, with embedded units at the level of chapter, scene, narrative zone, character, workplace event, and lexical or grammatical feature. The design is comparative rather than statistically representative. It seeks analytical transfer across selected genres and workplace domains, not universal claims about contemporary English literature.

5.2 Corpus scope and selection

The corpus will be named the Contemporary English Workplace Literature Corpus (CEWLC). It will contain approximately 30 to 45 literary works published between 1990 and 2025, with a target size of 2.5 to 4 million words, depending on copyright access and genre balance. The date range captures late twentieth-century restructuring, digital expansion, platformisation, and contemporary automation while remaining manageable for contextual reading.

Eligible texts must be written in English or published in an authorised English version that is explicitly analysed as translated literature. Because translation introduces additional stylistic variables, translated works will form a marked subcorpus rather than being silently pooled with works originally written in English. The main corpus will include novels, novellas, short-story collections, and dramatic texts in which work is a sustained narrative environment. A text qualifies when occupational action, institutional hierarchy, technical systems, or workplace communication contributes materially to plot, characterisation, or thematic structure.

Selection will use purposive maximum-variation sampling. The corpus will seek balance across workplace domains, literary genres, author gender, geographical varieties of English, publication periods, and degrees of technological intensity. Best-selling and prize-recognised texts may be included, but prestige will not be the sole criterion. Small-press and genre fiction should be considered where they provide substantial representations of work absent from the mainstream literary canon.

Exclusion criteria include texts in which employment is only biographical background; works dominated by abstract discussion of technology without workplace scenes; non-fiction memoir unless used as a comparison corpus; and texts unavailable in a legally analysable digital form. The study will maintain a screening log recording inclusion decisions and borderline cases.

5.3 Corpus architecture

The CEWLC will be divided into overlapping but analytically distinct subcorpora. The Industrial and Extractive Work Subcorpus will include manufacturing, energy, construction, mining, and maintenance settings. The Digital and Corporate Work Subcorpus will include software, finance, data, surveillance, administration, and platform management. The Service, Care, and Domestic Labour Subcorpus will include hospitality, retail, healthcare, education, domestic work, and emotional labour. The Logistics and Circulation Subcorpus will include warehouses, transport, delivery, shipping, and supply-chain work. The Scientific and Technical Work Subcorpus will include laboratories, engineering, research, medicine, and technical problem-solving. The Speculative Automation Subcorpus will include texts in which artificial agents, advanced automation, or cybernetic systems restructure work.

These categories are not genres and may overlap. A novel may belong to both logistics and speculative automation, for example. The metadata will record primary and secondary domains rather than force every text into one exclusive class. Genre categories—realist, satirical, dystopian, speculative, dramatic, and hybrid—will be coded separately.

The corpus will also be segmented by narrative zone. Narratorial prose, direct character speech, indirect speech, thought presentation, represented writing, embedded digital communication, and paratextual material will be stored as searchable layers. This segmentation is essential because workplace vocabulary in a manager's speech performs differently from the same vocabulary in ironic narration.

5.4 Text acquisition, copyright, and preprocessing

Texts will be acquired through licensed electronic editions, institutional databases, authorised digital archives, or publisher permissions. Copyrighted full texts will not be redistributed. The research repository will store only legally permissible materials, metadata, derived frequency tables, search scripts, and short quotations within fair-dealing or fair-use limits appropriate to publication jurisdiction.

Digital files will be converted to plain text while preserving chapter boundaries, paragraphing, speaker labels where available, and typographic signals relevant to speech or documents. Optical character recognition will be used only when necessary and every OCR-derived text will undergo manual quality checks. Front matter, publisher advertisements, tables of contents, indexes, and unrelated notes will be removed through a documented procedure. Hyphenation, apostrophes, quotation marks, and spelling variants will be normalised for search while the original text remains recoverable.

Tokenisation and lemmatisation will be conducted with tools appropriate to contemporary English prose. Automated part-of-speech tagging will be manually evaluated on a stratified sample because literary syntax, dialogue fragments, invented terms, and typography can reduce tagging accuracy. Proper names, technical abbreviations, and machine-generated strings will receive custom handling. Every processing decision will be archived in a reproducible script or change log.

5.5 Metadata

Each work will carry bibliographic and analytical metadata: author, publication year, edition, original language, genre, narrative person, dominant tense, setting, workplace domain, technological intensity, geographical location, author gender where publicly stated and ethically appropriate, total words, proportion of dialogue, and copyright status. Character metadata will include occupational role, hierarchical position, employment security, relation to technology, and narrative prominence.

The corpus will avoid treating demographic categories as self-evident textual facts. Author identity metadata will be based on reliable public information and used only when linked to a research question. Character categories will reflect textual representation rather than speculative classification.

5.6 Annotation framework

Manual and semi-automatic annotation will operate at four levels. The first is workplace event: instruction, briefing, task allocation, training, troubleshooting, maintenance, reporting, evaluation, discipline, hiring, dismissal, protest, care, service encounter, safety intervention, accident, and automated decision. The second is communicative function: command, request, warning, prohibition, permission, explanation, diagnosis, justification, disagreement, refusal, reassurance, concealment, escalation, and repair. The third is narrative presentation: direct speech, indirect speech, free indirect discourse, thought presentation, narratorial report, represented document, or digital message. The fourth is ideological-semantic category: efficiency, risk, expertise, replaceability, surveillance, responsibility, precarity, embodiment, resistance, and care.

Operational definitions will be supplied in a coding manual. A workplace event must involve occupational action or institutional relations, not merely occur in a workplace setting. A warning must identify or imply harm and attempt to alter action; general anxiety is not sufficient. An automated decision is coded only when a system's output changes access, evaluation, scheduling, employment, safety, or task sequence within the represented world.

Agency annotation will record grammatical subject, semantic actor, affected participant, institutional role, and whether agency is explicit, backgrounded, personified, or distributed. Modality annotation will distinguish obligation, permission, possibility, prediction, and epistemic certainty. Metaphor annotation will focus on recurrent conceptual domains relevant to work and technology, such as machine, organism, war, network, flow, game, and family.

5.7 Corpus analysis

Analysis will use a transparent combination of frequency, keyness, collocation, lexical bundles, n-grams, semantic tagging, and concordance review. AntConc, LancsBox, Sketch Engine, or comparable tools may be used, but only software actually employed will be reported. R or Python scripts will calculate normalised frequencies, dispersion, log-likelihood, log ratio, and confidence intervals where appropriate.

Internal comparison will be primary. Workplace-domain subcorpora, genre categories, publication periods, and narrative zones will be compared against one another. A carefully matched general-fiction reference corpus may be used to identify workplace-specific salience, but it will not be treated as neutral or universal. Reference-corpus composition must be documented, and interpretations will prioritise effect size and dispersion over p-values alone.

Keyword analysis will identify distinctive nouns, verbs, adjectives, and grammatical forms. Collocation will examine terms such as work, system, machine, human, safe, efficient, error, data, manager, body, and voice. Lexical-bundle analysis will investigate recurrent procedural and managerial phraseology. Modal and imperative searches will examine obligation and authority. Passive and impersonal constructions will be analysed for responsibility attribution. Speech-presentation patterns will be compared across hierarchical roles.

Every aggregate finding will be followed by concordance analysis and close reading of complete scenes. Representative, deviant, and narratively central cases will be selected through explicit criteria. The analysis will not infer irony from isolated lexical contrast; it will examine narratorial stance, focalisation, plot outcome, and intertextual context. Statistical salience will guide attention, not settle interpretation.

5.8 Genre and narrative comparison

Genre analysis will examine how realist, satirical, dystopian, speculative, and dramatic texts organise workplace discourse. Realist narratives may distribute technical detail across routine; satire may intensify managerial cliché; dystopia may literalise institutional metaphors; drama may foreground turn-taking and conflict; speculative fiction may personify systems or redistribute agency to nonhuman actors.

Narrative comparison will examine first- and third-person narration, internal and external focalisation, and temporal organisation. Retrospective narration may reinterpret earlier technical confidence as error. Present-tense narration may intensify procedural urgency. Multiple focalisation may expose conflict between managerial metrics and embodied experience. The corpus will therefore connect linguistic patterns to narrative architecture rather than treating texts as bags of words.

5.9 Reliability and interpretive transparency

Two trained coders will independently annotate 15 to 20 per cent of the corpus material selected across genres, workplace domains, and narrative zones. Cohen’s kappa will be used for two-coder nominal categories; Krippendorff’s alpha will be considered for multi-label or missing-data conditions. Agreement will be reported by category. Weak categories will be revised, merged, or removed rather than retained through vague claims of consensus.

Literary interpretation cannot be reduced to coding agreement, but interpretive claims can still be made transparent. The project will maintain analytic memos linking corpus patterns to selected passages, alternative readings, and rejected explanations. Negative cases will be reported. Where a pattern supports more than one interpretation, the article will identify the competing possibilities and explain why one is preferred.

5.10 Ethics and scholarly responsibility

The project involves texts rather than human participants, but ethical issues remain. Copyright restrictions, fair quotation, edition accuracy, author representation, and canon bias require explicit management. The corpus should not reproduce long passages from copyrighted works or distribute full texts. Author identity and character demographics should not be inferred from names or stereotypes. The study must also avoid treating literary depictions of occupational groups as factual evidence about those groups.

Computational analysis can create an appearance of objectivity that masks selection decisions. Corpus boundaries, excluded texts, failed searches, normalisation rules, and interpretive uncertainty will therefore be documented. The study will not use proprietary generative systems to fabricate quotations, fill missing text, or produce unverifiable literary evidence.

Table 1. Planned composition of the Contemporary English Workplace Literature Corpus

Subcorpus	Indicative material	Primary analytical purpose
Industrial and extractive work	Manufacturing, energy, construction, mining, maintenance	Material labour, hazard, machinery, responsibility
Digital and corporate work	Software, finance, data, administration, surveillance	Metrics, abstraction, managerial language, algorithmic authority
Service, care, and domestic labour	Hospitality, retail, healthcare, education, domestic work	Emotional labour, care, precarity, interpersonal hierarchy
Logistics and circulation	Warehouses, delivery, transport, shipping, supply chains	Timing, tracking, mobility, invisible coordination
Scientific and technical work	Laboratories, engineering, medicine, research	Expertise, evidence, diagnosis, procedural narration
Speculative automation	AI, robotics, cybernetic systems, automated institutions	Machine agency, replacement, posthuman labour, future imaginaries

Table 2. Core metadata for each literary work

Metadata group	Required fields
Bibliographic	Author, year, edition, publisher, original language, copyright status
Literary	Genre, narrative person, focalisation, tense, dialogue proportion
Workplace	Primary and secondary domains, occupational roles, hierarchy, employment security
Technology	Technological intensity, system types, human-machine relations
Corpus	Word count, file source, preprocessing status, excluded material

Table 3. Annotation categories and coding priorities

Category	Operational focus	Coding caution
Instruction	Directive tied to occupational action	Distinguish command from quoted procedure or internal rehearsal
Warning	Risk statement intended to alter action	Separate fear or foreshadowing from in-world warning
Diagnosis	Inference from evidence to cause	Record certainty and whether the narrative validates the inference
Automated decision	System output changes access, work, evaluation, or safety	Do not infer autonomy merely from personifying grammar
Responsibility attribution	Assigns, claims, shares, or suppresses agency	Analyse grammar, hierarchy, and plot outcome together
Resistance	Refusal, challenge, sabotage, reinterpretation, solidarity	Avoid treating every disagreement as political resistance

Table 4. Research questions, evidence, and analysis

Research focus	Primary evidence	Planned analysis
Workplace worlds and events	Metadata and event annotations	Normalised distribution and dispersion
Language of labour and technology	Narrative and dialogue subcorpora	Keywords, collocation, bundles, semantic fields
Authority and responsibility	Commands, modality, passives, reporting verbs	Grammar, role comparison, close scene analysis
Voice and focalisation	Speech/thought presentation and narrative zones	Distribution by role and narratological interpretation
Genre variation	Realist, satirical, dystopian, dramatic, speculative groups	Key-feature comparison and representative readings
Curricular application	Corpus outputs and teaching tasks	Pedagogical design without vocational reduction

6. Analytical Programme and Reporting Framework

Because the corpus has not yet been assembled, this section specifies the organisation of a future empirical report rather than presenting results. The first results section will profile the achieved corpus: number of works,

total words, genre distribution, publication periods, workplace domains, dialogue proportions, and missing or excluded materials. Planned and achieved composition will be distinguished.

The second section will map the literary workplace. It will report the distribution of settings, roles, communicative events, and technologies. Frequency will be normalised, but interpretation will consider narrative salience. A rare industrial accident may structure an entire novel, while frequent office routines may remain background texture.

The third section will analyse the language of authority and procedure. Imperatives, modals, managerial bundles, conditionals, and passive constructions will be compared across roles and narrative zones. Close readings will examine how formal authority differs from epistemic authority. A manager may control turn-taking while a technician controls causal explanation; a system may appear grammatically agentive while the plot reveals human design choices.

The fourth section will examine risk, error, and responsibility. Warning scenes, near misses, accidents, diagnostic disagreements, and reports will be traced as narrative sequences. The analysis will ask who notices danger, who is believed, how uncertainty is expressed, and how blame is assigned after failure. It will distinguish prospective risk language from retrospective explanation.

The fifth section will address automation and machine agency. Collocations and transitivity patterns will show how systems are animated, naturalised, feared, or trivialised. Close readings will test whether personification conceals institutional responsibility, creates uncanny affect, or grants nonhuman actors genuine narrative significance.

The sixth section will analyse labouring bodies and invisible work. Body-part clusters, sensory language, fatigue terms, maintenance verbs, care vocabulary, and temporal expressions will be compared across domains. This section will ask which forms of work are linguistically visible and which are backgrounded behind smooth technological surfaces.

The seventh section will examine voice and silence. Speech-presentation categories will identify which occupational groups receive direct quotation, interiority, narratorial sympathy, or merely functional mention. The analysis will consider interruption, reported speech, free indirect discourse, and represented documents as formal mechanisms for distributing authority.

The final synthesis will integrate these findings into a literary-industrial discourse framework. It will not rank texts according to ideological correctness. Instead, it will identify recurring formal strategies through which literature makes technological work intelligible, contested, or strange.

7. Literary-Industrial Discourse Framework

The final framework will contain eight dimensions derived from corpus evidence. Occupational world-building concerns the lexical and descriptive resources through which texts establish workplaces, tools, routines, and institutional environments. Procedural narration concerns the representation of sequence, instruction, maintenance, and task completion. Hierarchical voice concerns turn allocation, command, politeness, interruption, and access to direct speech. Epistemic authority concerns who can diagnose, predict, explain, and interpret technical evidence.

Risk and responsibility concern warnings, uncertainty, causal attribution, blame, and the naming or suppression of agents. Technological agency concerns personification, grammatical agency, automation, and the relation between system outputs and human decisions. Embodied labour concerns fatigue, injury, sensory experience, care, repetition, and the physical costs hidden by abstract efficiency language. Resistance and alternative futures concern refusal, repair, solidarity, sabotage, reinterpretation, and the imagination of work beyond managerial inevitability.

Each dimension will connect linguistic resources with literary functions. Imperatives may organise procedural narration but also characterise domination. Passive constructions may represent technical process or obscure culpability. Repeated efficiency collocations may create realism, satire, or ideological enclosure depending on context. Body-language clusters may individualise workers or reduce them to functional parts. Direct speech may empower a character, yet it can also expose coercion or self-deception.

The framework is not a checklist for judging whether a novel is “accurate.” Its purpose is to make comparison possible while preserving literary specificity. A text can be technically implausible but formally insightful; another can be procedurally detailed yet ideologically narrow. The analysis should show how each

work constructs its own relation among language, labour, and technology.

Table 5. Proposed literary-industrial discourse dimensions

Dimension	Observable resources	Interpretive question
Occupational world-building	Terms, objects, routines, spatial description	How is work made materially and institutionally legible?
Procedural narration	Sequence, condition, instruction, completion	How does the text narrate competent or failed action?
Hierarchical voice	Commands, address, interruption, quotation	Who is entitled to speak and be heard?
Epistemic authority	Diagnosis, evidence, certainty, prediction	Whose knowledge is validated by the narrative?
Risk and responsibility	Warnings, blame, passives, causal accounts	How are danger and accountability distributed?
Technological agency	Personification, active verbs, system outputs	Does technology conceal, redistribute, or acquire agency?
Embodied labour	Fatigue, pain, gesture, sensory and care language	What costs are hidden by abstract efficiency?
Resistance and alternatives	Refusal, repair, solidarity, counterfutures	How does the text imagine action beyond inevitability?

8. Implications for English Literature Curricula in Saudi Universities

The revised project has direct relevance to English literature programmes in Saudi Arabia, but its pedagogical value should not be confused with vocational training. Literature departments contribute to national development by cultivating interpretive judgment, historical awareness, linguistic sensitivity, ethical reasoning, and the capacity to analyse narratives that legitimise or contest social change. These capacities are particularly important when technological transformation is presented as inevitable and universally beneficial.

A corpus-assisted module on work and technology could be integrated into courses in contemporary fiction, stylistics, digital humanities, or literary theory. Students might compare keywords across realist and dystopian subcorpora, trace the speech patterns of managers and workers, examine how a novel personifies an algorithm, or analyse the grammar of responsibility after an accident. Such tasks make evidence explicit and encourage students to test interpretive claims rather than rely on thematic paraphrase.

Close reading must remain central. Concordance lines can identify patterns, but students should return to chapters, scenes, and whole-text structures. A productive sequence would move from hypothesis to corpus search, from aggregate pattern to passage analysis, and from passage analysis to historical and ideological interpretation. Assessment could include a reproducible mini-corpus project, a close-reading essay supported by concordance evidence, or a comparative analysis of narrative voice and technological agency.

The corpus can also support interdisciplinary collaboration without subordinating literature to engineering or business. Literature students might work with technical specialists to understand procedural language, while retaining authority over narrative and stylistic interpretation. Such collaboration can reveal where fictional representations simplify technology and where technical discourse itself relies on metaphor, narrative, and institutional assumptions.

Arabic-language discussion may be pedagogically useful in Saudi classrooms, but the manuscript itself remains in English and contains no Arabic script. Where Arabic concepts or titles are later introduced, they should be consistently Romanised and translated. The choice of classroom language should support literary understanding rather than be treated as a measure of cultural or intellectual worth.

The strongest curricular contribution is critical literacy. Students learn to recognise how words such as innovation, efficiency, flexibility, optimisation, and smart can function as evaluative terms rather than neutral descriptions. They also learn that literary form can disclose what managerial summaries omit: hesitation, fatigue, memory, shame, care, fear, and the unequal capacity to refuse.

9. Critical Discussion

Reframing the project around English literature changes the meaning of evidence. Authentic workplace corpora can support claims about recurrent professional practices; a literary corpus supports claims about representation. The distinction is fundamental. Fiction does not show what workers generally say. It shows what authors, narrators, genres, and publishing cultures select as meaningful or imaginable. A technically realistic dialogue remains mediated by plot and style, while an implausible scene may reveal a culture's anxieties about automation more clearly than a documentary transcript.

This distinction also strengthens rather than weakens the project. Literature can bring together scales that institutional documents separate. A single scene may connect a command to a bodily reaction, a technical error to a family's insecurity, or an algorithmic decision to a character's sense of self. Corpus analysis can reveal how often and where such connections recur, while close reading explains their aesthetic force.

The framework complicates celebratory Industry 4.0 discourse. Terms such as smart, autonomous, seamless, and efficient often suppress maintenance, energy use, human oversight, and institutional choice. Contemporary fiction can reproduce these terms through corporate characters, expose their limits through plot, or embed them in ironic narration. The analysis should not presume critique in advance, however. Some texts may endorse technological mastery; others may express nostalgia for labour relations that were themselves exploitative. Critical interpretation must remain open to ideological contradiction.

The project also challenges the assumption that English literature becomes socially relevant only when aligned with employability. Its relevance lies partly in resisting narrow utility. Literary texts cultivate attention to ambiguity, indirect consequence, competing perspectives, and the moral language of responsibility. These are not decorative capacities, but they should not be justified only as workplace skills. A humanities curriculum contributes by examining the stories through which societies understand work, value, progress, and human replaceability.

Corpus methods introduce their own risks. Quantification can privilege what is easy to count and understate silence, narrative structure, intertextuality, and historical nuance. Search terms can reproduce the researcher's expectations. Categories such as worker, manager, machine, and resistance are not always stable across genres. The framework addresses these risks through iterative coding, negative-case analysis, narrative-zone segmentation, and explicit separation of statistical salience from literary significance.

A further tension concerns canon formation. Contemporary workplace fiction is distributed unevenly across literary prestige, genre publishing, digital platforms, and regional markets. A corpus based only on prize lists may privilege professional-class anxieties and underrepresent manual, migrant, domestic, and service labour. A corpus based only on keyword-rich digital availability may overrepresent genre fiction or texts with accessible electronic editions. The selection protocol must therefore balance critical relevance, diversity, and legal access rather than pretend that any corpus is neutral.

Finally, machine agency requires careful interpretation. Fiction often grants technologies active verbs because human language readily personifies complex systems. Such grammar may intensify fear, reduce explanatory burden, or dramatise genuine constraints. Yet it can also absolve designers, managers, and institutions. The most revealing question is not whether the machine "really acts," but how the text distributes agency among code, infrastructure, policy, human intention, and narrative perspective.

10. Limitations

The proposed corpus will be bounded and cannot represent all contemporary English literature. Copyright restrictions may favour older, digitised, or institutionally accessible texts. Genre balance will remain imperfect, and sustained workplace narratives may be more common in some domains than others. The corpus may underrepresent unpublished, digital-native, regional, or small-press writing.

Text preparation can introduce errors. Electronic editions differ in typography and paratext; OCR may distort dialogue and technical terms; automated tagging may misclassify fragments, invented vocabulary, and non-standard speech. Manual checks reduce but do not eliminate these problems.

Annotation categories inevitably simplify literary ambiguity. A line can function simultaneously as instruction, threat, irony, and characterisation. Primary-function coding supports comparison but may suppress multiplicity. The project should therefore retain multi-label annotations where justified and use qualitative

memos to preserve interpretive complexity.

The study does not include reader-response data. Claims about likely reader effects will be cautiously grounded in textual cues and established stylistic theory, not presented as measured psychological outcomes. Later research could incorporate classroom studies, reading experiments, or interviews.

The Saudi curricular implications are conceptual rather than institutionally evaluated. The framework can guide course design, but its effectiveness would need to be tested with students and instructors under appropriate ethical approval. It should not be presented as a nationally validated curriculum model.

Finally, the period 1990–2025 contains rapid changes in technology and labour discourse. A corpus assembled at one moment may date quickly, particularly in relation to artificial intelligence, platform work, and digital surveillance. Collection dates, editions, and technological references must therefore be recorded, and the corpus should be designed for transparent extension.

11. Conclusion

A shift from technical-education discourse to English literature requires more than a revised title. It requires a new object of study: not authentic workshop communication, but the literary representation of work, technology, expertise, risk, and institutional power. This article has reconstructed the project accordingly.

The proposed Contemporary English Workplace Literature Corpus offers a reproducible architecture for combining corpus stylistics with narratology, speech and thought presentation, critical stylistics, and labour-oriented literary criticism. It defines selection criteria, subcorpora, narrative zones, metadata, annotation categories, corpus procedures, reliability measures, copyright safeguards, and reporting principles. It also establishes a strict boundary between representation and workplace reality, preventing fictional dialogue from being treated as documentary evidence.

Once implemented, the framework can show how contemporary English literature makes technological work audible and visible: who gives orders, who interprets systems, whose warnings are ignored, how machines acquire grammatical agency, how responsibility is obscured, and how bodies register the costs of efficiency. It can also reveal differences among realist, satirical, dystopian, dramatic, and speculative forms.

For English literature curricula in Saudi universities, the project offers a method for connecting close reading to contemporary debates about automation and labour without reducing literature to vocational instruction. Corpus evidence can sharpen interpretation, while literary form preserves ambiguity, ethical conflict, and historical depth. The resulting study would position English literature not at the margin of technological transformation, but as a discipline capable of analysing the narratives through which transformation is justified, feared, resisted, and imagined otherwise.

Funding

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU2026/02/44998)

References

- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (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 & Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508088962>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Cermakova, A., & Mahlberg, M. (2021). Gender inequality and female body language in children's literature. *Digital Scholarship in the Humanities*, 36(Supplement_2), ii72–ii77. <https://doi.org/10.1093/llc/fqaa051>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165–193. <https://doi.org/10.1093/applin/amv011>
- Egbert, J., & Biber, D. (2023). Key feature analysis: A simple, yet powerful method for comparing text varieties. *Corpora*, 18(1), 121–133. <https://doi.org/10.3366/cor.2023.0275>
- Gillings, M., Learmonth, M., & Mautner, G. (2024). Taking the road less travelled: Corpus-assisted discourse studies in organisational research. *British Journal of Management*, 35(4), 1667–1679. <https://doi.org/10.1111/1467-8551.12816>
- Green, C. (2017). Introducing the Corpus of the Canon of Western Literature: A corpus for culturomics and stylistics. *Language and Literature*, 26(4), 282–299. <https://doi.org/10.1177/0963947017718996>
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Hatzel, H. O., Biemann, C., & Reiter, N. (2023). Machine learning in computational literary studies. *It - Information Technology*, 65(4–5), 200–217. <https://doi.org/10.1515/itit-2023-0041>
- Hughes, J. M., Foti, N. J., Krakauer, D. C., & Rockmore, D. N. (2012). Quantitative patterns of stylistic influence in the evolution of literature. *Proceedings of the National Academy of Sciences*, 109(20), 7682–7686. <https://doi.org/10.1073/pnas.1115407109>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- Kilgariff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Lazzeretti, C. (2019). 'I've been nowhere and done nothing': The characterization of Daisy Mason in the British drama *Downton Abbey*. *Series - International Journal of TV Serial Narratives*, 5(1), 33–44. <https://doi.org/10.6092/issn.2421-454X/8966>
- Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora*, 2(1), 1–31. <https://doi.org/10.3366/cor.2007.2.1.1>
- Mahlberg, M. (2012). Dickens, the suspended quotation and the corpus. *Language and Literature*, 21(1), 51–65. <https://doi.org/10.1177/0963947011432058>
- Mahlberg, M. (2015). Literary style and literary texts. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 346–361). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.020>
- Mahlberg, M., Conklin, K., & Bisson, M.-J. (2014). Reading Dickens's characters: Employing psycholinguistic methods to investigate the cognitive reality of patterns in texts. *Language and Literature*, 23(4), 369–388. <https://doi.org/10.1177/0963947014543887>
- Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*, 4(2), 204–227. <https://doi.org/10.1075/etc.4.2.03mah>
- McIntyre, D. (2015). Towards an integrated corpus stylistics. *Topics in Linguistics*, 16(1), 59–68. <https://doi.org/10.2478/topling-2015-0011>
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig,

- P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>
- O'Halloran, K. (2007). The subconscious in James Joyce's 'Eveline': A corpus stylistic analysis that chews on the 'Fish hook'. *Language and Literature*, 16(3), 227–244. <https://doi.org/10.1177/0963947007072847>
- Partington, A., Duguid, A., & Taylor, C. (2013). Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies. John Benjamins. <https://doi.org/10.1075/scl.55>
- Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE*, 10(10), e0137041. <https://doi.org/10.1371/journal.pone.0137041>
- Scott, M. (1997). PC analysis of key words—and key key words. *System*, 25(2), 233–245. [https://doi.org/10.1016/S0346-251X\(97\)00011-0](https://doi.org/10.1016/S0346-251X(97)00011-0)
- Semino, E., & Short, M. (2004). Corpus stylistics: Speech, writing and thought presentation in a corpus of English narratives. Routledge. <https://doi.org/10.4324/9780203494073>
- Stockwell, P., & Mahlberg, M. (2015). Mind-modelling with corpus stylistics in David Copperfield. *Language and Literature*, 24(2), 129–147. <https://doi.org/10.1177/0963947015576168>
- Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5–24. <https://doi.org/10.1177/0963947005048873>
- Wijitsopon, R. (2022). Corpus stylistics and colour symbolism in *The Great Gatsby* and its Thai translations. *Language and Literature*, 31(1), 54–77. <https://doi.org/10.1177/09639470211048159>

Appendix A. CEWLC Inclusion Record

- Work identifier and bibliographic source
- Copyright and access status
- Publication year and edition
- Original language and translation status
- Genre and narrative form
- Primary and secondary workplace domains
- Technological intensity and system types
- Sustained workplace relevance: yes/no with justification
- Word count and dialogue proportion
- Preprocessing and quality-control notes
- Inclusion, exclusion, or borderline decision
- Reviewer initials and decision date