Check for updates

Journal of Management
Vol. XX No. X, Month XXXX 1–28
DOI: 10.1177/01492063261465289
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions

Making Words Count: Computational Linguistics in Management Research

Joseph J. Simpson
The University of Texas Rio Grande Valley
Richard A. Hunt
Virginia Tech
David M. Townsend
Virginia Tech
Judy Rady
University of Kansas
Elham Asgari
Michigan Tech
Mohammad Rady
Virginia Tech
Daniel J. Beal
Virginia Tech

In recent years, a rapidly growing number of management scholars are using computational linguistics (CL) to analyze vast corpora of naturally occurring organizational language. Yet the rapid adoption of CL methods has outpaced the development of shared standards for linking theoretical constructs, linguistic data, and computational measurement. We synthesize 353 articles published in leading management journals for 2013–2025, identifying four recurring challenges shaping the credibility of CL-based research: misalignment between constructs and textual corpora, unnecessary model complexity, the opacity of large language model workflows, and insufficient transparency in reporting. Building on these patterns, we develop four integrative design principles—construct-language fit, minimal sufficient complexity, LLM sensitivity and auditability, and disclosure-as-assessability—and identify key research implications that reposition CL as a theoretically consequential lens for management scholarship.

Current affiliation: Judy Rady is now affiliated to Syracuse University.

Supplemental material for this article is available with the manuscript on the JOM *website.*

Corresponding author: Joseph J. Simpson, The University of Texas Rio Grande Valley, 402 N Sugar Rd, Edinburg, TX 78541.

E-mail: jjsimpson@vt.edu

Keywords: *computational linguistics; natural language processing; large language models; research methods; artificial intelligence; machine learning*

Introduction

Making words count in management research requires more than impressive computational tools; it requires aligning theoretical constructs, linguistic evidence, and analytic methods. The ability to readily access and carefully assess digitized information has transformed the conceptual, empirical, and methodological terrain of management research, reflecting an explosion in the availability of source documents and a proliferation of new analytical tools that both enable and challenge scholars. Worldwide, individuals and organizations generate more than 300 million terabytes of new data every day (Edge Delta, 2024), encompassing content such as corporate filings, patent abstracts, employee texts, quarterly earnings call transcripts, emails, social media threads, internal chat logs, wearable sensor transcripts, and AI-generated text. This digital information provides scholars with unprecedented access to naturally occurring language, reflecting a vast array of managerial and organizational phenomena—for example, employee and customer behavior, strategic decision-making, organizational routines, stakeholder engagement, and employee wellbeing—captured in real time (Amaya & Holweg, 2024; Guzman & Li, 2023; Shalpegin et al., 2025; Woo, Tay, & Oswald, 2024; Zhang, Yu, & Marin, 2021). Far from being a passing methodological fad, computational linguistics (CL)—the analysis and synthesis of language and speech (Schubert, 2020)—is an indispensable toolkit for analyzing how actors communicate, influence, and enact organizational reality through language, as evidenced by a surge in peer-reviewed articles using CL methods across leading management journals (Padmanabhan, Fang, Sahoo, & Burton-Jones, 2022; von Krogh, Roberson, & Gruber, 2023).

Although CL methods are now widely used, the field lacks a shared methodological infrastructure that links theoretical constructs, linguistic data, and computational measurement (Nyberg et al., 2025). As a result, scholars often adopt potent analytical tools without a clear framework for evaluating when and how those tools produce valid theoretical inferences. Scholars now harvest language from an ever-widening array of sources and increasingly rely on large language models (LLMs) to analyze organizational texts (e.g., Antons, Breidbach, Joshi, & Salge, 2023; Kobayashi, Mol, Berkers, Kismihók, & Den Hartog, 2018; Shalpegin et al., 2025). Yet, growing concerns about reliability and validity accompany these tools, particularly given the challenges of controlling and interpreting opaque model outputs (Nelson, 2020). Language is not merely a data source for organizational research; it is a core mechanism through which actors construct identities, signal strategic intent, negotiate power, and coordinate collective action. Given this, CL expands not only the scale with which language can be analyzed but also the range of theoretical questions scholars can address about cognition, meaning, and organizing. Although recent studies have initiated essential conversations regarding guidelines for effective CL use (e.g., Campion & Campion, 2023; Speer, Oswald, & Putka, 2026), more work is needed to identify the challenges and best practices in applying these tools across micro- and macro-level research (e.g., Haans & Mertens, 2024). This makes it crucial to

address the question: How can scholars effectively utilize CL to generate theoretically rigorous and methodologically credible insights?

Our central insight is that the credibility and effectiveness of CL research hinges on the alignment among three elements: the theoretical construct, the linguistic signal through which that construct is expressed, and the computational method used to measure it. When these elements are misaligned—for example, when theory implies contextual meaning, but methods capture only word frequency—sophisticated analytical tools can produce results that superficially appear precise, yet whose deeper theoretical meaning may be specious. Building on this insight, we synthesized 353 CL-based studies published in leading management journals to identify recurring alignment problems, and derived four integrative principles designed to guide the alignment of theory, language, and computational analysis. The first principle, construct–language fit, emphasizes that linguistic corpora must plausibly contain traces of the construct being theorized. The second, minimal sufficient complexity, holds that model sophistication should follow from the linguistic mechanism implied by theory rather than from the novelty of available algorithms. The third, LLM sensitivity and auditability, addresses the inferential fragility introduced by prompt design and model configuration when LLMs are used for annotation or measurement. The fourth, disclosure-as-assessability, holds that documentation is not a validity-enhancing procedure but a precondition for others to evaluate whether validity was achieved. Taken together, these principles function as durable decision rules that clarify how theoretical constructs, linguistic data, and computational techniques can be systematically aligned in management research.

Building on this framework, our review makes three contributions to management scholarship. First, we develop an integrative map of computational linguistics research in management by analyzing studies across major journals, and identifying the dominant methodological trajectories, recurring inferential risks, and emerging patterns in how scholars deploy CL tools for micro- and macro-level research. This synthesis provides a conceptual decision framework that aligns CL methods with the theoretical mechanisms they are intended to capture. Second, we advance methodological standards for CL research by introducing a two-tiered reporting framework that distinguishes between minimum viable disclosure and exemplary disclosure practices, thereby clarifying what constitutes assessable and cumulative CL-based evidence. Third, we outline a forward-looking research agenda that follows from the principles identified in our review: treating computational linguistics as organizational infrastructure, strengthening inference under conditions of distortion and strategic silence, and developing the institutional and methodological conditions required for cumulative CL science. In line with Kunisch, Denyer, Bartunek, Menz, and Cardinal's (2023) typology of review articles, these contributions collectively demarcate the domain of CL in management, classify its methodological approaches, interpret evolving research patterns, and problematize persistent blind spots that limit the field's ability to generate reliable and theoretically meaningful insights.

Review Methodology

Consistent with the procedures used in previously published review articles in the *Journal of Management* (e.g., Bolinger, Josefy, Stevenson, & Hitt, 2022; Hymer & Smith, 2024), the core scope of our review included all topically relevant works from outlets listed in the UTD

and TAMUGA journal lists. We also included articles from *Organizational Research Methods* and premier operations management journals, given their prominence in publishing research on CL. The final list consisted of: *Academy of Management Journal*, *Administrative Science Quarterly*, *Journal of Applied Psychology*, *Journal of International Business Studies*, *Journal of Management*, *Journal of Operations Management*, *Management Science*, *Manufacturing & Service Operations Management*, *Operations Research*, *Organization Science*, *Organizational Behavior and Human Decision Processes*, *Organizational Research Methods*, *Personnel Psychology*, *Production and Operations Management*, and *Strategic Management Journal.*

We then built a keyword list containing terms common to the broader literature:[1] "Computational linguistics", "CATA", "Computer-Aided Text Analysis", "Natural Language Processing", "DICTION", "NLP", "Text Analysis", "Textual Analysis", "Text Analytics", "Textual Analytics", "Sentiment Analysis", "Topic Mode*", "Text Mining", "Bag of Words", "Tokenization", "Stemming", "Lemmatization", "Semantic Analysis", "Latent Dirichlet Allocation", "LDA", "Named Entity Recognition", "NER", "Part-of-Speech Tagging", "POS Tagging", "Dependency Parsing", "Syntactic Parsing", "Coreference Resolution", "Text Summarization", "Text Classification", "Word Embeddings", "Word2Vec", "GloVe", "Document Embeddings", "Contextual Embeddings", "BERT", "Distributional Semantics", "Text Clustering", "Lexical Analysis", "Discourse Analysis", "Word Sense Disambiguation", "Semantic Similarity", "Dialogue System", "Speech Recognition", "Transformer", "Generative Model", "LLM", "GPT", "N-Gram", "LIWC", "NVivo", "Bibliometric".

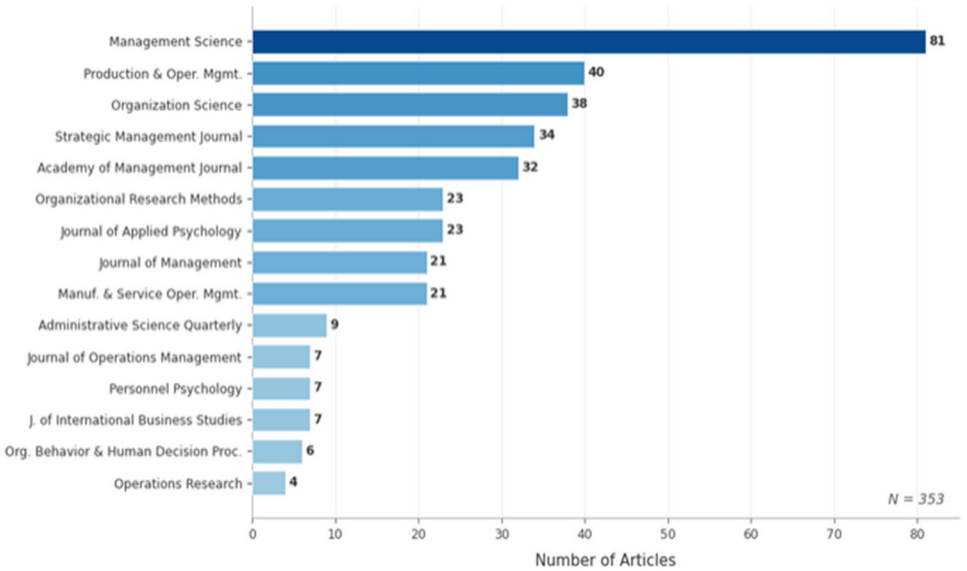
Consistent with best practices (e.g., Paruchuri, Hoempler, Cowen, Cannella, & Nahm, 2024), we utilized the Web of Science (WoS) to collect all articles relevant to our journal and keyword list. Because our review focuses more on state-of-the-art CL methods, we restricted our search to articles published in the years 2013–2025, inclusive, identifying a set of 353 articles related to our search terms and focal journals.[2]

Literature Scope and Article Selection Process

From the WoS, we downloaded all article information (e.g., Article Title, Publication Title, Keywords, and Abstract) as RIS files, storing the "full record" for each result. We filtered the initial set of articles by first removing book reviews, commentaries, duplicates, and calls for papers. However, we included other review articles and methodological articles for their relevance to this *JOM* Review. Next, we coded any article that obviously used CL analysis as part of its empirical strategy (1 = Yes; 0 = No). We then evaluated the interrater reliability of our article coding. Across three judges, our Cohen's Kappa was .74, .87, and .74 on the first attempt. Fleiss' Kappa was .79. These scores indicate sufficient agreement across judges (Fleiss, 1981). As a result of this coding process, we reduced the total number of articles to 225. Since WoS often contains partial records for some journals for some publication years, we conducted a manual search for relevant articles, adding relevant articles and updating the review corpus all relevant articles published through December 31, 2025. Through this process, we identified 128 additional articles, resulting in a total of 353 articles.

Next, we categorized and analyzed the distribution of articles in our set to determine the number of CL articles by journal. The top three journals were: *Management Science*,

Figure 1
Number of Computational Linguistics Articles by Journal



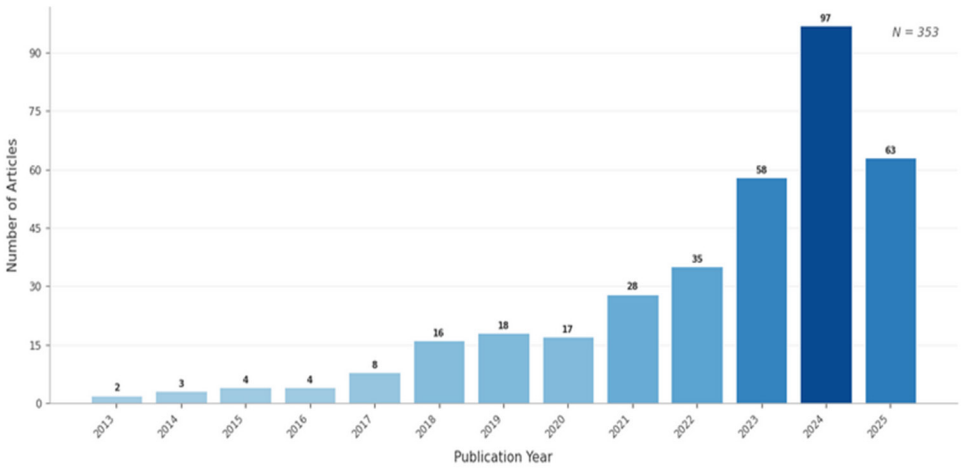
Production and Operations Management, and *Organization Science*. The journals with the fewest articles were *Operations Research*, *Organizational Behavior and Human Decision Processes*, and *Journal of International Business Studies*. Figure 1 summarizes these journal frequency distributions. Figure 2 displays the number of CL articles by year in top management journals for 2013–2025. CL publications experienced a sharp rise starting in 2018, and specifically peaking in 2024, surpassing all other years in terms of the number of articles, with a 60% increase from 2023 to 2024 alone. In total, these articles analyzed more than 2.3 billion text documents.

Understanding Computational Linguistics and its Challenges

Defining CL in Management Research

As noted from the outset, CL is the study and application of computational methods to interpret, understand, and use human language (Hirschberg & Manning 2015; Schubert, 2020). This scope encompasses all phases of analysis, including identifying textual artifacts from text collection, conversion, and preprocessing to evaluation and reporting of results. These techniques and methods have existed for over 50 years, as evidenced by the founding of the *American Journal of Computational Linguistics* (*AJCL*), now *Computational Linguistics*, which highlights the field's enduring nature. Early articles in *AJCL* covered topics ranging from pattern matching in natural language processing (Colby, Parkison, & Faught,

Figure 2
Number of computational linguistics articles by year



Note. The decline from 2024 to 2025 is due to indexing timing for articles collected in early 2025. Articles from 2024 that are first available online often change to a 2025 publication date once they are officially assigned to an issue. This phenomenon has the effect of slightly inflating the 2024 count and deflating the 2025 count for a period of time.

1974) to concept extraction and text generation (Kittredge, 1983); subjects that management scholars with a background in CL today will recognize. The Association for Computational Linguistics traces its origins even earlier, to the founding of the Association for Machine Translation and Computational Linguistics in 1962, indicating a community of interest that predated AJCL. Although the most familiar question–answering systems today (e.g., ChatGPT) are relatively recent phenomena utilized by hundreds of millions of people around the globe, the general concept of using computers to read and answer questions about texts has existed for nearly 50 years (e.g., Evens & Smith, 1978). It was not until much later that scholars began to leverage advances from CL to the discipline of management.

We organize CL methods according to the type of linguistic signal each method can analyze—a distinction that determines which theoretical constructs a given tool can credibly measure. Doing so sets the stage for understanding and then resolving the field's previously mentioned "alignment" problems. The CL landscape comprises three primary "signaling" techniques used to identify, analyze, and process the structural, semantic, and contextual cues (or signals) in human language, making it interpretable by machines. *Lexical-signal methods*, including bag-of-words (BoW) dictionaries, sentiment lexicons, and rule-based extraction, operate on word presence and frequency. These tools provide a direct, transparent mapping between linguistic markers and theoretical claims, making them well suited for cases where the focal construct is defined by vocabulary-level phenomena such as tone, hedging, or agentic language. For instance, a basic dictionary-based sentiment tool assigns a fixed value to individual words and aggregates these values, so a simple implementation would score "not happy" as positive because it scores "not" as neutral and "happy" as positive. Lexical tools,

particularly modern ones, can be extended to catch local interactions such as negation—but handling is itself lexical, depending upon which terms count as negators, the scope window, and whether the negator precedes the affected word or words. *Distributional-signal methods* such as topic models (LDA, STM) and static word embeddings (Word2Vec, GloVe) capture statistical regularities in how words co-occur within linguistic contexts, revealing latent themes and semantic proximity without modeling full sentential composition. These methods go beyond individual word counts, producing context-independent representations of words rather than modeling meaning at the sentence level. *Compositional-signal methods*, inclusive of transformer-based models (BERT, RoBERTa), fine-tuned classifiers, and large language models, compute a separate representation for each occurrence of a word from its surrounding context, so the same word form can be represented differently from one passage to the next. A compositional method would correctly classify "not happy" as negative because it models how words interact within the sentence, not because an analyst supplied a negation rule. When research designs require combining domain-specific rules with learned semantic representations, hybrid pipelines that span multiple signal types may be warranted.

The distinctions among these signal types are not based on chronological sophistication (i.e., their sequential order) but rather on *what kind of meaning* each method can access. Lexical and distributional methods both assign single representations to each word type, differing only in whether that representation is fixed by hand (lexical) or learned from corpus-wide co-occurrence (distributional). Compositional methods differ in that they recompute a word's representation for each occurrence in the text, so identical word forms can receive different representations depending on context. This matters because the appropriate method depends on the linguistic mechanism implied by the focal theory, not on the novelty of the algorithm. Table 1 organizes computational linguistics methods according to the linguistic mechanism implied by theory, clarifying how construct definitions determine the appropriate level of model complexity.

As Table 1 illustrates, each signal type carries distinct advantages. Lexical methods offer transparency, interpretability, and direct reproducibility; properties that make them preferable, not merely acceptable, when the focal construct is vocabulary-level. Distributional methods reveal thematic structure and conceptual relationships that no word-counting approach can detect. Compositional methods enable disambiguation of meaning in context, but at the cost of reduced transparency. All three method families warrant comparable validation rigor; what reduced transparency changes is not how much validation a compositional measure deserves, but how difficult its validity is to demonstrate and scrutinize. In sum, these method families constitute complementary options, and the choice among them should be driven by the linguistic mechanism the theory implies rather than by methodological fashion or algorithmic novelty. We develop this matching logic more fully below as the principle of minimal sufficient complexity (Principle 2).

Patterns and Problems in Computational Linguistics Research

Building on the previous section where we discussed the core capabilities of CL, we now analyze existing applications of CL in management research. Our analysis reveals four recurring problems that shape the credibility of CL-based research: (i) misalignment between constructs and textual corpora; (ii) unnecessary model complexity; (iii) the opacity of large

Table 1
Choosing Computational Linguistics Methods Based on the Linguistic Mechanism Implied by Theory

Linguistic Mechanism Implied by Theory	Method Family	Representative Tools	Best Suited When the Focal Construct. . .	Illustrative Article(s)
Lexical signals: Meaning conveyed through word presence or frequency	Bag-of-words dictionaries	LIWC, DICTION, custom word lists, sentiment lexicons	. . .is defined by which words appear and how often (e.g., tone, hedging, agentic vs. communal language)	Ridge, Hill, Ingram, Kolomeitsev, and Worrell (2024); Madera, Hebl, and Martin (2009)
	Rule-based extraction	Regex, pattern matching, gazetteer-based NER	. . .is a surface-level entity or linguistic pattern (e.g., firm names, dates, section headers in filings)	Jung, Zhou, and Smith (2025)
	Supervised classifiers (BoW/TF-IDF features)	SVM, Naive Bayes, Random Forest	. . .has labeled training data and vocabulary-level features are sufficient for prediction	Harrison, Josefy, Kalm, and Krause (2023); Hickman, Liff, Rottman, and Calderwood (2026)
Distributional signals: Meaning reflected in word co-occurrence patterns	Topic models	LDA, STM, LSA, correlated topic models	. . .is a latent theme or category defined by recurring co-occurrence patterns among words	Schmiedel, Müller, and Vom Brocke (2019)
	Static word embeddings	Word2Vec, GloVe, Doc2Vec, FastText	. . .involves conceptual similarity, semantic distance, or domain coverage	Cong, Liang, Zhang, and Zhu (2025)
Compositional signals: Meaning depends on how words combine in context	Contextual embeddings / transformers	BERT, RoBERTa, DeBERTa	. . .depends on negation, qualification, or polysemy that changes meaning at the sentence level	Carlson (2023); Hernandez and Nie (2023)
	Fine-tuned classifiers	BERT-based classifiers, prompt-tuned models, few-shot learning	. . .requires sentence-level semantic judgment trained on domain-specific examples	Siano (2025)
	LLM-based annotation and measurement	GPT-4, Claude, Llama	. . .requires holistic interpretation similar to human coding (e.g., impression management strategies)	Carlson and Burbano (2026); Speer et al. (2026)
Hybrid designs: Constructs spanning multiple linguistic mechanisms	Hybrid and ensemble pipelines	Rule-based + ML stacking, multi-model workflows	. . .requires domain-specific rules combined with learned semantic representations across signal types	Choudhury, Wang, Carlson, and Khanna (2019)

Note. Methods are organized according to the linguistic mechanism implied by theory rather than technological sophistication. Lexical methods measure word presence and frequency. Distributional models learn a single representation for each word type from co-occurrence patterns, revealing latent themes and semantic proximity. Compositional methods compute a context-specific representation for each word occurrence, enabling negation detection, disambiguation, and passage-level inference. The use of context during training (as in distributional embeddings) or in local rules (as in lexical negation handling) does not by itself make a method compositional; what distinguishes compositional methods is that they produce a different representation for the same word form depending on its context. This classification operationalizes the principle of minimal sufficient complexity (Principle 2): lexical methods are most transparent when the focal construct can be captured through relatively stable linguistic markers, whereas compositional methods become necessary when valid measurement depends materially on interpreting how words interact within sentences or passages. Both over-engineering (applying complex models to lexical constructs) and under-engineering (applying lexical methods to compositional constructs) risk misaligning theoretical constructs with computational measurement. LLM-based methods also introduce auditability concerns addressed by Principle 3.
LIWC = Linguistic Inquiry and Word Count; NER = named entity recognition; BoW = bag of words; TF-IDF = term frequency-inverse document frequency; SVM = support vector machine; LDA = latent Dirichlet allocation; STM = structural topic model; LSA = latent semantic analysis; GloVe = Global Vectors for Word Representation; BERT = Bidirectional Encoder Representations from Transformers; RoBERTa = Robustly Optimized BERT Pretraining Approach; DeBERTa = Decoding-enhanced BERT with Disentangled Attention; LLM = large language model; GPT = Generative Pre-trained Transformer; ML = machine learning.

Table 2
Data Sources and Usage

Data-Source Family	% of Articles ($N = 353$)	Corpora Examples	Illustrative Studies
Social Media	30%	Twitter/X posts; Yelp and TripAdvisor reviews; Reddit threads; Glassdoor reviews; Kickstarter pages	Schmidt, Wuttke, Ball, and Heese (2020) *JOM*; Zhang and Luo (2023) *MS*
Corporate Filings and Events	21%	SEC 10-K/10-Q MD&A sections; DEF 14A proxy statements; StreetEvents earnings-call transcripts	Bae, Belo, Li, Lin, and Zhao (2023) *MS*; Parker et al. (2024) *SMJ*
Press and News	13%	Factiva/LexisNexis newspaper articles; CNBC interview transcripts; activist letters to SEC; analyst reports	Dewan, Simons, and Wernicke (2024) *OS*; Yoganarasimhan and Iakovetskaia (2024) *MS*
Employee Text	10%	Performance-appraisal narratives; applicant essays and video-interview transcripts; internal Slack or e-mail; negotiation transcripts	Speer (2018) *PP*; Lix, Goldberg, Srivastava, and Valentine (2022) *MS*
Academic Articles	7%	Published academic articles used as analytical corpora	Alley, Kadolkar, Gupta, Cortina, and Winsler (2024) *JOM*
Patent Data	4%	USPTO abstracts; patent citations and claims; technology-space embeddings	Arts, Cassiman, and Gomez (2018) *SMJ*
Other	15%	Sensor logs; web-scraped organization websites; customer call transcripts	Fan et al. (2023) *JAP*; Green, Dalal, Fyffe, Zaccaro, Putka, and Wallace (2023) *JAP*

Note. Percentages reflect primary data source; articles employing multiple source types were coded by their predominant corpus.
SEC = U.S. Securities and Exchange Commission; 10-K = annual report filed with the SEC; 10-Q = quarterly report filed with the SEC; MD&A = management's discussion and analysis; DEF 14A = definitive proxy statement filed with the SEC; USPTO = U.S. Patent and Trademark Office. *JOM = Journal of Management; MS = Management Science; SMJ = Strategic Management Journal; OS = Organization Science; PP = Personnel Psychology; JAP = Journal of Applied Psychology.*

language model workflows; and (iv) insufficient transparency in reporting computational pipelines. These four problems parallel the stages of a CL research project—from selecting text, to choosing an analytical method, to configuring a model, to reporting results—and we organize our analysis of the CL literature accordingly.

Data Sources and Construct-Corpus Alignment

The first problem begins with the field's choice of linguistic data. Table 2 reveals that scholars rely on seven broad families of textual data, each carrying unique benefits and inferential risks, but collectively raising important questions about over-reliance on convenience sampling. Social media posts are the single most common source (30%, $n = 106$), well-suited for studying cultural beliefs (Corritore, Goldberg, & Srivastava, 2020) and public sentiment (Zhang et al., 2021), although platform algorithms, demographic skews, strategic manipulation, and, increasingly, the proliferation of AI-generated content introduce selection biases

that call for careful validation (Manzoor, Chen, Lee, & Smith, 2024). Corporate filings such as 10-K filings and earning calls (e.g., Liang, Cavusoglu, & Hu, 2023; Parker, Short, Titus, Gong, & Nahm, 2024) account for 21% of studies ($n = 75$). While their quasi-standardized templates facilitate large-scale comparisons, their language is also shaped by disclosure rules, legal review, and investor-facing incentives, which can dampen or reshape the linguistic traces relevant to the focal construct. Press and media sources constitute 13% ($n = 45$), employee and applicant texts 10% ($n = 35$), patent documents 4% ($n = 14$), and the remaining categories account for the balance.

A gap across all source families is that most studies do not justify their choice of linguistic text. Justifying corpus selection should address five considerations: (a) Theory: why the language in this corpus is expected to contain traces of the focal construct; (b) Authorship: who produced the text and under what incentives or constraints, including whether it was human-authored or machine-generated; (c) Audience: who the text was written for, since language registers shift with intended readership; (d) Medium: whether the source is written or transcribed from speech, since these registers differ systematically; and (e) Level of Construct: whether the theoretical claim is operationalized at the word, sentence, document, or corpus level, and whether the chosen unit of analysis matches. Explicitly addressing these dimensions transforms corpus selection from an unstated convenience decision into a defensible design choice—and provides the foundation for diagnosing construct-corpus misalignment before analytical choices compound it.

Method Selection and Model Complexity

Once a corpus is selected, the second problem concerns how scholars choose among analytical tools. Rather than cataloging applications by substantive topic, we organize this assessment around method families to show how method-selection problems manifest across levels of analysis. Lexical-signal methods grounded in word-counting remain widely used, and for good reason. They are most appropriate when the focal construct is indicated primarily by the presence, absence, or relative frequency of identifiable linguistic markers, and when those markers have sufficiently stable meanings within the corpus being studied. In such cases, the inferential task is not to recover the full meaning of a sentence, but to detect whether particular kinds of language are used more or less often—for example, hedging terms, gratitude expressions, first-person pronouns, or agentic versus communal descriptors. Micro-level studies have used bag-of-words dictionaries to code agentic and communal content in recommendation letters (Madera et al., 2009), self-verification motives from job interviews (Moore, Lee, Kim, & Cable, 2017), and expressions of gratitude from survey responses (Locklear, Taylor, & Ambrose, 2021). At the macro level, Eklund and Mannor (2021) constructed strategic issue categories using dictionary-based methods, while other studies use sentiment and tone measures to analyze chief executive officer (CEO) communications (Liang et al., 2023; Parker et al., 2024).

At the distributional level, unsupervised methods, such as topic modeling, operate on latent co-occurrence patterns rather than pre-defined word lists, and have been productively applied to reveal thematic structure (Chen & Mankad, 2024). The critical question, then, is not whether context ever matters—it almost always does—but whether valid measurement of the focal construct in the empirical setting depends materially on interpreting how words

interact within sentences and passages. When it does, lexical approaches become insufficient. Constructs such as strategic intent, impression management, or emotional framing often depend on negation, qualification, contrast, or sequencing, such that the same term can imply different meanings across utterances. Under those conditions, counting word frequencies to infer managerial intent can yield erroneous conclusions.

It is in these cases that the first and second problems converge. The issue is not that words are ever fully context-free; rather, it is that researchers sometimes rely on lexical tools when contextual variation is systematic enough to alter the construct being measured. Scholars routinely apply dictionary-based tools to constructs whose definitions imply compositional meaning. In such cases, the relevant diagnostic is whether plausible shifts in surrounding language would change how a reasonable reader interprets the focal term in ways that matter for the theoretical claim. A general dictionary measuring "risk," for example, may capture different constructs in a pharmaceutical earnings call than in an employee safety survey. In similar fashion, a sentiment lexicon scoring "aggressive" may indicate interpersonal hostility in one context but competitive intensity in another (Short, Broberg, Cogliser, & Brigham, 2010). Few studies report diagnostics that would reveal whether such contextual instability is central enough to threaten construct validity.

To address these problems, a growing body of work uses transformer-based models, fine-tuned classifiers, and contextual embeddings to capture semantic relationships that context-free methods cannot. At the micro level, these tools have been used to assess leadership behaviors (Green et al., 2023), work attitudes (Speer, Perrotta, Tenbrink, Wegmeyer, Delacruz, & Bowker, 2023), cognitive ability (Hickman, Tay, & Woo, 2024), and personality (Fyffe, Lee, & Kaplan, 2024; Hernandez & Nie, 2023), while HR researchers have examined CL tools for personnel selection (Koenig et al., 2023) and bias mitigation (Campion et al., 2024). At the macro level, embedding models are increasingly used to examine conceptual relationships; but, if not carefully validated against management theory, they risk conflating superficial word similarity with genuine conceptual similarity (Aceves & Evans, 2024). These methods demonstrate the greatest incremental value when constructs require disambiguating meaning across utterances rather than counting discrete markers—precisely the conditions under which the second problem is absent, and the tool's sophistication is warranted.

These challenges illuminate two important failure modes in choosing appropriate CL tools: under-engineering (i.e., applying lexical-signal tools to compositional constructs) and over-engineering (i.e., applying compositional-signal tools to lexical constructs). As compositional methods rapidly proliferate across the management field, we face increasing risks from unnecessary model complexity, which obscures the construct-to-measure mapping, and makes validity claims harder to evaluate. The problems are amplified in situations where CL methods fail to account for what organizations choose *not* to say, such as when employees withhold information (Dyne, Ang, & Botero, 2003; Morrison & Milliken, 2000). Emerging research on the role of strategic silence, euphemism, and deliberate omission indicates that what is left unsaid can be as consequential as explicit speech (Qin, Luo, Schifeling, & Wang, 2024; Suslava, 2021; Brauer, Wiersema, & Binder, 2023). Yet, most CL tools are built to detect word presence rather than meaningful absence. In these cases, the inferences researchers draw from analyzing organizational language are systematically shaped and limited by the affordances of their methods: when tools measure only what is said, what is left unsaid becomes invisible.

The Opacity of LLM Workflows

The third problem concerns opacity in how LLM-based measurement is implemented. As scholars adopt large language models for annotation, classification, and construct measurement, key design choices—such as model version, prompt wording, system instructions, examples, and generation settings—often remain hidden or underreported. This opacity matters not because it is separate from validity and reliability, but because it obscures the basis on which those qualities must be evaluated. Our analysis of the literature suggests that the adoption trajectory differs significantly across levels. Micro-level researchers often bring a well-established psychometric tradition to bear on new tools, routinely reporting reliability, validity, and bias diagnostics (Short, McKenny, & Reid, 2018). Macro-level research, by contrast, is far less likely to report comparable validation evidence, even as it scales context-sensitive models across thousands of texts. The result is an asymmetry in which the field's most advanced tools receive the least scrutiny precisely where they are deployed most extensively.

Several studies productively combine methods across signal types (e.g., Bezrukova, Thatcher, Jehn, & Spell, 2012; Stajkovic & Stajkovic, 2025), although multi-method designs create additional validation demands because each component introduces its own assumptions. One notable trend in micro-level research is an increasing focus on how CL techniques should and should not be used. For example, Hickman, Huynh, Gass, Booth, Kuruzovich, and Tay (2025) proposed a four-stage model to assess and mitigate algorithmic biases, and micro-level studies that explicitly analyze validity and reliability provide important exemplars for all organizational researchers (Hickman et al., 2026; Koenig et al., 2023; Speer et al., 2026). Yet, this validation tradition disproportionately clusters in micro-level research, creating an imperative for macro-level studies to adopt comparable standards.

Transparency and Disclosure

The fourth problem concerns insufficient disclosure of how CL-based measures are produced. Across our corpus, many studies do not report the intermediate steps linking raw text to construct measurement. In 262 cases (74%), authors omit descriptions of preprocessing pipelines, including tokenization rules, stop-word lists, stemming or lemmatization decisions, model parameters, and computing environment. In 191 studies, constituting approximately 54% of our sample, authors did not share code, schemes, or data. These reporting gaps make it difficult to assess scalability, reproduce analyses across different texts, or determine whether findings reflect meaningful patterns or artifacts of undocumented decisions.

These omissions are especially consequential because CL-based measurement often involves an indirect, model-dependent link between text features and the intended theoretical construct (Hickman, Thapa, Tay, Cao, & Srinivasan, 2022). In our review, 168 of 353 articles (47.6%) offered limited or no evidence of discriminant, convergent, predictive, or out-of-sample validity. When validation evidence is limited, disclosure becomes even more important because it allows readers to evaluate how measures were constructed and where inferential risks may have entered. This gap likely reflects the growing accessibility of high-level CL toolkits that enable researchers to obtain results without fully understanding what intermediate steps entail, rather than a deliberate decision to withhold information (Nyberg et al., 2025).

Hyde, Bachura, Bundy, Gretz, and Sanders (2024) provide an instructive exemplar of what stronger disclosure looks like by making their code publicly available, documenting preprocessing decisions and parameter specifications, and providing sufficient detail for readers to reproduce the full construct-measurement pipeline. Ultimately, the assessability of CL-based research depends on three practices: (i) documenting preprocessing choices; (ii) justifying each analytic technique; and (iii) releasing complete code and data when feasible. Disclosure does not itself ensure sound measurement, but, without it, validity claims remain unevaluable—and cumulative inference impossible (Bergh, Sharp, Aguinis, & Li, 2017). No matter how sophisticated the tool, it remains essential to ground CL efforts in organizational theories that specify the mechanism linking language to the construct of interest (Choudhury et al., 2019; Hannigan, Haans, Vakili, & Jennings, 2019).

Overall, the four problems documented in this section—construct-corpus misalignment, unnecessary model complexity, the opacity of LLM workflows, and insufficient transparency in reporting—are not isolated shortcomings but interconnected features of a rapidly evolving methodological landscape. The following section distills these problems into a set of integrative principles designed to function as durable decision rules for CL research in management.

From Review to Design: Integrative Principles for Computational Linguistics Research

From Patterns to Principles

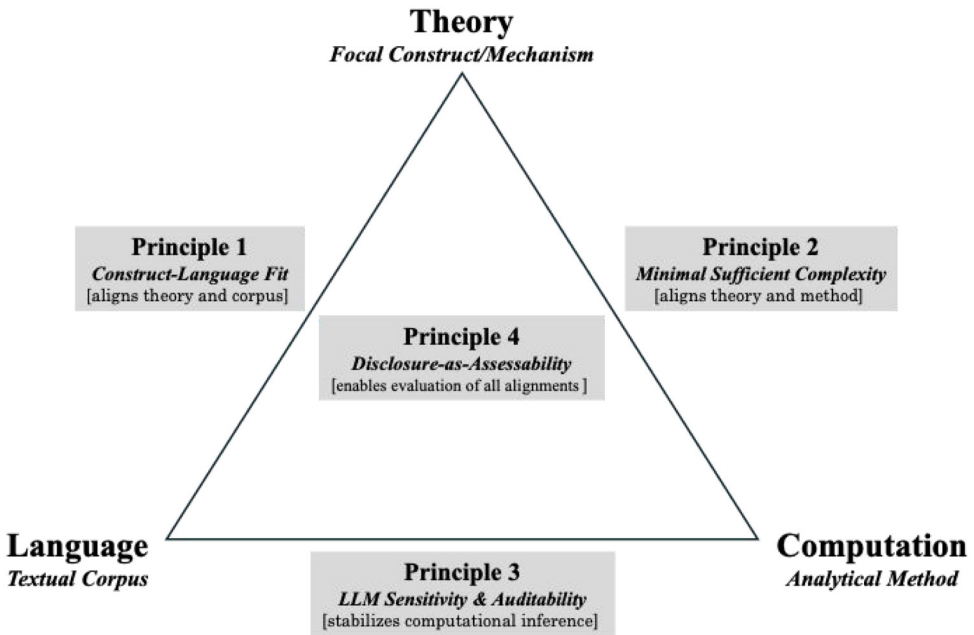
In the previous section, we identified four interconnected problems across our corpus of 353 CL studies. We now turn to a description of four integrative principles designed to function as decision rules (Nyberg et al., 2025): construct-language fit, minimal sufficient complexity, LLM sensitivity and auditability, and disclosure-as-assessability. Each principle addresses one of the four problems documented in the preceding section, and each emerges from regularities that become visible only when micro and macro applications are examined jointly rather than in isolation. Together, they provide a normative framework for resolving the alignment problems documented in the preceding section. Figure 3 summarizes the integrative framework that emerges from our review, illustrating how the four principles connect theory, language, and computational analysis.

As Figure 3 demonstrates, CL-related problems in management and organizational research arise when Theory, Language, & Computation are misaligned. The four Principles facilitate alignment in ways that are highly instrumental to the use of CL tools and the advancement of the field.

Principle 1: Construct-Language Fit

Principle 1 addresses the first problem identified in our review: construct-corpus misalignment. The credibility of any CL-based inference depends on a three-way alignment among the focal construct, the textual data used to measure it, and the analytical method applied to that data. We refer to this alignment as construct-language fit. Although the importance of aligning construct and operationalization may seem axiomatic, our review shows that such alignment is often assumed rather than demonstrated. Recall that 47.6% of studies

Figure 3
Framework for aligning theory, language, and computation in organizational research



in our sample offered limited or no evidence of construct validation, and most studies did not justify their choice of linguistic text. These are not independent problems. They are symptoms of a field that has not yet developed shared standards for diagnosing whether the textual data being analyzed actually contains traces of the construct being theorized.

We recommend three complementary approaches for diagnosing construct-language fit. The first is to determine the theoretical abstraction level of the focal construct: whether the researcher is measuring directly observable linguistic behavior (e.g., emotion word choice, hedging frequency) or an underlying latent state (e.g., strategic intent, organizational culture). This abstraction level is a property of the theory, not of the construct label, so the same label can demand different methods depending on how it is theorized. "Tone," for example, is well served by a lexical method when it is conceptualized as the relative frequency of positive and negative terms, but it implies a compositional construct when it is conceptualized as a stance that emerges from how statements qualify, contrast, or undercut one another. Construct-language fit is therefore assessed against the construct as the theory invokes it, not against the construct in name.

The second is to assess the linguistic register of the source corpus—the degree to which the text reflects how focal actors actually communicate about the construct of interest, as opposed to how they communicate in a genre governed by different conventions. A 10-K filing reflects what securities law requires management to disclose, but these filings do not

necessarily reflect what management believes or intends to do. A Glassdoor review reflects what a departing employee chooses to make public, shaped by platform norms and audience expectations, but these reviews do not always reflect the culture or climate of the organization. The third is to assess the semantic granularity of the chosen method—whether the CL tool captures meaning at the level of individual words (lexical signal), at the level of patterns among co-occurring words (distributional signal), or at the level of full sentences and passages where meaning depends on how words interact with one another (compositional signal). As Table 1 illustrates, these three signal types correspond to distinct method families, each suited to different types of constructs.

When these three dimensions are misaligned, measurement validity is compromised regardless of the CL tool's sophistication. Using a context-free, general dictionary to operationalize a latent construct that manifests through compositional phrasing (e.g., using word frequency to measure "strategic intent") elevates the risk of conflating surface vocabulary with deep meaning. Similarly, a context-sensitive model can fail when the text it learns from does not reflect how the focal actors actually express the construct of interest. Fine-tuning a transformer on earnings call transcripts to classify emotional states, for instance, trains the model on language that has been filtered through legal review and investor-relations coaching. The resulting classifier may learn to detect the conventions of corporate disclosure rather than the psychological states the researcher intends to measure. In practice, diagnosing construct-language fit requires attention to the considerations we identified. When these are absent—as they are in the majority of studies we reviewed—the field cannot distinguish valid measurement from systematic artifact.

Principle 2: Minimal Sufficient Complexity

Principle 2 addresses the second problem identified in our review: unnecessary model complexity. As compositional-signal methods become widely accessible, management scholars face a broader menu of increasingly powerful tools. The central methodological question is therefore not whether more sophisticated models exist, but whether they are necessary for the inferential task at hand. When a study asks whether certain words appear more or less frequently in a corpus—that is, a lexical question about vocabulary—a simple word-counting approach provides a direct, transparent answer. That is why lexical approaches remain well suited for constructs such as agentic versus communal language in recommendation letters (Madera et al., 2009), self-verification motives in job interviews (Moore et al., 2017), and tone in CEO communications (Liang et al., 2023; Parker et al., 2024). By contrast, contextual models are most valuable when the construct depends on sentence-level interpretation, as in work attitudes, leadership behaviors, or personality assessment (Green et al., 2023; Hernandez & Nie, 2023; Speer et al., 2023). Problems arise when scholars reverse this matching logic—either by using contextual models where relatively stable lexical markers are sufficient, or by relying on word-counting methods where meaning depends on negation, qualification, contrast, or sequencing. We formalize this observation as a principle of minimal sufficient complexity: the appropriate level of model complexity should be determined by the linguistic mechanism the theory implies, not by the novelty or availability of the algorithm.

This principle guards against two related failure modes. *Over-engineering* occurs when researchers deploy transformer models for tasks in which dictionaries or simpler supervised models would provide equally defensible answers with greater transparency. The risk is not

merely inefficiency: when the model is more complex than the construct demands, the additional parameters can obscure the construct-to-measure mapping and make validity harder to evaluate independently. Under-engineering occurs when researchers apply lexical-signal methods to constructs whose meaning is inherently compositional, producing measures that may appear reliable while missing the contextual features on which the construct depends. A bag-of-words approach cannot distinguish strategic negativity from genuine distress, or a hedged commitment from a confident one, because these distinctions reside in relationships among words rather than in the words themselves. As increasingly accessible contextual models make the first error more tempting, and long-standing dictionary habits sustain the second, the central methodological task is to match method complexity to theoretical mechanism rather than to methodological fashion.

Principle 3: LLM Sensitivity and Auditability

The opacity of LLM workflows documented in the preceding section motivates a principle specific to the means of deploying large language models for annotation, classification, and construct measurement. We formalize this as "LLM sensitivity and auditability": when LLMs are used to classify or score textual data, researchers must demonstrate that inferences are robust to the configuration choices governing model output, and must document those choices for independent evaluation. This principle addresses a form of inferential fragility distinct from the complexity-matching problem in Principle 2. Even when a compositional-signal method is theoretically warranted, LLM implementations introduce configuration decisions that can alter results in ways that are difficult to detect and rarely reported. Prompt wording is the most consequential, where small changes in phrasing, task framing, or example ordering can substantially shift classification outputs, yet most studies treat prompts as fixed instruments rather than as design parameters requiring sensitivity analysis (Shalpegin et al., 2025). Model version is a second source of fragility, as commercial providers update weights without notice, meaning that results may not be reproducible even with identical prompts. Temperature and sampling parameters introduce stochastic variation across successive passes. The interaction among these choices creates a combinatorial space of configurations that is rarely explored systematically.

The practical consequence is that two teams studying the same construct on the same corpus can obtain meaningfully different results through equally defensible configuration choices, without either recognizing the divergence as an implementation artifact. This fragility does not invalidate LLM-based measurement, but it does impose a distinct demand, which we term auditability, where researchers should demonstrate that LLM-based results remain stable across plausible choices about prompts, model versions, and generation settings. At minimum, LLM-based studies should report the exact model's name and version, the full prompt text including system instructions and few-shot examples, temperature and sampling parameters, and sensitivity checks demonstrating stability across plausible alternative prompt formulations. The micro-level validation tradition offers instructive models and reporting standards that are useful for the entire field. For example, Hickman, Langer, Saef, and Tay (2025) advanced a staged framework for assessing algorithmic bias, and Speer et al. (2026) developed reliability standards for CL measurement directly applicable to LLM workflows. The imperative is for macro-level research, where LLM adoption is accelerating most rapidly, to adopt comparable standards that treat prompt design and model configuration as reportable design choices rather than background implementation details.

Principle 4: Disclosure for Assessability

The chronic disclosure gaps documented throughout our review motivate a fourth principle, which we refer to as "disclosure-as-assessability." Transparency does not create validity, but it determines whether validity claims can be assessed, and therefore whether cumulative inference is possible (Bergh et al., 2017). This distinction matters because the field's current norms often conflate the two—treating disclosure as if it were a validity-enhancing procedure rather than the infrastructure that allows others to evaluate whether validity was achieved. Without documentation of preprocessing choices, model parameters, and prompt specifications, reviewers and replicators cannot determine whether a study's inferences are sound.

This logic applies with particular force to CL research, where the chain from raw text to construct measurement involves an unusually large number of under-documented decision points. As we reported in the previous section, 74% of studies in our sample omit sufficient descriptions of preprocessing pipelines, and 54% do not share coding schema or data. The consequence of this opacity is not necessarily that these studies draw erroneous inferences; rather, it is that their conclusions cannot be independently evaluated. An undocumented preprocessing pipeline is not necessarily flawed, but it is unevaluable—much as an unstated identification assumption in causal inference is not necessarily wrong but is unverifiable. Both render the study's conclusions a matter of faith rather than evidence. The principle of disclosure-as-assessability, therefore, positions documentation not as an ancillary virtue to be practiced when convenient, but as a precondition for scientific inference. Table 3 below provides our two-tier framework for resolving documentation decision points.

Hyde et al. (2024) demonstrate what this standard looks like in practice: public code, documented preprocessing decisions, parameter specifications, and sufficient detail for readers to reproduce the full construct-measurement pipeline. The distance between this standard and the field's current practices—where the majority of studies omit most of this information—underscores the scale of the challenge. Closing this gap is a necessary condition for CL research to produce findings that the field can reliably build upon.

The Interdependence of Principles

Taken together, these principles clarify that the credibility of computational linguistics research in management does not depend on algorithmic sophistication alone, but on the alignment among theory, language, and analytic technique. Construct-language fit ensures that the text under analysis contains meaningful traces of the theoretical construct. Minimal sufficient complexity calibrates model sophistication to the linguistic mechanism implied by theory. LLM sensitivity and auditability address whether LLM-based results remain robust across plausible choices about prompts, model versions, and generation settings. Disclosure-as-assessability addresses a different question: whether a study provides enough information about its analytic pipeline for others to evaluate, interrogate, or approximate it. The two are related but distinct. A study may fully disclose its workflow yet fail to demonstrate robustness, or it may report robustness checks without providing enough information for others to assess how those checks were conducted.

These four principles are not independent prescriptions to be applied in sequence. Construct-language fit is unverifiable without the disclosure standards specified by Principle 4. Minimal sufficient complexity is unjustifiable without the alignment specified by Principle 1, because the "right" level of complexity depends on the nature of the construct, which in turn depends on whether the text actually contains traces of it. LLM auditability is a special

Table 3
Design and Disclosure Decision Matrix for Computational Linguistics Research

Design Question	What the Author Must Decide	Tier 1: Minimum Viable Disclosure	Tier 2: Exemplary Disclosure
Does this corpus plausibly contain the construct?	Whether the text source is a valid site for observing the focal construct	Identify the text source, time period, sample size, and focal construct	Justify corpus choice using theory, authorship, audience, medium, and level of construct; explain what the corpus can and cannot reveal
What kind of linguistic signal does the theory imply?	Whether the construct is primarily lexical, distributional, or compositional	Name the method family and explain why it matches the construct's linguistic mechanism	Explain why adjacent method families were not selected; compare against at least one simpler alternative on a validation subset
Could preprocessing choices materially change the result?	Whether tokenization, stop-word removal, stemming, lemmatization, or normalization affect meaning	Report all preprocessing steps	Provide code or pseudocode and show sensitivity to plausible preprocessing alternatives
If using an LLM, how fragile is the measurement?	Whether outputs depend on prompt wording, model version, or generation settings	Report model name/version, prompt text, temperature, top-p, and access date	Show sensitivity across plausible prompts and repeated runs; report inter-prompt or inter-run agreement
What evidence shows that the measure captures the construct?	Whether the operationalization has credible construct validity	Report at least one validation check, such as human comparison, hold-out performance, or convergent/discriminant evidence	Report multiple validation forms, classification diagnostics, and ideally out-of-sample validation on a different corpus or period
Can another scholar evaluate or approximate this study?	Whether the study is assessable even if full replication is impossible	State whether code and data are available; explain any restrictions	Share full code, preprocessing pipeline, and data or synthetic substitutes sufficient for independent evaluation where feasible

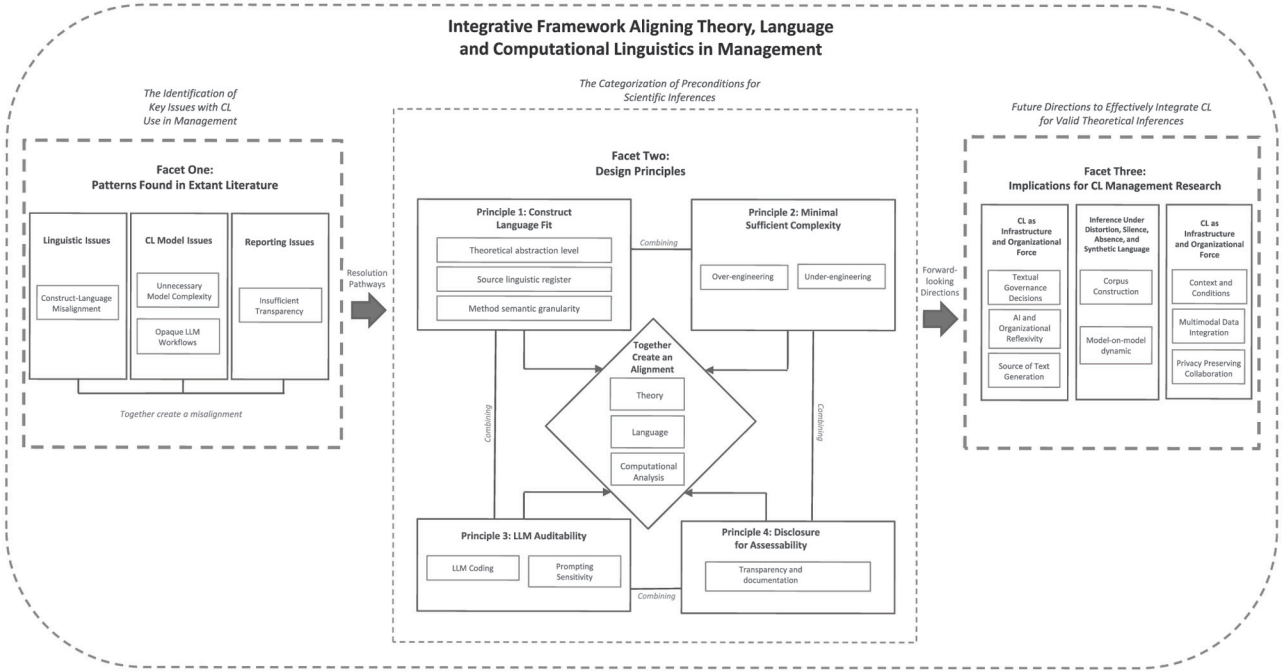
case of the general disclosure requirement, intensified by the opacity of specific tools. And disclosure-as-assessability is rendered hollow if what is disclosed reveals no attention to fit or complexity calibration. Partial compliance is insufficient: a study that documents its preprocessing pipeline but applies a context-free method to a compositional construct has disclosed a flawed design rather than a sound one.

The four principles transform CL tools and technologies into a coherent methodological infrastructure that supports cumulative organizational research. This transformation also highlights a broader shift in which CL is increasingly embedded not only in researchers' analytical workflows, but also in the organizational processes that researchers study—shaping how firms communicate, how stakeholders interpret signals, and how fields converge on shared rhetorical forms. In the following section, we identify key implications arising from CL's dual role as both an analytical tool and an organizational infrastructure.

Implications for the Future of Computational Linguistics in Management and Organizational Research

Our review suggests that the next phase of CL will be shaped less by incremental gains in predictive accuracy than by whether the field can turn CL's engineering power into findings that other scholars can evaluate, replicate, and build on. Figure 4 summarizes the broader implications that emerge from our review.

Figure 4
Computational linguistics in management and organizational research: patterns and principles



Building on the integrative principles developed above, we organize these implications into three domains: implications for theory; implications for research design and inference; and implications for the cumulative advancement of management and organizational science. Together, these domains extend the central premise emerging from our review: that the value of CL lies not in increasing analytical power, but in the principled alignment of what is theorized, what is observed in text, and what is computationally extracted from it. In doing so, they reposition CL as part of the methodological and socio-technical infrastructure through which organizational phenomena are expressed, analyzed, and understood.

Implications for Theory: Language, Absence, and Meaning in Organizations

The tools and procedures attendant to the computation of linguistic structures and content not only expand the scale at which language can be analyzed, but, in a myriad of ways, CL reshapes how management and organizational scholars theorize the relationship between language, meaning, and action. This is true at both the micro-level (i.e., individuals, dyads, and teams) and the macro-level (i.e., firms, sectors, markets, and populations) of analysis. Equipped with increasingly sophisticated tools, scholars will have the capacity to extract insights from the structure and content of language that have previously passed largely unnoticed. Nowhere is this more apparent than in analyses of when and how language is used, or not used, in organizational settings (Butler, 2021; Dimitrov, 2019; Dimitrov, Reilly, & Sarkar, 2022). Across levels and throughout the studies we reviewed, language is typically treated as a positive signal, meaning it typically comprises observable textual traces that can be used to infer key drivers of decisions, actions, and outcomes, including attitudes, sentiments, intentions, or strategic positioning. Yet, the principles emanating from our review—particularly those pertaining to construct–language fit—highlight a complementary and under-theorized dimension of organizational communication regarding the role and impact of absence or silence. A regimented focus on what is explicitly communicated certainly provides a certain set of insights, but just because that which is communicated is accessible to scholars for analysis does not mean that is the entire story. That which is not said—as well as that which is intentionally or unintentionally excluded, filtered, or suppressed—may be as theoretically consequential as that which is observed.

This insight follows from the structure of many organizational communication processes. As shown in our review of data sources, widely used corpora such as earnings calls, corporate disclosures, and public-facing communications are not neutral reflections of organizational reality. Rather, they are shaped by legal, reputational, and institutional constraints that govern both expression and omission (Carlos & Lewis, 2018; Suchman, 1995). Similarly, a verbatim transcript of what is said in a project team meeting is unlikely to fully convey how power dynamics, inter-personal relationships, and unstated contextual factors determine what is said and what is left unsaid (Dyne et al., 2003; Morrison, 2023; Szkudlarek & Alvesson, 2023). Yet, both the said and the unsaid are essential building blocks of cognitive linguistics (Tyler, 1978), which is a close cousin of computational linguistics. Linguistic data often exhibit systematic absences that are endogenous to the phenomena under study. The omission of risk-related language, the avoidance of controversial topics, or the suppression of dissent may reflect deliberate personal, tactical, and strategic choices rather than measurement error. Under these conditions, absence is not missing data at all; rather, it is theoretically meaningful silence.

Most CL methods, particularly those relying on observable word frequencies or semantic patterns, are not designed to detect the existence of absence directly, much less its role and impact. This becomes a validity problem in the specific case where the focal construct is itself partly constituted by what is left unsaid. Collective-focus, for instance, can be conveyed both by emphasizing the collective and by deemphasizing the individual, so operationalizing it through the frequency of collective terms alone, without attention to a reduced use of individual-focused terms, risks conflating observed language with the underlying construct. Where the theory does not implicate absence, ignoring silence poses no such threat. A study linking the use of third-person pronouns to organizational citizenship behavior, for example, need not measure the absence of first- and second-person pronouns, because its construct does not depend on them. The risk, then, is not that any study ignoring silence is flawed, but that studies whose constructs implicate absence must account for it to preserve construct-language fit. Quite critically, even with context-sensitive models and LLM-based approaches, this inherent limitation persists, creating a high risk of generating seemingly valid interpretations of what is communicated while remaining insensitive to what is omitted.

Increasing methodological sophistication does not eliminate this problem, although it may obscure it by producing coherent representations of incomplete signals. It falls upon the shoulders of scholars to monitor and mitigate these effects, which will in turn improve the quality and usefulness of theories. In the end, recognizing silence as theoretically meaningful opens new avenues for research. Rather than treating language as a transparent window into organizational processes, scholars can conceptualize communication as a selective filtering mechanism through which actors actively construct both presence and absence. This perspective builds on foundational theories of legitimacy, impression management, and symbolic interaction (Goffman, 1959; Suchman, 1995) and extends them to algorithmically mediated environments. AI is likely to dramatically accelerate the importance of accounting for selective filtering (Chalmers, Hunt, Pachidi, Potočnik, & Townsend, 2026). As organizations increasingly anticipate how their communications will be parsed by automated systems, algorithms, regulatory monitors, or AI-driven sentiment tools, individuals and organizations will adjust not only what they say, but what they leave unsaid. In such contexts, silence becomes a strategic resource that shapes both human and machine interpretation, leading to new perspectives on how absence and silence influence the relationship between observed language and underlying constructs across levels of analysis.

Implications for Research Design and Inference: CL Under Conditions of Distortion

A second major implication of our review concerns the conditions under which CL methods yield, or fail to yield, valid inference in organizational research. Across the studies we examined, a common assumption is that textual data provide sufficiently accurate representations of underlying constructs, enabling computational analysis to recover meaningful patterns (Hannigan et al., 2019; Nelson, 2020). However, the principles developed in this review—consisting of construct–language fit, minimal sufficient complexity, and LLM sensitivity and auditability—highlight that this assumption is often extremely fragile to tool selection, investigative procedures, and the complications of selective disclosure. In many

organizational contexts, linguistic data are systematically distorted, whether through strategic communication, institutional constraints, or algorithmic mediation. Under these conditions, the central challenge is not extracting signal from noise but understanding how the signal itself is transformed before it becomes observable.

As our review reveals, distortion arises at multiple stages of the data-generating process. Organizational texts are frequently shaped by interpersonal censoring, legal vetting, reputational concerns, hierarchical editing, and platform-specific norms (Gorwa & Guilbeault, 2020; Suslava, 2021). These processes do not merely attenuate linguistic signals (Jaeger & Buz, 2017); they often reshape them by altering both content and distribution in ways that are endogenous to the constructs under study. For example, earnings call language reflects not only managerial beliefs but also impression management and expectation-setting (Huang, Teoh, & Zhang, 2014; Mayew & Venkatachalam, 2012). Similarly, internal communications between members of an organization may exhibit convergence or suppression driven by power dynamics or organizational culture (Detert & Edmondson, 2011). In such settings, observed text may function as a transformed representation of the focal construct, but it may also capture a related yet distinct construct generated by the communication environment itself, such as strategic self-presentation, genre compliance, or reputational management. The inferential task, therefore, is not simply to recover an underlying construct from distorted language, but to determine whether the text operationalizes the intended construct, a transformed proxy for it, or a plausible rival construct altogether. Doing so requires validation strategies that examine the nomological pattern of the resulting measure against both the focal construct and theoretically credible alternatives.

The increasing use of context-sensitive models and LLMs further complicates this inferential landscape, dramatically escalating the risk of biased, spurious, or utterly unreliable measures. Even when best practices are employed, and scholars are able to capture some semblance of semantic nuance (Devlin, Chang, Lee, & Toutanova, 2019), the methods introduce new sources of uncertainty that are often underreported. As emphasized by Principle 3, concerning LLM sensitivity and auditability, model outputs depend on prompt design, parameterization, and training data—factors that materially influence results yet are rarely disclosed (Carlson & Burbano, 2026; Shalpegin et al., 2025). When LLMs are used for annotation or classification, they effectively act as secondary data-generating processes, translating raw text into structured variables through mechanisms that are largely, if not wholly, opaque to scholars using them. This creates a layered inferential problem in which researchers must account for both distortion in the original text and distortions that are introduced by the model itself, in what can best be described as a sort of compound distortion in which even the most basic assumptions about measurement can be inadvertently violated, including stable construct representation and independent error (Egami, Fong, Grimmer, Roberts, & Stewart, 2018; Grimmer & Stewart, 2013). As a result, even sophisticated models may produce internally consistent but theoretically misaligned findings. This reinforces Principle 2, minimal sufficient complexity. In many situations, simpler and more interpretable models may yield more credible inference because their assumptions are transparent and testable (Rudin, 2019).

Not all of this is bad news for scholars. There are numerous ways in which recognizing that distortion is endogenous to linguistic data opens new opportunities for research design. Rather than treating distortion as noise, scholars can explicitly theorize and model it as part of the phenomenon under study. For instance, variation in distortion across contexts may

itself be theoretically informative, such as situations involving regulated and unregulated communication environments. Similarly, sensitivity analyses that vary prompts, preprocessing choices, or model architectures can serve not only as robustness checks but also as tools for theory development by revealing how alternative representations of language shape varied conclusions and the underlying drivers of diverse decisions and actions. In this sense, CL methods enable a more reflexive approach to inference, in which computational models are treated as objects of inquiry rather than neutral instruments (Nelson, 2020). At the micro level, these developments raise important questions about how individuals adapt communication under conditions of algorithmic visibility and evaluation. At the macro level, scholars may wish to highlight how institutional and technological infrastructures shape the distribution and transformation of language across organizations and fields (Fourcade & Healy, 2017). In both cases, CL requires a shift from viewing text as a stable input to treating it as a dynamic signal embedded in socio-technical systems.

Implications for the Advancement of Management and Organizational Science: Building an Assessable Computational Linguistics Infrastructure

A third implication of our review concerns the conditions under which research employing CL can accumulate into a coherent and reliable body of knowledge. Although the rapid adoption of CL methods has expanded the scope of management research, our analysis reveals a growing gap between methodological sophistication and the field's ability to evaluate, replicate, and extend empirical findings. This gap is evident in the widespread absence of documentation surrounding preprocessing pipelines, model configurations, and validation procedures—reflecting a broader absence of shared infrastructure for assessable inference. The logic of disclosure-as-assessability (Principle 4) provides a foundation for addressing this challenge. As our review shows, transparency is not simply a reporting norm; it is a necessary precondition for evaluating whether theoretical claims are empirically warranted (Aguinis, Ramani, & Alabduljader, 2020; Bergh et al., 2017). Without optics into the intermediate steps linking raw text to construct measurement, editors, reviewers, and subsequent researchers cannot determine whether findings reflect meaningful patterns or artifacts of undocumented decisions. Importantly, this challenge is not reducible to individual scholarly choices and behaviors. It is part of a broader shift in the epistemic structure of management research, in which increasingly complex and computationally intensive methods create asymmetries in who can produce, evaluate, and extend empirical work (Short et al., 2018). As advanced models and large-scale datasets become concentrated within specific research communities, the risk is not only fragmentation but also epistemic stratification, in which some scholars generate findings that others cannot readily scrutinize (Christensen, Freese, & Miguel, 2019).

Addressing this challenge requires moving beyond general calls for transparency toward shared standards that enable systematic assessment of CL-based research. Our two-tiered disclosure framework, introduced under Principle 4, offers one step in this direction: Tier 1 establishes a baseline of documentation that enables evaluation of core design choices, while Tier 2 extends this standard by encouraging the sharing of code, data (where feasible), and sensitivity analyses that demonstrate robustness across specifications. At the same time, the increasing use of proprietary models, restricted datasets, and privacy-sensitive organizational data complicates the implementation of open science norms. This constraint highlights the need for

alternative mechanisms of assessability, such as synthetic data, federated learning, and standardized reporting templates that preserve confidentiality while enabling independent evaluation. The key requirement is not universal openness, per se, but rather norms of transparency that provide sufficient information for others to understand, evaluate, and approximate the analytical process. Journals, reviewers, and scholarly communities play a central role in establishing these expectations, and, as computational methods become more deeply embedded in management research, the standards governing their use will shape not only the credibility of individual studies but also the trajectory of knowledge development in the field.

Conclusion

Ultimately, making words count in management research depends not on more powerful tools, but on using them in ways that advance the theoretical ambitions of the field. The four principles developed in this review—construct-language fit, minimal sufficient complexity, LLM sensitivity and auditability, and disclosure-as-assessability—provide durable decision rules for aligning theoretical constructs, linguistic evidence, and computational methods in ways that support credible and cumulative inference. Together with the signal-type taxonomy that organizes methods by the kinds of meaning they can access and the two-tiered reporting framework that calibrates disclosure to methodological complexity, these principles reframe CL not as a set of tools but as a methodological infrastructure shaping how organizational phenomena are conceptualized, measured, and evaluated. As our review demonstrates, this infrastructure must contend with the theoretically consequential roles of silence and absence, the systematic distortion of linguistic signals before they become observable, and the epistemic conditions required for findings to accumulate into reliable knowledge. As CL becomes more widely adopted, the central risk is that methodological sophistication substitutes for theoretical clarity, yielding findings that are precise but not meaningful, or scalable but not cumulative. Addressing this risk requires a reflexive approach—one that treats language as a selectively constructed signal, and computational models as part of the inferential process rather than neutral instruments (Grimmer & Stewart, 2013; Hannigan et al., 2019; Nelson, 2020). In this respect, our framework aligns with and extends broader calls for transparency, robustness, and theoretical grounding in management scholarship (Aguinis et al., 2020; Bergh et al., 2017).

ORCID iDs

Joseph J. Simpson https://orcid.org/0000-0002-3035-2969
Richard A. Hunt https://orcid.org/0000-0002-8893-2616
David M. Townsend https://orcid.org/0000-0001-7903-7831
Daniel J. Beal https://orcid.org/0000-0003-4750-2430

Notes

1. As many of these terms are technical, we provide a glossary in the appendices (see Appendix A1 in the online supplemental material) for readers. The glossary is intended to provide readers with a clear, plain-language explanation of these terms, rather than a polished definition.

2. We originally selected 2013–2023 for two reasons. First, CL works were sparse prior to 2013. Second, we wanted to cover a minimum of 10 years, consistent with prior reviews, and added 2024 and 2025 during the manuscript revision process.

References

Note: The references in the Reference List include only those that are referenced in the text. Please refer to Appendix A2 in the online supplemental material for the reference list used for our computational linguistics review.

Aceves, P., & Evans, J. 2024. Mobilizing conceptual spaces: How word embedding models can inform measurement and theory within organization science. *Organization Science*, 35: 788-814.

Aguinis, H., Ramani, R. S., & Alabduljader, N. 2020. Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organizational Research Methods*, 23: 465-488.

Alley, L., Kadolkar, I., Gupta, A., Cortina, J. M., & Winsler, K. P. 2024. Grammatical redundancy in scales: Using the "ConGRe" process to create better measures. *Journal of Management*. Advanced online publication.

Amaya, J., & Holweg, M. 2024. Using algorithms to improve knowledge work. *Journal of Operations Management*, 70: 482-513.

Antons, D., Breidbach, C. F., Joshi, A. M., & Salge, T. O. 2023. Computational literature reviews: Method, algorithms, and roadmap. *Organizational Research Methods*, 26: 107-138.

Arts, S., Cassiman, B., & Gomez, J. C. 2018. Text matching to measure patent similarity. *Strategic Management Journal*, 39: 62-84.

Bae, J. W., Belo, F., Li, J., Lin, X., & Zhao, X. 2023. The opposing effects of complexity and information content on uncertainty dynamics: Evidence from 10-K filings. *Management Science*, 69: 6313-6332.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization*, 15: 423-436.

Bezrukova, K., Thatcher, S., Jehn, K. A., & Spell, C. S. 2012. The effects of alignments: Examining group faultlines, organizational cultures, and performance. *Journal of Applied Psychology*, 97: 77.

Bolinger, M., Josefy, M., Stevenson, R., & Hitt, M. 2022. Experiments in strategy research: A critical review and future research opportunities. *Journal of Management*, 48: 77-113.

Brauer, M., Wiersema, M., & Binder, P. 2023. "Dear CEO and Board": How activist investors' confidence in tone influences campaign success. *Organization Science*, 34: 1487-1508.

Butler, J. 2021. *Excitable speech: A politics of the performative*. London: Routledge.

Campion, E. D., Campion, M. A., Johnson, J., Carretta, T. R., Romay, S., Dirr, B., & Mouton, A. 2024. Using natural language processing to increase prediction and reduce subgroup differences in personnel selection decisions. *Journal of Applied Psychology*, 109: 307-338.

Campion, M. A., & Campion, E. D. 2023. Machine learning applications to personnel selection: Current illustrations, lessons learned, and future research. *Personnel Psychology*, 76: 993-1009.

Carlos, W. C., & Lewis, B. W. 2018. Strategic silence: Withholding certification status as a hypocrisy avoidance tactic. *Administrative Science Quarterly*, 63: 130-169.

Carlson, N. 2023. Differentiation in microenterprises. *Strategic Management Journal*, 44: 1141-1167.

Carlson, N., & Burbano, V. 2026. The use of LLMs to annotate data in management research: Foundational guidelines and warnings. *Strategic Management Journal*, 47: 699-725.

Chalmers, D., Hunt, R., Pachidi, S., Potočnik, K., & Townsend, D. 2026. The acceleration of AI: Rethinking organization and work in an era of rapid technological change. *Journal of Management Studies*, 63: 285-314.

Chen, L., & Mankad, S. 2024. A structural topic and sentiment-discourse model for text analysis. *Management Science*, 71: 5419-6318.

Choudhury, P., Wang, D., Carlson, N., & Khanna, T. 2019. Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, 40: 1705-1732.

Christensen, G., Freese, J., & Miguel, E. 2019. *Transparent and reproducible social science research: How to do open science*. Berkeley, CA: University of California Press.

Colby, K. M., Parkison, R. C., & Faught, B. 1974. Pattern-matching rules for the recognition of natural language dialogue expressions. *American Journal of Computational Linguistics*. Microfiche 5: 1-82.

Cong, L. W., Liang, T. Y., Zhang, X., & Zhu, W. 2025. Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information. *Management Science*, 71: 9869-10753.

Corritore, M., Goldberg, A., & Srivastava, S. B. 2020. Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly*, 65: 359-394.

Detert, J. R., & Edmondson, A. C. 2011. Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, 54: 461-488.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. 2019. *BERT: Pre-training of deep bidirectional transformers for language understanding*. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics.

Dewan, Y., Simons, T., & Wernicke, G. 2024. The ideological imperative: Corporate social responsibility and news media coverage of firms. *Organization Science*, 35: 1930-1955.

Dimitrov, R. 2019. Explicit and implicit strategies of silence. *Journal of Advertising and Public Relations*, 2: 19-31

Dimitrov, R., Reilly, P., & Sarkar, S. 2022. Silencing algorithms: The role of absence in digital communication systems. *Information, Communication & Society*, 25: 787-805.

Dyne, L. V., Ang, S., & Botero, I. C. 2003. Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of Management Studies*, 40: 1359-1392.

Edge Delta. 2024. *Breaking down the numbers: How much data does the world create daily in 2024?* https://edgedelta.com/company/blog/how-much-data-is-created-per-day?utm_source=perplexity

Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., & Stewart, B. M. 2018. How to make causal inferences using texts. *Science Advances*, 8: Article 42.

Eklund, J., & Mannor, M. 2021. Keep your eye on the ball or on the field? Exploring the performance implications of strategic attention. *Academy of Management Journal*, 64: 1685-1713.

Evens, M., & Smith, R. N. 1978. A lexicon for a computer question-answering system. *American Journal of Computational Linguistics*. Microfiche 83: 1-93.

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. 2023. How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology*, 108: 1277-1299.

Fleiss, J. L. 1981. Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5: 105-112.

Fourcade, M., & Healy, K. 2017. Seeing like a market. *Socio-Economic Review*, 15: 9-29.

Fyffe, S., Lee, P., & Kaplan, S. 2024. "Transforming" personality scale development: Illustrating the potential of state-of-the-art NLP. *Organizational Research Methods*, 27: 265-300.

Goffman, E. 1959. *The presentation of self in everyday life*. New York, NY: Anchor Books.

Gorwa, R., & Guilbeault, D. 2020. Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12: 225-248.

Green, J., Dalal, R., Fyffe, S., Zaccaro, S., Putka, D., & Wallace, D. 2023. An empirical taxonomy of leadership situations. *Journal of Applied Psychology*, 108: Article 1515.

Grimmer, J., & Stewart, B. M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21: 267-297.

Guzman, J., & Li, A. 2023. Measuring founding strategy. *Management Science*, 69: 101-118.

Haans, R. F., & Mertens, M. J. 2024. The internet never forgets: A four-step scraping tutorial, codebase, and database for longitudinal organizational website data. *Organizational Research Methods*, 29: 113-141.

Hannigan, T., Haans, R., Vakili, K., & Jennings, P. 2019. Topic modeling in management research: Rendering theory from textual data. *Academy of Management Annals*, 13: 586-632.

Harrison, J. S., Josefy, M. A., Kalm, M., & Krause, R. 2023. Using supervised machine learning to scale human-coded data: A method and dataset in the board leadership context. *Strategic Management Journal*, 44: 1780-1802.

Hernandez, I., & Nie, W. 2023. The AI-IP: Minimizing the guesswork of personality scale item development through artificial intelligence. *Personnel Psychology*, 76: 1011-1035.

Hickman, L., Huynh, C., Gass, J., Booth, B., Kuruzovich, J., & Tay, L. 2025. Whither bias goes, I will go: An integrative, systematic review of algorithmic bias mitigation. *Journal of Applied Psychology*, 110: 979-1000.

Hickman, L., Langer, M., Saef, R. M., & Tay, L. 2025. Automated speech recognition bias in personnel selection: The case of automatically scored job interviews. *Journal of Applied Psychology*, 110: 846-858.

Hickman, L., Liff, J., Rottman, C., & Calderwood, C. 2026. The effects of the training sample size, ground truth reliability, and NLP method on language-based automatic interview scores' psychometric properties. *Organizational Research Methods*, 29: 40-72.

Hickman, L., Tay, L., & Woo, S. E. 2024. Are automated video interviews smart enough? Behavioral modes, reliability, validity, and bias of machine learning cognitive ability assessments. *Journal of Applied Psychology*, 110: 314-335.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. 2022. Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25: 114-146.

Hirschberg, J., & Manning, C. D. 2015. Advances in natural language processing. *Science*, 349: 261-266.

Huang, X., Teoh, S., & Zhang, Y. 2014. Tone management. *The Accounting Review*, 89: 1083-1113.

Hyde, S., Bachura, E., Bundy, J., Gretz, R., & Sanders, W. 2024. The tangled webs we weave: Examining the effects of CEO deception on analyst recommendations. *Strategic Management Journal*, 45: 66-112.

Hymer, C. B., & Smith, A. D. 2024. Making exceptions exceptional: A cross-methodological review and future research agenda. *Journal of Management*, 50: 2374-2402.

Jaeger, T. F., & Buz, E. 2017. Signal reduction and linguistic encoding. In E. M. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics*: 38-81. Wiley.

Jung, J., Zhou, W., & Smith, A. 2025. From textual data to theoretical insights: Introducing and applying the word-text-topic extraction approach. *Organizational Research Methods*, 28: 182-209.

Kittredge, R. I. 1983. Semantic processing of texts in restricted sublanguages. In N. Cercone (Ed.), *Computational linguistics*: 45-58. Oxford: Pergamon Press.

Kobayashi, V., Mol, S., Berkers, H., Kismihók, G., & Den Hartog, D. 2018. Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, 21: 766-799.

Koenig, N., Tonidandel, S., Thompson, I., Albritton, B., Koohifar, F., Yankov, G., Speer, A., Hardy, J. H., III, Gibson, C., Frost, C., Liu, M., McNeney, D., Capman, J., Lowery, S., Kitching, M., Nimbkar, A., Boyce, A., Sun, T., Guo, F., Min, H., Zhang, B., Lebanoff, L., Phillips, H., & Newton, C. 2023. Improving measurement and prediction in personnel selection through the application of machine learning. *Personnel Psychology*, 76: 1061-1123.

Kunisch, S., Denyer, D., Bartunek, J. M., Menz, M., & Cardinal, L. B. 2023. Review research as scientific inquiry. *Organizational Research Methods*, 26: 3-45.

Liang, P., Cavusoglu, H., & Hu, N. 2023. Customers' managerial expectations and suppliers' asymmetric cost management. *Production & Operations Management*, 32: 1975-1993.

Lix, K., Goldberg, A., Srivastava, S. B., & Valentine, M. A. 2022. Aligning differences: Discursive diversity and team performance. *Management Science*, 68: 8430-8448.

Locklear, L., Taylor, S., & Ambrose, M. 2021. How a gratitude intervention influences workplace mistreatment: A multiple mediation model. *Journal of Applied Psychology*, 106: Article 1314.

Madera, J. M., Hebl, M. R., & Martin, R. C. 2009. Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, 94: Article 1591.

Manzoor, E., Chen, G. H., Lee, D., & Smith, M. D. 2024. Influence via ethos: On the persuasive power of reputation in deliberation online. *Management Science*, 70: 1613-1634.

Mayew, W. J., & Venkatachalam, M. 2012. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, 67(1): 1-43.

Moore, C., Lee, S., Kim, K., & Cable, D. 2017. The advantage of being oneself: The role of applicant self-verification in organizational hiring. *Journal of Applied Psychology*, 102: Article 1493.

Morrison, E. W. 2023. Employee voice and silence: Taking stock a decade later. *Annual Review of Organizational Psychology and Organizational Behavior*, 10: 79-107.

Morrison, E. W., & Milliken, F. J. 2000. Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management Review*, 25: 706-725.

Nelson, L. K. 2020. Computational grounded theory: A methodological framework. *Sociological Methods & Research*, 49(1): 3-42.

Nyberg, A. J., Schleicher, D. J., Bell, B. S., Boon, C., Cappelli, P., Collings, D. G., Dalle Molle, J. E., Feuerriegel, S., Gerhart, B., Jeong, Y., Korsgaard, M. A., Minbaeva, D., Ployhart, R. E., Tambe, P., Weller, I., Wright, P. M., & Yakubovich, V. 2025. A brave new world of human resources research: Navigating perils and identifying grand challenges of the GenAI revolution. *Journal of Management*, 51: 2677-2718.

Padmanabhan, B., Fang, X., Sahoo, N., & Burton-Jones, A. 2022. Machine learning in information systems research. *MIS Quarterly*, 46: iii-xix.

Parker, O., Short, C., Titus, V., Gong, K., & Nahm, P. I. 2024. Accentuate the positive? Strategic negativity amid the hazard of high expectations. *Strategic Management Journal*, 45: 1851-1874.

Paruchuri, S., Hoempler, E. A., Cowen, A. P., Cannella, A. A., Jr., & Nahm, P. I. 2024. Governance failure and governance under failure: Reviewing the role of directors in organizational misconduct. *Journal of Management*, 50: 2237-2265.

Qin, M. S., Luo, X., Schifeling, T., & Wang, Y. 2025. When corporate silence is costly: Negative consumer responses to corporate silence on social issues. *Strategic Management Journal*, 46: 994-1026.

Ridge, J. W., Hill, A. D., Ingram, A., Kolomeitsev, S., & Worrell, D. L. 2024. Avoidance and aggression in stakeholder engagement: The impact of CEO paranoia and paranoia-relevant cues. *Academy of Management Journal*, 67: 1392-1415.

Rudin, C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1: 206-215.

Schmidt, C. G., Wuttke, D. A., Ball, G. P., & Heese, H. S. 2020. Does social media elevate supply chain importance? An empirical examination of supply chain glitches, Twitter reactions, and stock market returns. *Journal of Operations Management*, 66: 646-669.

Schmiedel, T., Müller, O., & Vom Brocke, J. 2019. Topic modeling as a strategy of inquiry in organizational research. *Organizational Research Methods*, 22: 941-968.

Schubert, L. 2020. Computational linguistics. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford, CA: CSLI.

Shalpegin, T., Browning, T. R., Kumar, A., Shang, G., Thatcher, J., Fransoo, J. C., Holweg, M., & Lawson, B. 2025. Generative AI and empirical research methods in operations management. *Journal of Operations Management*, 71: 578-587.

Short, J. C., Broberg, J., Cogliser, C., & Brigham, K. 2010. Construct validation using computer-aided text analysis. *Organizational Research Methods*, 13: 320-347.

Short, J. C., McKenny, A. F., & Reid, S. W. 2018. More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5: 415-435.

Siano, F. 2025. The news in earnings announcement disclosures: Capturing word context using LLM methods. *Management Science*, 71: 9831-9855.

Speer, A. B. 2018. Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, 71: 299-333.

Speer, A., Oswald, F., & Putka, D. 2026. Reliability evidence for AI-based scores in organizational contexts. *Organizational Research Methods*, 29: 412-440.

Speer, A., Perrotta, J., Tenbrink, A., Wegmeyer, L., Delacruz, A., & Bowker, J. 2023. Turning words into numbers: Assessing work attitudes using natural language processing. *Journal of Applied Psychology*, 108: 1027-1045.

Stajkovic, A. D., & Stajkovic, K. 2025. A summer of protest: Using event system theory to test an intersectional leadership advantage. *Journal of Management*, 51: 1201-1230.

Suchman, M. C. 1995. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20: 571-610.

Suslava, K. 2021. "Stiff business headwinds and uncharted economic waters": The use of euphemisms in earnings conference calls. *Management Science*, 67: 7184-7213.

Szkudlarek, B., & Alvesson, M. 2023. Doing silence: How silence is produced in meetings. *Academy of Management Learning & Education*, 23: 304-324.

Tyler, S., 1978. *The said and the unsaid*. New York, NY: Academic Press.

Von Krogh, G., Roberson, Q., & Gruber, M. 2023. Recognizing and utilizing novel research opportunities with artificial intelligence. *Academy of Management Journal*, 66: 367-373.

Woo, S., Tay, L., & Oswald, F. 2024. Artificial intelligence, machine learning, and big data. *Personnel Psychology*, 77: 1387-1402.

Yoganarasimhan, H., & Iakovetskaia, I. 2024. From feeds to inboxes: A comparative study of polarization in Facebook and Email news sharing. *Management Science*, 70: 6461-6472.

Zhang, C., Yu, M. C., & Marin, S. 2021. Exploring public sentiment on enforced remote work during COVID-19. *Journal of Applied Psychology*, 106: 797.

Zhang, M., & Luo, L. 2023. Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Science*, 69: 25-50.