Tampereen yliopisto

Teemu Helenius

ORGANIZING KNOWLEDGE

Speaking the same language

Master's Thesis
Faculty of Management and Business
Examiner: Krista Sorri
Examiner: Osku Torro
July 2026

ABSTRACT

Teemu Helenius: Organizing Knowledge: Speaking the Same Language
Master's thesis
Tampere University
Master's Programme in Information and Knowledge Management
July 2026

Organizations coordinate action through shared vocabulary, but the meaning of that vocabulary is constituted through ongoing practice rather than fixed in written definitions. Large language models, deployed increasingly for organizational knowledge management, are trained on a fixed corpus and represent a synchronic snapshot of language at a point in time. This thesis investigates whether such models can participate in the maintenance of organizational meaning, and what limits their reliable use in practice.

A theoretical framework is developed, synthesizing three dimensions along which organizational meaning is constituted: a temporal dimension, describing whether a term's meaning is treated as fixed or continuously revised; a structural dimension, describing how abstract a term's representation is and how many senses it folds together; and a distributional dimension, describing how widely a term's governing code is shared beyond the group that produced it. Drawing on Peircean pragmatism, Saussure's synchronic/diachronic distinction, organizational semiotics, and computational linguistic evidence of semantic drift, the thesis argues that the gap between organizational and machine-represented meaning is a structural property of synchronic systems operating in diachronic environments, not a limitation correctable by better models or more careful prompting.

An empirical study, following a semiotic methodology adapted from Mingers and Willcocks' four-phase framework, examines an AI classification product tested against an enterprise knowledge base. Investigating the case revealed that no metadata category in the knowledge base carried an explicit definition, and that organizational vocabulary divides into a globally standardized layer, decodable by an outside agent, and a locally negotiated layer, decodable only through participation in organizational practice. Only the first layer could be brought into classification testing at all. Testing this layer against an LLM classifier showed that agreement with an independently constructed human baseline tracked how structurally open a category's definition was, and that adding explicit structural constraints reduced this variation directly, independent of whether a category's broader organizational meaning was otherwise resolved.

These findings are read as a direct empirical encounter with the theoretical boundary described above, rather than as a limitation of the research design: the categories that would have tested the gap most directly could not be evaluated from outside the organization at all. The thesis proposes that organizational knowledge management should pair categories with explicit statements of purpose and distributional status, and route locally negotiated categories away from automated classification by design.

Keywords: knowledge management, information systems, large language models, organizational semiotics, semiotics, semantics, artificial intelligence, classification

The originality of this thesis has been checked using the Turnitin Originality Check service.

TIIVISTELMÄ

Teemu Helenius: Tiedon organisointi: saman kielen puhuminen
Diplomityö
Tampereen yliopisto
Tietojohtamisen DI-ohjelma
Heinäkuu 2026

Organisaatiot koordinoivat toimintaansa jaetun sanaston avulla. Tämän sanaston merkitys rakentuu jatkuvassa käytännön toiminnassa, ei kirjallisissa määritelmissä. Suuret kielimallit, joita hyödynnetään yhä enemmän organisaation tiedonhallinnassa, on koulutettu aineistolla, joka edustaa kielen tilaa tietyssä ajankohdassa. Tässä opinnäytetyössä tutkitaan, voivatko tällaiset mallit osallistua organisaation merkitysten ylläpitoon, ja mitkä tekijät rajoittavat niiden luotettavaa käyttöä.

Työssä kehitetään teoreettinen viitekehys yhdistämään kolme ulottuvuutta, joiden kautta organisaation merkitys rakentuu. Ajallinen ulottuvuus kuvaa, kohdellaanko käsitteen merkitystä vakiintuneena vai jatkuvasti uudelleenarvioitavana. Rakenteellinen ulottuvuus kuvaa, kuinka abstrakti käsitteen esitystapa on, ja kuinka monta merkitystä käsite sisältää. Leviämisen ulottuvuus kuvaa, kuinka laajalti käsitettä ohjaava koodi on levinnyt sen tuottaneen ryhmän ulkopuolelle. Työssä hyödynnetään Peircen pragmatismia, Saussuren synkronis-diakronista erottelua, organisaatiosemiotiikkaa ja laskennallisen lingvistiikan empiiristä näyttöä semanttisesta muutoksesta. Työssä esitetään, että organisaation ja koneellisesti esitetyn merkityksen välinen kuilu on synkronisten järjestelmien rakenteellinen ominaisuus diakronisissa ympäristöissä, ei puute, jonka parempi malli tai huolellisempi kehotteiden laadinta voisi korjata.

Empiirinen tutkimus, joka noudattaa Mingersin ja Willcocksin nelivaiheisesta viitekehyksestä sovellettua semioottista metodologiaa, tarkastelee tekoälypohjaista luokittelutuotetta yrityksen tietopankissa. Tapauksen tarkastelu paljasti, ettei yksikään tietopankin metatietokategoria sisältänyt eksplisiittistä määritelmää. Organisaation sanasto jakautuu globaalisti standardoituun kerrokseen, jonka ulkopuolinen toimija pystyy tulkitsemaan, ja paikallisesti neuvoteltuun kerrokseen, joka on tulkittavissa vain organisaation sisäiseen toimintaan osallistumalla. Vain globaalisti standardoitu kerros voitiin ottaa mukaan luokittelutestaukseen. Tämän kerroksen testaaminen kielimallipohjaisella luokittelijalla osoitti, että kielimallin tuotoksen ja ihmisen tekemän arvion välinen yksimielisyys vaihteli sen mukaan, kuinka rakenteellisesti avoin kategorian määritelmä oli. Lisäksi eksplisiittisten rajoitteiden lisääminen vähensi hajontaa suoraan. Tämä ei kuitenkaan tarkoita, että kategorian semanttinen merkitys olisi ratkaistu organisaation näkökulmasta.

Näitä löydöksiä tulkitaan empiirisenä kohtaamisena teoreettisen kuilun kanssa, ei tutkimusasetelman rajoitteena: kategoriat, jotka olisivat testanneet kuilua suorimmin, eivät olleet lainkaan arvioitavissa organisaation ulkopuolelta. Työ esittää, että organisaation tiedonhallinnassa kategoriat tulisi yhdistää eksplisiittisiin tarkoituksen ja leviämistilanteen kuvauksiin, ja ohjata paikallisesti neuvotellut kategoriat pois automaattisesta luokittelusta jo suunnitteluvaiheessa.

Avainsanat: tiedonhallinta, tietojärjestelmät, suuret kielimallit, organisaatiosemiotiikka, semiotiikka, semantiikka, tekoäly, luokittelu

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

USE OF AI IN THESIS

I have utilized AI tools in my thesis:

☐ No
☒ Yes

According to my declaration, I have used the following artificial intelligence applications during the thesis process:

Names and versions of AI applications: Anthropic Claude Sonnet 5, Google Gemini 3.1 Pro, and the case company's proprietary AI classification tool.

Purpose of use: Claude Sonnet 5 and Gemini 3.1 Pro were used to support learning and the refinement of written text. Their uses included: helping identify the research gap and shape the initial research idea, which was subsequently checked against existing literature to confirm it was genuinely unaddressed; explaining theoretical concepts drawn from the different fields the thesis engages with, often comparing the two models' explanations against each other, with all explanations verified against the primary/original texts before being relied on; suggesting and locating literature sources during Chapter 1-2 research, with all suggested sources checked against the original texts before citation; aiding comprehension of English-language source material, since the researcher's first language is Finnish; brainstorming structure and arguments; and editing and proofreading written text. The case company's proprietary AI tool served a different function: it is the empirical object of study itself, the LLM-based document classification product whose outputs are tested against a human baseline, as described in the methodology. No customer document content was input into Claude or Gemini at any stage; all work involving customer data was conducted exclusively through the case company's proprietary tool, consistent with company data policy. All empirical data curation and analysis, including grouping classification outputs and assessing agreement against the human baseline, was performed manually by the researcher.

Sections in which AI has been used: Claude Sonnet 5 and Gemini 3.1 Pro were used across Chapters 1-6 for the purposes above. The case company's proprietary tool was used specifically within the empirical study in Chapters 3 and 4, where its classification outputs constitute the primary data collected and analyzed.

I am aware that I am fully responsible for the entire content of my thesis, including the parts in which AI has been utilized, and I accept responsibility for any possible violations of ethical guidelines.

PREFACE

It was not obvious I would in fact do a master's degree and complete a thesis. My friends were taking bets on when I'd go back to university after two years of pursuing my personal interests in form of startups since completing the bachelor's degree in information technology. Going back to university was primarily motivated to find like-minded people, ultimately co-founders, and that became a year and a half long intense journey as the chairperson of Tampere Entrepreneurship Society (TRES). That period did not progress my degree that much, but it did grow me more as a person and most importantly it became a community of like-minded people and a spark of culture that will do great things.

After TRES, the same drive turned toward the thesis. It became an obsession project overwhelmed by the number of dimensions available primarily in organization, sociologic, linguistic, semiotic, artificial intelligence, and philosophy studies. Unmanageable, but somehow, I found a way to connect a lot of dots and create a path through different worlds and find a path out. Sort of world building, or modeling. While inspired by great scientists like Shannon, Peirce, or by the talks like Hamming's "You and your research", or by Karpathy's blog post on "A Survival Guide to a PhD" I must admit, I still do not know what great research, or rather a master's thesis, looks like today compared to Shannon's little under 90 years ago, in the intersection of the organizational field of theories. Regardless, the lack of examples often means doing something new.

I'm thankful for more people than I can name in this short preface, from the case organization to my friends, colleagues, but most importantly Krista for her guidance and my wife Ilona for her support and love.

Lempäälä, 31.7.2026

Teemu Helenius

TABLE OF CONTENTS

1. INTRODUCTION 1
    1.1 Outline 3
2. THEORETICAL BACKGROUND 5
    2.1 Language as a Coordination Mechanism 5
    2.2 Semiotic meaning theory 7
    2.3 Shannon's Information theory and LLM background 8
    2.4 Three dimensions of organizational meaning 10
    2.5 The Organizational Semantic Gap 12
3. METHODOLOGY 15
    3.1 Methodological framework 15
    3.2 Research design 18
    3.3 Case Context 19
    3.4 Data collection 20
    3.5 Data analysis 23
4. ANALYSIS 26
    4.1 Investigating the three worlds 26
    4.2 Interaction of the worlds: the local/global divide 28
    4.3 Generating the hypothesis 30
    4.4 Testing the hypotheses: baseline validation 31
    4.5 Summary of findings 34
5. DISCUSSION 37
    5.1 Assessing the hypotheses against literature 37
    5.2 The semantic access barrier as the phenomenon 40
    5.3 Sketching an idealized design 42
6. CONCLUSIONS 46
    6.1 Central Thesis 46
    6.2 Theoretical and Empirical Contribution 47
    6.3 Practical Proposals 48
    6.4 Limitations 49
    6.5 Future Research 50
REFERENCES 52
APPENDIX A 56
APPENDIX B 57

LIST OF TABLES

Table 1: 4As framework adapted (Mingers & Willcocks, 2017)..... 18
Table 2: 12 steps adapted (Mingers & Willcocks, 2017)..... 24
Table 3: Material-social interactions examined in this section..... 28
Table 4: Test Run 1 results (twenty documents)..... 32
Table 5: Test Run 2 results (seventeen documents)..... 33
Table 6: Summary of empirical findings..... 35
Table 7: Idealized counterfactuals for each identified mechanism..... 43
Table 8: From counterfactual design to practical proposals..... 49

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
LLM	Large Language Model
RQ	Research Question
RAG	Retrieval-augmented generation

1. INTRODUCTION

Organizations depend on shared language to coordinate action. When a team agrees on what "done" means, or what "urgent" means, they are not assigning labels to objective truths but arriving at negotiated, locally shared agreements, tied to a specific real-world context, that become the common ground making joint action possible. This is fundamentally a coordination problem: the knowledge required for good decisions is distributed across people who cannot fully articulate or transmit it, and shared vocabulary is what makes acting together possible despite this (Hayek, 1945). Coordination fails not primarily when people lack information, but when they mean different things by the same words. A project that one team considers complete and another considers unfinished does not suffer from an information deficit. It suffers from a semantic problem: the two teams are not speaking the same language.

This problem is as old as organizations, and managing it has always required active effort. Shared vocabularies must be built, maintained, and renegotiated as circumstances change through practice (Orlikowski, 2002; Wenger, 1998). What has changed in recent years is the scale and speed at which organizational knowledge is produced, and the tools organizations reach for to manage it. Organizations across industries face a rapidly growing volume of unstructured information, such as internal documents and communications, that traditional keyword search and manual categorization struggle to keep pace with (Karakurt & Akbulut, 2026). Artificial intelligence (AI) and large language models (LLM) have entered knowledge management as a response to this problem of scale: research on applying LLMs to enterprise knowledge management and document automation has grown rapidly since 2020, though most enterprise deployments remain experimental (Karakurt & Akbulut, 2026).

This promise is not unfounded. Large language models are genuinely capable of processing language at organizational scale, producing outputs that are fluent and statistically consistent with human language (Brown et al., 2020; OpenAI et al., 2023; Singh et al., 2026). But processing language and participating in the meaning of language are not the same thing (Searle, 1969, 1980, 2009). An organization's vocabulary is not a stable, evenly distributed set of definitions waiting to be applied. It is a living system of concepts whose meaning is constituted through use (Wittgenstein, 1953) and consequence, re-

vised on an ongoing basis as circumstances change (Peirce, 1877, 1878). When an organization deploys a large language model to manage its knowledge, it implicitly assumes the model interprets its terms the way the organization currently intends them. When the model reads "risk" or "customer" or "priority," it draws on statistical regularities in how words co-occur in general text, rather than from the specific agreements the organization has negotiated (Harris, 1954; Sahlgren, 2008).This assumption, and what it costs an organization when it goes unexamined, is the problem this thesis addresses.

The central claim is that there is a semantic gap between how organizations produce and maintain meaning and how large language models represent it. This gap can be described along three dimensions: temporal, whether a term's meaning is treated as fixed or as continuously revised; structural, how abstract a term's representation is and how many senses it folds together; and distributional, how widely a term's governing code is shared beyond the group that produced it. Across all three, the gap is a structural property of synchronic systems deployed in diachronic environments (Saussure, 1916/2011), not a technical limitation that better models or more careful prompting can close. LLMs are trained on a fixed corpus and capture a statistical snapshot of how language was used at the time of training (Brown et al., 2020; Radford et al., 2018). Organizations update the meaning of their vocabulary continuously, through a consequence-driven feedback loop of action, failure, doubt, negotiation, and revision (Peirce, 1877, 1878). A large language model cannot participate in this feedback loop for two related reasons: it has no mechanism for updating its representation of a term once training ends (Brown et al., 2020; Radford et al., 2018), and participating in the loop itself, acting, observing consequences, and revising accordingly, requires an agent capable of intentional action, which a transformer-based model, by its structure, lacks (Gubelmann, 2024).

The empirical part of the thesis observes this gap through an unexpected route, following a semiotic methodology adapted from Mingers and Willcocks' four-phase framework for information systems research (Mingers & Willcocks, 2017). Investigating a real enterprise knowledge base surfaced a first, preliminary observation: none of its metadata categories carried an explicit definition anywhere in the system, only a name and a data format. Examining why revealed a sharper pattern. Organizational vocabulary divided into two kinds: categories anchored in a code shared widely enough, beyond the organization, that an outside researcher could construct a working definition for them, and categories anchored in a code negotiated entirely inside the organization's own practice, with no accessible external equivalent. Only the first kind could be brought into classification testing at all; the second could not be evaluated from outside and could not be tested, a boundary visible in the category data itself before any classification took place.

Testing the first kind against an LLM classifier then showed a further pattern: agreement between the model's output and a human baseline was high wherever a category's correct answer took a fixed structural form, and dropped, in varying degrees, wherever a definition left room for more than one plausible reading, a variation that narrowed again once the definition itself was made more structurally explicit. Read together, these observations are treated in this thesis as a direct empirical instance of the theoretical gap described above, rather than as a limitation of the research design.

The thesis is organized around one main research question (RQ): can large language models participate in the maintenance of organizational meaning, and what limits their reliable use in organizational knowledge management? Three supporting questions structure the investigation of this question across the theoretical and empirical parts of the thesis.

RQ1: How do humans and large language models differ in the way they organize and maintain meaning?

RQ2: What challenges arise when large language models are applied to organizational knowledge tasks in practice?

RQ3: What are the implications of these challenges for the division of labor between humans and AI in organizational knowledge management?

These questions will guide the investigation of organizational meaning, outlined next.

1.1 Outline

The thesis proceeds in six chapters. Chapter 2 develops a theory of organizational meaning as constituted through consequence-driven practice, synthesizing this into three dimensions along which meaning can be described: temporal, structural, and distributional. It establishes that large language models are structurally excluded from the feedback loop through which organizational meaning is maintained, and examines why this exclusion tends to stay hidden rather than announcing itself.

Chapter 3 sets out the methodology: a semiotic approach adapted from Mingers & Willcocks (2017) four-phase framework, applied to a real enterprise knowledge base rather than a constructed test set. It establishes the case context, the criteria used to select document classes and metadata categories for testing, and the protocol for investigating the three worlds and validating results against a human baseline, together with the methodological justification for adapting the framework's original producer-consumer model to the three dimensions developed in Chapter 2.

Chapter 4 executes this protocol. Investigating the three worlds surfaces two preliminary findings: no category in the knowledge base carried an explicit definition (Finding 1), and organizational vocabulary divides into a globally standardized layer, decodable by an outside agent, and a locally negotiated layer, decodable only through organizational practice (Finding 2). Only the first layer could be brought into classification testing. Testing this layer against an LLM classifier under two successive rounds of definitional specificity shows that agreement with a human baseline tracks how structurally open a category's definition is (Finding 3), and that constraining that structure directly reduces variation, independent of whether the category's broader organizational meaning is otherwise resolved (Finding 4).

Chapter 5 assesses these findings against the theoretical framework of Chapter 2, addressing RQ1 and RQ2 directly, and argues that the study's design encountered the semantic gap as a barrier to access rather than as an observable failure in production, a limitation treated here as itself a form of evidence for the theoretical claim. It closes by sketching an idealized counterfactual design in which the mechanisms behind each finding would not arise, beginning the answer to RQ3.

Chapter 6 concludes by stating the thesis's theoretical and empirical contribution, converting Chapter 5's counterfactuals into a concrete practical proposal for pairing organizational categories with explicit statements of purpose and distributional status, completing the answer to RQ3, and closes with the study's limitations and the further empirical work needed to extend it.

2. THEORETICAL BACKGROUND

This chapter examines the theoretical foundations of organizational language and its intersection with artificial intelligence. Section 2.1 begins by reframing language from a static storage medium into a dynamic coordination mechanism sustained by social practice and speech acts. Section 2.2 then introduces semiotic meaning theory, utilizing Peirce's pragmatic maxim to explain how meaning is continuously negotiated through action and consequence. Following this, Section 2.3 explores Shannon's information theory and the underlying mechanics of Large Language Models, contrasting the synchronic, distributional architecture of LLMs with the diachronic evolution of human vocabulary. Section 2.4 synthesizes these theoretical strands to evaluate organizational meaning across three distinct dimensions: temporal, structural, and distributional. Finally, Section 2.5 concludes by defining the organizational semantic gap, examining the practical consequences and hidden risks that emerge when AI systems operate without access to the shared practices that ground these three dimensions of meaning.

2.1 Language as a Coordination Mechanism

Knowledge management literature has predominantly treated language as a storage and retrieval medium: a way to capture what people know so it can be shared and reused (Alavi & Leidner, 2001; Nonaka & Takeuchi, 1995). This view motivates taxonomies, ontologies, and document repositories. The underlying assumption is that knowledge exists in people's heads and language is the vessel that moves it between them, and what is stored arrives at its destination intact, meaning unchanged. Designing good knowledge management is then a question of designing good vessels. This view is not wrong about what language can do. It is wrong about what language primarily does in organizations.

The more fundamental problem language solves in organizations is coordination under conditions of distributed knowledge. Hayek's (1945) argument about economic knowledge applies directly: the knowledge required for good decisions is not held by any single person or system (Hayek, 1945). It is distributed across people who do not fully know what they know, and who cannot fully articulate it even when they do. Polanyi captured the boundary of this problem precisely: we know more than we can tell (Polanyi, 1967). Tacit knowledge, local expertise, and situation-specific judgment resist capture not because documentation tools are inadequate but because much of what people know is inseparable from the practice in which it was acquired (Tsoukas, 1996). No central

system can aggregate the knowledge of an entire organization without distorting it in the process (Tsoukas, 1996). Organizations are, in Tsoukas's terms, distributed knowledge systems in a strong sense, decentered and lacking any single overseeing mind capable of holding what the organization collectively knows (Tsoukas, 1996). This is not a temporary problem to be solved by better documentation tools. It is a structural feature of how knowledge exists in human activity.

If distributed knowledge cannot be fully aggregated, the question becomes how organizations act coherently despite it. The answer is shared language. A shared term functions as what Star calls a boundary object: flexible enough to be adapted to the local needs of different groups, while robust enough to preserve a common identity as it moves between them (Star, 1989). When people share a vocabulary, they can act together without needing to fully explain their reasoning each time. A shared term carries compressed, pre-negotiated meaning that allows people to coordinate across contexts, roles, and time without reopening every assumption. Shared language does not eliminate the distributed character of knowledge. It makes coordination possible despite it, by giving people enough common ground to act on without requiring complete mutual transparency. (Star, 1989)

What this means for how language functions in organizations becomes clearer through speech act theory. Austin and Searle established that language is not only constative, describing states of affairs, but performative: it does things in the world (Austin, 1962; Searle, 1969). Promises, commitments, directives, and declarations are speech acts that create organizational reality rather than merely describing it. For example, when a manager says "this is a priority," they are not reporting a pre-existing fact about the world. They are allocating attention, setting expectations, and constituting accountability. The statement changes what people do. This performative dimension is the most organizationally consequential aspect of language, and it is the dimension most absent from how machine language processing works (Gubelmann, 2024). A language model can reproduce the surface form of such statements, but it cannot perform the act itself: genuine speech acts require an agent with intentions, and a language model, whatever its fluency, is the wrong kind of thing to have them (Gubelmann, 2024). It has no part in the organizational reality such statements create.

The implication for this thesis follows directly. If language in organizations is primarily a coordination mechanism whose function is performed through use and consequence rather than storage and retrieval, then the question for knowledge management is not "how do we capture what people know" but "how do we maintain the shared vocabulary that makes coordination possible." These are different problems. The first is a representation

problem, and it is the problem that most knowledge management tools, including AI tools, are designed to address. The second is a social and semantic maintenance problem, one that requires participation in the ongoing practice through which meaning is tested, updated, and stabilized through action and its consequences..

This second problem, maintaining rather than merely storing organizational vocabulary, has its own literature within knowledge management and information systems research, and situating the present argument within it clarifies what this thesis adds. Wenger's account of communities of practice describes how a shared repertoire of concepts and vocabulary emerges from sustained joint engagement in an activity, and is renegotiated as the community's practice evolves rather than fixed by any single act of definition (Wenger, 1998). A term used within a community of practice carries meaning because members have jointly worked out what it picks out through repeated situations, not because a dictionary entry was consulted (Wenger, 1998). Organizational semiotics, the tradition within information systems research gives this idea a more precise vocabulary: a norm is a generalized disposition shared by members of a community, and organized behavior is norm-governed behavior in the sense that signs acquire their organizational meaning through the norms a community holds about how to act on them (Stamper et al., 2000). Both accounts converge on the point this thesis has been building toward from a different direction: organizational vocabulary is sustained by an ongoing social practice, not stored as a static artifact, and an agent outside that practice, human or machine, has no direct route to the norms that give a locally negotiated term its meaning.

2.2 Semiotic meaning theory

If meaning is maintained through practice rather than stored in representations, it requires a theory of what meaning is. The dominant frameworks in information science and language technology rest on assumptions that are worth making explicit before accepting.

Peirce's pragmatic maxim provides the foundation this thesis builds on. Peirce proposed that the meaning of a concept is the sum of its conceivable practical effects: what difference it makes in practice to apply one concept rather than another (Peirce, 1878). This is not a philosophical abstraction but a working definition with direct organizational implications. To know what a word means is to know what people do differently because of it (Peirce, 1878). A concept whose application makes no practical difference has no meaning in the relevant sense (Peirce, 1878). A concept that changes how people act, allocate resources, or assign responsibility means precisely those changes. Organization studies

arrive at a related conclusion from empirical rather than philosophical grounds: Orlikowski's study of distributed product development found that organizational knowing is not a stable, storable disposition but an ongoing accomplishment, continuously reconstituted as people act, and inseparable from the practice in which it occurs (Orlikowski, 2002).

Peirce's triadic model of the sign develops this further (Chandler, 2021). The three elements, representamen, object, and interpretant, are less important individually than in their relationship. The interpretant is the effect the sign produces: not a mental image but a disposition to act. Meaning is not a fixed property of a word. It is what the word makes people do, and it changes as the consequences of acting on the word change. For organizational purposes, this means that the meaning of a term like "customer" or "risk" is constituted by the organizational practices the term activates. When the practices change, the meaning changes with them, whether or not anyone updates a definition document.

Peirce described this process as semiosis: a continuous chain of interpretation in which each sign generates a further sign (Atkin, 2023). Meaning is never finished. It is an ongoing process that updates through use and consequence. The same word can mean something slightly different in each context without losing recognizability, because both parties carry enough shared history to bridge the gap. This is not a defect of natural language. It is exactly the feature that makes language useful as a coordination tool across varying contexts, roles, and situations.

2.3 Shannon's Information theory and LLM background

The contrast with Shannon's theory of communication sharpens what is at stake. Shannon defined information as a measure of uncertainty reduction across a set of possible states (Shannon, 1948). The definition is precise and powerful, but it was explicitly scoped to the syntactic level: the accurate transmission of symbols regardless of what they mean. Shannon and Weaver acknowledged the semantic and pragmatic levels of communication but stated that they fell outside the engineering problem they were solving (Shannon & Weaver, 1949). This was a reasonable boundary for signal transmission. The difficulty arises when the same paradigm is extended to the management of organizational meaning, where the semantic and pragmatic levels are precisely what matters.

Large language models inherit this problem through the distributional hypothesis. Harris proposed that words occurring in similar linguistic contexts have similar meanings, and that meaning can therefore be modeled from the statistical distribution of language (Harris, 1954). This hypothesis became the foundation of modern language modeling. LLMs

extend it at scale: trained on large corpora, they learn to predict which words follow which other words, producing outputs that are statistically consistent with human language (Brown et al., 2020; OpenAI et al., 2023; Radford et al., 2018; Singh et al., 2026). The results are fluent and often useful. But distributional models are not models of meaning in the Peircean sense. They are models of meanings in text (Sahlgren, 2008).This limitation compounds the one Polanyi identified in the previous section. Distributional models can only learn from language that was written down. Expressed knowledge is already a subset of what people know (Polanyi, 1967), and publicly available text is a further subset of what organizations ever commit to writing. The distributional model learns from the residue of the residue of organizational knowledge.

Boisot's concept of codification names a related idea from an organizational rather than a linguistic starting point (Boisot, 1998). Boisot models knowledge along an axis running from uncodified, meaning embedded in concrete, particular circumstances, to codified, meaning abstracted into an explicit, general form that can be applied across situations (Boisot, 1998). A distributional model of language performs a related operation on words: it represents a word's meaning as its position relative to every other word it co-occurs with, itself a highly abstracted structure built by discarding the particulars of any single utterance in favor of a pattern that holds across a whole corpus (Harris, 1954). Distributional language modeling and Boisot's codification axis describe the same basic move applied to different objects: both trade the particular for the general to make a pattern usable outside the situation that produced it.

The synchronic and diachronic distinction introduced by Saussure names the structural dimension of this problem (Saussure, 1916/2011). Synchronic analysis describes language as it exists at a moment in time. Diachronic analysis describes how language changes through use over time. Saussure's own preference was explicitly for synchronic analysis. This preference is the direct intellectual lineage of why distributional language modeling is synchronic by design: the paradigm inherits Saussure's framing of language as a system to be analyzed at a moment in time. For organizations, however, the diachronic dimension is not merely temporal. It is the consequence feedback loop through which meaning evolves. People act on words, observe what happens, and update their understanding accordingly. New projects expose ambiguities in existing terms. Strategic shifts change what words are expected to pick out. Coordination failures reveal where the vocabulary was imprecise. The organization learns through these episodes, and its vocabulary moves with that learning.

This is not merely a theoretical possibility. Computational linguistics has independently documented that word meanings measurably drift over historical time, and that the drift

follows identifiable patterns. Analyzing word embeddings trained on separate time slices of large historical corpora, Hamilton, Leskovec, and Jurafsky (2016) show that words shift in meaning at rates that vary systematically with frequency and polysemy, with rarer and more multi-sense words drifting faster (Hamilton et al., 2016). Their study treats semantic change as an empirical, measurable phenomenon in exactly the sense Saussure's diachronic axis describes, and confirms that a model trained on language from one period misrepresents the meaning words carry in a later period, in proportion to how much drift has occurred between the two. What Hamilton et al. (2016) document at the scale of general language over decades, this thesis argues happens at the scale of organizational vocabulary over far shorter cycles: an LLM's training cutoff functions as exactly the kind of fixed time slice their methodology depends on, except that the organizationally relevant drift it misses can occur in months rather than decades, wherever a term's meaning is actively being renegotiated in practice.

Sahlgren explicitly points out the strength, and the weakness of distributional language models; "By grounding the representations in actual usage data, distributional approaches only represent what is really there in the current universe of discourse. When the data changes, the distributional model changes accordingly; if we use an entirely different set of data, we will end up with an entirely different distributional model." (Sahlgren, 2008). LLMs are trained on a fixed corpus and represent a synchronic snapshot of language at the time of training (Brown et al., 2020; Radford et al., 2018). They do not participate in the consequence-driven feedback loop between training runs. The gap between a synchronically frozen model and a diachronically evolving organization does not stay constant. It widens in proportion to how much the organization has learned through experience since the model last encountered its language.

2.4 Three dimensions of organizational meaning

The theoretical strands developed above are easier to hold together as three dimensions along which the meaning of a term can be described, rather than as a list of separate authors and arguments. Each dimension has its own literature and its own way of showing why a synchronic system runs into difficulty in a diachronic environment.

The first is a temporal dimension: whether a term's meaning is treated as fixed at a moment or continuously revised through use. Saussure's synchronic and diachronic axes name the two poles directly (Saussure, 1916/2011), and Peirce's account of semiosis names the mechanism that moves a term from one state to the other: each act of interpretation produces a further sign, so meaning is continuously regenerated through the chain of interpretants a community produces rather than settled once (Atkin, 2023). This

is not only a philosophical claim. Hamilton, Leskovec, and Jurafsky (2016) show empirically that word meanings drift measurably over historical time, at a rate that depends on how frequently a word is used and how many senses it carries (Hamilton et al., 2016). A simple example makes this concrete: "cloud" carried a single meteorological sense for most of its history and now carries a second, unrelated sense referring to remote computing infrastructure. A snapshot taken before cloud computing existed and one taken after would represent two different meanings, and neither is wrong on its own terms. It is simply frozen at a different point on the temporal axis.

The second is a structural dimension: how abstract or concrete a term's representation is, and how many distinct senses it has folded into one form. Boisot's codification axis names the organizational side of this, from uncodified knowledge tied to a specific situation to codified knowledge abstracted into a general form (Boisot, 1998). Harris's distributional hypothesis names the linguistic side of the same move (Harris, 1954), and the connectionist architectures behind large language models inherit this commitment directly (Brown et al., 2020; Radford et al., 2018): meaning is represented as a position in a relational structure rather than as a symbol tied to one referent, which is why a model can be fluent about a term's typical usage while carrying nothing about the specific situation an organization needs addressed on a given day. Hamilton et al.'s finding that polysemous words drift faster sits on this same dimension (Hamilton et al., 2016), since polysemy is a direct measure of how much a form has abstracted, folding several situation-specific meanings into one shape. "Priority" works the same way in an organization: its abstract sense (something that matters more than other things) is stable, but its concrete sense (which of this week's forty open tasks counts as one) is exactly what a general representation cannot carry.

The third is a distributional dimension: how widely a term's meaning is shared beyond the group that produced it. Boisot's diffusion axis names this directly (Boisot, 1998). Organizational semiotics gives it a social mechanism: a norm is a generalized disposition shared by members of a community, and it is the norm, not the word alone, that fixes what a sign means for that community (Stamper et al., 2000). Wenger's (1998) communities of practice describes how such norms are produced and sustained through repeated joint engagement rather than through any single act of definition (Wenger, 1998), and Mingers and Willcocks (2017) name the shared result of this process the code: the system of meanings a producer and a consumer must hold in common for communication to succeed (Mingers & Willcocks, 2017). "Document date" sits at the diffused end of this dimension, since its code is the Gregorian calendar, shared far beyond any one

organization. An internal project phase label sits at the undiffused end, its code held only by the group whose practice produced it.

These three dimensions are not independent. The Boisot and Hamilton et al. connections already show the structural dimension interacting with the temporal one: more abstract, more polysemous terms are also the ones whose meaning moves fastest through time. But the dimensions are distinguishable enough to do separate analytical work, and together they specify what a synchronic model deployed against organizational vocabulary cannot track. It cannot follow a term as its meaning moves through time. It cannot distinguish an abstract sense from the concrete one an organization needs on a given day. And it has no access to the socially bounded code that gives an undiffused term its meaning in the first place. A model can be fluent in the atemporal snapshot sense of a word while missing all three dimensions at once, and because fluency is what a reader notices, the absence of the other two tends to go unnoticed too.

2.5 The Organizational Semantic Gap

The previous sections established the structural claim: LLMs are synchronic systems that cannot participate in the diachronic process through which organizational meaning evolves. This section examines what that claim looks like in organizational practice — how the gap manifests, why it tends to stay hidden, and why the risk of its remaining invisible is unevenly distributed.

Wittgenstein's rule-following argument is the appropriate starting point, because it makes the practical problem precise (Wittgenstein, 1953). Wittgenstein observed that no finite set of examples can uniquely determine a rule. For any rule and any set of past applications, there is always an alternative interpretation consistent with all prior cases while diverging on the next one (Kripke, 1982). This is not merely a philosophical puzzle. It describes exactly what happens when an LLM applies an organizational concept in a new context: its interpretation may be fully consistent with all prior usage in its training data while diverging from what the organization would currently intend. The model has no access to the cases that came after training, nor to the coordination failures that would have revealed where the concept needed revision. From the model's perspective, there is no divergence. The rule is being followed. It is simply being followed according to a different learned history.

This divergence does not announce itself. In organizational practice, the semantic gap presents not as obvious error but as a situation where a human and an LLM each operate with a different internalized meaning of the same term, without a shared reference point

from which to detect the conflict. The difficulty is compounded by the fact that correctness in organizational language is not established by inspecting a term in isolation (Orlikowski, 2002). It is established through collective action and consequence. A word means what the organization has learned it to mean through using it and observing what happened (Orlikowski, 2002). Neither the human nor the model can resolve a semantic disagreement simply by consulting the word itself, because the word's meaning is not in the word. This is the Chinese room problem reframed (Searle, 1980, 2009): not a single agent simulating understanding, but two agents each proceeding with confidence, operating with invisibly different semantic models and no common ground from which to surface the divergence.

This pattern is a specific instance of a more general finding in human-automation interaction research. Parasuraman and Manzey's review of automation bias and complacency concludes that people systematically under-monitor automated systems and over-rely on their output, and that this tendency occurs in both naive and expert operators and is not reliably corrected by training alone (Parasuraman & Manzey, 2010). Whether the gap surfaces or stays hidden is not a fixed property of the situation. It depends substantially on the epistemic confidence of the human involved (Logg et al., 2019), though organizational context and confidence appear to reduce rather than eliminate the underlying tendency Parasuraman and Manzey describe. A person with deep organizational context and confidence in their own understanding will notice when the model's output diverges from organizational intent. The friction becomes explicit, and the gap is identified. A person with less organizational context, or less confidence in their own judgment relative to a fluent and authoritative-seeming machine, may defer to the model's output without recognizing the divergence (Allen & Choudhury, 2022). The gap stays hidden, and the model's semantic interpretation is quietly adopted as correct. The practical consequence of this asymmetry is significant. Organizations are not populated uniformly by people with deep organizational context and knowledge. Junior staff, recent hires, people crossing into unfamiliar domains, and anyone who encounters an AI-produced output without the background to evaluate it is exposed to the semantic gap in its most dangerous form (Kellogg et al., 2025): one that reads as plausible and goes unchallenged.

The compounding dynamic is worth naming explicitly. When a person with insufficient context accepts a model's output, that output may enter organizational circulation (Crossan et al., 1999): it becomes a document, a decision input, a classification that others rely on. The model's semantic interpretation gets embedded in the organization's knowledge base (Walsh & Ungson, 1991). Later users encounter it as part of the record

and have no way of knowing it diverged from organizational intent at the point of production. The gap is not just hidden in any single instance. Over time, it can silently become the organization's official meaning.

This is the organizational reality of the semantic gap. It is dynamic rather than static, socially distributed rather than uniformly present, and it feeds back into the language environment it operates in. Managing it requires more than improving the distributional model quality. It requires organizational attention to where the gap is most likely to stay hidden, and deliberate design of the human role in knowledge work with this risk in view.

3. METHODOLOGY

This chapter details the methodology used for the empirical investigation. Section 3.1 establishes the methodological framework, adapting Mingers and Willcocks' four-phase semiotic approach (Appreciate, Analyse, Assess, Act) to integrate Peirce's triadic model of the sign with the temporal, structural, and distributional dimensions of meaning. Section 3.2 describes the qualitative, observational research design, which utilizes a single embedded case study rather than a constructed test set. Section 3.3 introduces the case context, outlining the AI classification product and the enterprise customer's document repository used as the testing environment. Section 3.4 outlines the data collection process, detailing the selection and filtering of document classes alongside the construction of category definitions. Finally, Section 3.5 defines the data analysis protocol, explaining how the classification outputs are evaluated against an independently constructed human baseline.

3.1 Methodological framework

The empirical study follows a structured semiotic methodology adapted from Mingers and Willcocks' four-phase framework for information systems research: Appreciate the research situation, Analyse the research material using semiotics, Assess the validity and plausibility of the potential explanatory mechanisms, and Act to bring about change if necessary (Mingers & Willcocks, 2017). The framework offers a disciplined progression from problem identification through interpretation to actionable proposals, and it shares a foundation with the theoretical framework developed in Chapter 2: both draw on Peirce's triadic model of the sign, in which meaning is constituted through the relation between representamen, object and interpretant rather than fixed within the sign itself (Mingers & Willcocks, 2017).

The adaptation departs from Mingers and Willcocks' original application in one respect. Their framework organizes analysis around three worlds, personal, social/cultural and material, connected through producer and consumer roles engaged in the transmission and interpretation of a shared code (Mingers & Willcocks, 2017). This thesis reinterprets two of those worlds through two of the dimensions developed in section 2.4, and treats the third dimension, the temporal, as a lens applied across the framework rather than as a third world-to-dimension substitution.

The social/cultural world corresponds to the distributional dimension: the norms a community holds about a sign, and how widely those norms extend beyond the group that produced them, following Stamper's account of organizational semiotics and Wenger's account of communities of practice (Stamper et al., 2000; Wenger, 1998).

The material world corresponds to the structural dimension: the artifacts through which meaning is technologically mediated, including the documents, the classification categories as name-datatype pairs, and the language model itself as a structurally abstracted, synchronic snapshot of prior usage.

The temporal dimension is not substituted for the personal world in the same way. Mingers and Willcocks's framework does not exclude time: sociation, embodiment, and sociomateriality are all, in principle, processes that unfold over time, and nothing in their account rules out asking how a producer's or consumer's relation to a code changes historically. But time is not made an explicit, operative axis in their model; it is present in the ideas without being foregrounded as a dimension of analysis. This thesis foregrounds it, and does so specifically as a property of use between people rather than of private, individual interpretation. This choice follows from the theoretical position already established in section 2.5: Wittgenstein's rule-following argument denies that meaning can be secured by an individual's private, internal act of interpretation in the first place, since a private rule is as underdetermined as a public one (Kripke, 1982; Wittgenstein, 1953). Meaning is constituted publicly, through what a community does with a term and revises about it over time, not privately, through what any one person takes it to mean at a given moment (Kripke, 1982; Wittgenstein, 1953). The temporal dimension, in this sense, operates through the social and material worlds, in how norms shift through use and how codified artifacts fall out of step with that shift, rather than through a separate personal world of subjective change.

The code, in Mingers and Willcocks' terms the shared system of meanings that producer and consumer must hold in common for communication to succeed, is treated here as internally divided along the distributional dimension rather than uniform: part of it is anchored in conventions shared globally, well beyond the organization, and part of it is anchored in conventions negotiated locally, inside organizational practice. This division becomes the operative distinction in the empirical findings (Chapter 4, Finding 2).

The adaptation described above requires justification beyond disclosure, since departing from an established framework is a substantive methodological decision rather than a stylistic one. Three considerations ground it.

First, the departure is targeted rather than wholesale. What is retained from Mingers and Willcocks (2017) is the four-phase progression (Appreciate, Analyse, Assess, Act) and its underlying commitment to Peirce's triadic model of the sign, in which meaning arises from the relation between representamen, object, and interpretant rather than residing in any one of them (Mingers & Willcocks, 2017). What is altered is a single interpretive layer: the three worlds and the producer-consumer relationship through which Mingers and Willcocks operationalize them. Their own application of this layer is built around sociation, embodiment, and sociomateriality, categories developed for technologically-mediated communication in general and illustrated in their paper through cases such as social media and business process outsourcing (Mingers & Willcocks, 2017). These categories carry the structural and distributional dimensions this thesis needs, but they do not foreground the temporal dimension as an explicit, operative axis. They were not designed to distinguish a code that is static because it was learned once from a code that is live because it is still being negotiated, and that distinction is exactly what this thesis's research questions require, since the central theoretical claim rests on the difference between synchronic and diachronic systems of meaning (Saussure, 1916/2011). Reinterpreting the three worlds along that axis is the minimum change needed to make the framework answer a question it was not originally built to ask, not an arbitrary substitution.

Second, this kind of targeted reinterpretation is a recognized move in interpretive information systems research rather than an idiosyncratic one. Klein and Myers's (1999) principle of contextualization requires researchers to critically reflect on how the social and historical setting of a phenomenon bears on both the phenomenon and the research process examining it (Klein & Myers, 1999). The deployment of large language models against organizational vocabulary did not exist in its current form when Mingers and Willcocks published their methodology, and there is no reason to assume categories built for a broader class of ICT-mediated communication transfer unchanged to a case defined specifically by the tension between a frozen training corpus and a living organizational practice. Adapting the worlds to carry that distinction is what the contextualization principle asks a researcher to do, not a departure from it.

Third, the adaptation is consistent with the methodological stance of Mingers's own work. Mingers (2001) argues against treating a single framework as fixed and complete, proposing instead that research is strengthened when methods are combined and adjusted to fit the problem at hand (Mingers, 2001). Reinterpreting one element of a framework while preserving its core structure, in service of a specific and stated theoretical question,

follows the same logic the framework's own author has argued for (Mingers & Willcocks, 2017).

The adaptation is also, in a modest sense, vindicated by what it made visible. The empirical boundary reported as Finding 2, the division between a globally standardized and a locally negotiated layer of organizational vocabulary, is a direct product of applying the distributional dimension, developed in section 2.4 and mapped onto the social/cultural world above, to the case material. It is not obvious that Mingers and Willcocks's original framing would have surfaced this same boundary with the same clarity, since their producer-consumer model does not treat a code's distribution, how widely its meaning extends beyond the group that produced it, as an explicit, operative axis of analysis in its own right.

Table 1 shows where each phase is executed across the thesis.

Phase	Goal	Chapter
Appreciate	Identify the research situation, the actors involved, and the research questions	Chapter 3
Analyse	Investigate the personal, material and social worlds in which the categories function; investigate how the worlds interact; generate explanations	Chapter 4
Assess	Validate results against a human baseline; confirm, eliminate, or revise hypotheses; develop an idealized design in which the problem would not occur	Chapter 4 & 5
Act	Disseminate findings; propose action to improve organizational knowledge management practice	Chapter 6

Table 1: *4As framework adapted (Mingers & Willcocks, 2017)*

3.2 Research design

Consistent with Steps 1 and 2 of the Appreciate phase, collecting initial data and identifying semiotic questions, problems and challenges, then generating semiotic research questions from them, the empirical component of this thesis is designed as an observational case study. The goal is not to benchmark AI performance against a standard, but to observe what happens when AI is applied to organizational knowledge tasks in a real

organizational environment. The design is qualitative, exploratory and explanatory, organized around a single embedded case involving two organizational actors: the company that developed an AI classification product and the enterprise customer whose knowledge base provided the testing environment.

The research design follows directly from the research questions. The questions concern how semantic gaps manifest in organizational practice and what it implies for human and AI roles in knowledge work. These questions cannot be answered by controlled experiments that measure accuracy in isolation. They require observation of the phenomena in a real organizational knowledge environment, where the categories, documents, and definitions carry the organizational history the theoretical framework predicts will matter. A constructed test set would remove precisely the properties that make the phenomena observable.

The empirical study of the AI classification product against real organizational documents from the customer knowledge base was conducted in two phases. The phases were designed to observe what happens when organizational categories are applied by an LLM with varying degrees of definitional specificity: the first phase used general definitions constructed from category names alone, and the second used more precise definitions incorporating format constraints where the category permitted them.

Classification was chosen as the task because it is the most fundamental operation in knowledge organization: assigning a document or knowledge artifact to a category is the minimal act of applying a concept. To ask a machine to classify is to ask it to apply a concept consistently. To evaluate the result is to ask what that concept means and what counts as correct. Both acts expose semantic gaps in directly observable ways.

3.3 Case Context

The case context completes the Appreciate phase by establishing the actors whose practice and interests bear on the categories under study, and the tension between them that motivates the research questions. The case company develops a knowledge management platform serving enterprise clients. The product originated as a document management system and has evolved to include workflows, integrations, and AI-assisted automation. The specific product under study is an LLM-based classification module designed to automate the assignment of document metadata: given a document and a set of organizational categories, the system assigns values to each category without human intervention. The company's practical interest in this study is direct. Classification work

is currently manual and time-consuming, and they want to understand how well the AI product performs relative to human judgment as a step toward automating it.

The knowledge base used for testing belongs to a customer of the case company, an enterprise organization. This organization's document repository constitutes the empirical environment for the testing phase. Its categories, documents, and metadata structures were not designed for research purposes. They were built through organizational practice over time and reflect the knowledge management needs and accumulated decisions of the people who created and maintained them.

The use of a real organizational knowledge base rather than a constructed test set is central to the observational design. A constructed set could be controlled for ambiguity, coverage, and definitional completeness. The point of this study is that those properties are exactly what organizational knowledge bases lack, and that this absence is not a flaw to be corrected but the normal condition in which AI classification products are deployed. The empirical environment had to carry the organizational history the theoretical framework predicts will matter, and a constructed environment could not have provided that.

The case sits at the intersection of the thesis's central tension. The case company is attempting to automate a knowledge organization task using a synchronic model. The customer organization holds a knowledge base whose categories carry meaning constituted through organizational practice. Observing the AI product operating on this knowledge base provides the empirical conditions in which the phenomena described in the theoretical framework can be examined directly: what happens to organizational concepts when an LLM is asked to apply them, and what is required to evaluate whether it has done so correctly.

3.4 Data collection

Identifying suitable documents required navigating the customer knowledge base across several hundred document classes, most of which proved unusable for the study, for reasons worth reporting, since they show what a real organizational knowledge base looks like under scrutiny for internal consistency. Many classes contained no documents at all, or only a handful, well short of what a comparable test set would require. Among classes with enough volume to consider, most still failed on consistency: documents nominally belonging to the same class differed in layout, in which metadata fields were populated, and in content, typically because the class had been used under one template

for a period and then migrated to a new template with different fields and structure, leaving only two or three documents consistent with any single version. Two classes were identified where this was not the case: a set of twenty documents and a set of seventeen, each sharing a single template, a consistent metadata schema, and comparable length and content across the whole set. This consistency was a precondition for the study rather than an incidental convenience: without it, disagreement between the LLM and the human baseline could not be attributed to how a category was interpreted, since it could equally be explained by the documents themselves varying in what information they contained or where it appeared.

Category selection followed a comparable filtering logic, applied to an initial pool of twenty-four candidate categories drawn from the metadata fields associated with the two selected document classes. Fourteen of these were excluded because the documents themselves did not reliably contain the information a category required for a classification decision to be made, by a human or by an LLM, regardless of how the category might be defined. Information Classification was one such case, excluded for two independent reasons: the documents did not consistently record what information classification they held, and, separately, the category's meaning depended on a locally negotiated code for which no external definition could plausibly be constructed. Both obstacles applied to this category, and either alone would have been sufficient to exclude it from the candidate pool.

Of the remaining ten candidates, five were selected for the first test run (Name, Project, Owner, Document date, Customer) on the basis of business relevance: categories judged more useful for downstream classification work, such as identifying a specific project, were prioritized over categories such as keyword tags whose practical value for the task was less clear.

Before data collection could proceed, a preliminary finding emerged. None of the selected categories had explicit definitions attached to them in the knowledge base. Categories existed as name-and-datatype pairs only: a label and a data format, nothing more. No statement of purpose or intended use was recorded for any of them. To proceed with data collection, general definitions were constructed by the researcher from category names and datatypes alone, without access to organizational intent. Construction followed a deliberate, consistent template across categories, a single-sentence definition beginning "A category identifying/specifying...", to avoid introducing variance between categories that came from differences in phrasing rather than from differences in what a category actually asked for. Definitions were drafted individually for each category and then revised together for consistency of structure and register. Content was drawn from

three sources used in combination: informal consultation with staff at the case company on what a category was understood to mean in cases where this was not evident from the name alone, for instance clarifying how Owner was understood to differ from the superficially similar category Author; the datatype attached to the category, which constrained the range of plausible definitions directly, since a date-typed category could not denote a person; and comparison against the content of other metadata fields on the same documents, to establish what kind of information a category already appeared to hold relative to related categories. These researcher-constructed definitions were provided to the LLM alongside each classification task as the operative specification of what each category meant (see Appendix A).

Category selection for the second test run was not a repetition of this process but a targeted follow-up, built directly on what the first run had shown. Test run one's variation was concentrated in categories such as Customer, where no explicit definition existed and no connection to the organizational practice that would fix such a definition was accessible to an outside researcher. Rather than test Customer again under the same conditions, the second run replaced it with Project id and dropped Name for a related but distinct reason: a document's name functions largely as a human-readable summary of properties the other categories already capture directly, useful for a person's own recall but not for a system retrieving on structured, machine-legible categories. Order id was added alongside Project id, both categories sharing Customer's underlying access problem, no locally negotiated definition available from outside, but a different structural property, a known, bounded string format. This substitution was a direct test of the structural dimension developed in section 2.4: if polysemy, a term folding several possible senses into one abstract form, is a driver of interpretive variation, a category constrained to a single structural pattern should show less of it than a category left semantically open, independent of whether its organizational meaning is otherwise locally negotiated or globally standardized.

Document date, Project, and Owner were retained from the first run as an internal check: since they had already shown high agreement under general definitions, holding them constant let the second run isolate the effect of the newly introduced, structurally constrained categories rather than attribute any change in agreement to an unrelated shift in the category set. For two categories where the correct answer was structurally bounded, Project id and Order id, definitions were revised to include explicit format constraints specifying the expected string pattern (see Appendix B). The second evaluation was conducted against the class of seventeen documents using five categories: Project, Project id, Owner, Document date, and Order id.

3.5 Data analysis

This section sets out the data analysis protocol applied to the data described in section 3.4: how the three worlds are investigated (Steps 3 to 5), how their interaction is examined to generate hypotheses (Steps 6 and 7), and how those hypotheses are tested against a human baseline (Step 8). The actual analysis, including the findings it produced, are reported in Chapter 4. Steps 9 and 10, the remainder of Assess, confirm, eliminate, or revise that hypothesis against the literature and sketch an idealized design in which the problem would not arise; these are taken up in Chapter 5. Steps 11 and 12 (Act) state the resulting contribution and propose action to improve organizational knowledge management practice; these close the thesis in Chapter 6.

Table 2 sets out this same arc at the level of individual steps.

Step	Description	Section
1	Collect initial data; identify semiotic questions, problems, challenges	3.2-3.4
2	Generate semiotic research questions	1
3	Investigate personal world	4.1
4	Investigate material world	4.1
5	Investigate social world	4.1
6	Investigate how the worlds interact	4.2
7	Generate hypotheses	4.3
8	Validate results against human baseline	4.4
9	Confirm, eliminate, or revise hypotheses	5.1-5.2
10	Develop an idealized design in which the problem would not occur	5.3

Step	Description	Section
11	State contribution; disseminate findings	6.2,6.5
12	Propose action to improve organizational knowledge management practice	6.3

Table 2: *12 steps adapted (Mingers & Willcocks, 2017)*

Steps 3 to 5 investigate the documents, categories, and knowledge base described in section 3.4 across the three worlds Mingers and Willcocks (2017) identify, interpreted through the dimensions developed in section 2.4 (Mingers & Willcocks, 2017). The personal world (Step 3) is investigated only to the extent needed to identify the relevant actors, the case company's team and the customer organization's staff, without treating either group's private interpretation as an independent analytical category, consistent with the position on public rather than private meaning established in section 3.1. The material world (Step 4) is investigated through the structural dimension: the documents, the categories as name-datatype pairs, and the classification module through which categories are applied to them. The social world (Step 5) is investigated through the distributional dimension: the code, meaning the vocabulary of category names and whatever conventions, global or local, give them meaning for the people who rely on them.

Step 6 investigates how these worlds interact: specifically, whether the code investigated in Step 5 is uniform across the category set or divides along the distributional dimension into a globally accessible layer and a locally negotiated layer, and what that division implies for the material world's capacity to carry the code correctly, since a classification module can only apply a code it has access to. Step 7 generates a hypothesis from this interaction: if the code divides in this way, categories anchored in the globally accessible layer should be classifiable using an externally constructed definition, while categories anchored in the locally negotiated layer should not, regardless of how carefully such a definition is written.

Step 8 tests this hypothesis by validating LLM classification output against an independently constructed human baseline. For each document, the researcher read the document alongside the same definition later provided to the LLM and recorded a baseline value for each of the five categories in turn before moving to the next document. This baseline was recorded before the corresponding LLM output was reviewed, so that the baseline could not be anchored by the model's answer; general familiarity with the tool's

behavior had been gained beforehand through informal testing on other, unrelated documents while definitions were being drafted, but no document or category included in the study was scored by the researcher after seeing the corresponding LLM output.

For each category, the researcher logged every distinct value the LLM produced across the document set and grouped identical outputs into types, following a strict rule that any difference in wording, however small, produced a separate type. What counts as agreement differs by category type. For syntactically bounded categories, this reduces to a straightforward correct/incorrect judgment against the pattern specified in the definition. For semantically open categories, agreement is not judged against a single correct answer, since none is determinable from the definition alone, but against whether a given output constitutes a plausible interpretation of the definition as written, one that could reasonably be derived from the document under the stated definition, rather than one that merely differs in form from other outputs. This makes the evaluation procedure artificial in a deliberate sense: it is designed to surface variation in how an underspecified definition gets interpreted, on the understanding that the human baseline is itself one interpretation constructed outside organizational context, not an authoritative ground truth. The type counts this procedure produced are reported in Chapter 4.

4. ANALYSIS

This chapter presents the empirical analysis of the case study, applying the semiotic framework to the organizational knowledge base. Section 4.1 investigates the personal, material, and social worlds through the structural and distributional dimensions, establishing that the platform's material schema lacks explicit category definitions. Section 4.2 examines the interaction between these worlds, demonstrating how the material representation conceals the division between globally standardized and locally negotiated organizational vocabulary. Building on these interactions, Section 4.3 generates two hypotheses concerning how structural openness and explicit format constraints affect a language model's classification agreement. Section 4.4 then tests these hypotheses against an independent human baseline across two test runs, confirming that classification agreement tracks directly with syntactic boundedness and that structural constraints eliminate interpretive variation. Finally, Section 4.5 summarizes the four empirical findings and their underlying mechanisms, establishing the foundation for the theoretical assessment in the subsequent chapter.

4.1 Investigating the three worlds

Steps 3 to 5 investigate the personal, material, and social worlds Mingers & Willcocks (2017) identify, drawing on the three dimensions of organizational meaning developed in section 2.4: the structural dimension, how abstract or concrete a category's representation is and how many senses it folds together; the distributional dimension, how widely a category's governing code is shared beyond the group that produced it; and the temporal dimension, whether that code is treated as fixed or as continuously revised through use. These three dimensions do not map onto the three worlds one-for-one. The material world is investigated here through the structural dimension, and the social world through the distributional dimension. The personal world, consistent with the position established in section 3.1, is investigated only to identify the relevant actors, not through any dimension taken to reveal a private, subjective meaning, since meaning is treated throughout this thesis as constituted publicly rather than privately (Kripke, 1982; Wittgenstein, 1953). The temporal dimension is not applied to a single world in this section at all: it becomes analytically active only once the material and social worlds are shown interacting, and is developed as such in section 4.3, once section 4.2 has established what that interaction looks like. This section works through the personal, material, and social worlds in turn.

The personal world (Step 3) is occupied by two groups whose relationship to the categories differs in kind: the case company's team, who produce the categories as structural fields in the platform, fixed in advance of any document they will later be applied to; and the customer organization's staff, whose ongoing practice would, over time, be what fixes what a category like Owner or Customer actually picks out in this organization, if anything does. Consistent with the position established in section 3.1, this world is investigated only to identify these two groups; neither group's private interpretation is treated as an independent object of analysis.

The material world (Step 4) is occupied by the documents, the categories as name-datatype pairs, and the classification module that applies the latter to the former. Investigated through the structural dimension, the material world here shows a single, striking property: the classification module represents every category identically, as an empty field awaiting a value, regardless of what kind of code fills that field once a value is assigned. A category whose governing code is the Gregorian calendar and a category whose governing code is an internal organizational policy are stored in exactly the same structural form, a label and a datatype. The platform's material representation carries no marker of which kind of code, in Step 5's terms, a given category actually depends on. This follows from what codification, in Boisot's sense (section 2.4), does to a category by definition: abstracting it into a general, structured form strips away precisely the contextual particulars that would otherwise signal how locally or globally that category's meaning is anchored.

The social world (Step 5) is occupied by the code itself: the vocabulary of category names and whatever conventions give them meaning for the people who rely on them. Investigated through the distributional dimension, the social world was the source of the first empirical observation of the study. Navigating the customer's knowledge base surfaced several hundred metadata categories in total, and none of them, at any point, had an explicit definition attached. Categories existed purely as name-and-datatype pairs, with no statement of purpose or intended use recorded anywhere for any of them (Finding 1). This absence is not a documentation failure the organization neglected to correct; it is consistent with the theoretical position of section 2.5, that a category's meaning lives in the practice of using it, not in a declarative statement prepared for a reader who, from inside the practice, was never expected to need one. The reader who does need one, an outside researcher, a newly onboarded employee, or a language model trained on text from outside this specific practice, is exactly the reader this absence disadvantages, and section 4.2 takes up what happens when that disadvantage is made operational.

4.2 Interaction of the worlds: the local/global divide

Step 6 examines how the material and social worlds interact, and how the choices of the actors identified in the personal world (Step 3) produced that interaction. This is worth stating plainly before the empirical detail, since without an explicit frame the interaction is easy to misread as a sequence of filtering decisions rather than as an analytical claim in its own right. It is not the filtering itself that constitutes Step 6. It is what the filtering had to work around: a material world that, by design, gives no signal of the social world's real structure.

Two distinct collisions between the material and social worlds are examined below, summarized in Table 3 before being developed in full.

Interaction	What collides	Result
Representation gap	The material world's schema stores a category as a name and a datatype only; it has no field capable of holding a definition, a stated purpose, or a governing code	Finding 1: no category in the knowledge base carries an explicit definition, independent of whether its underlying code is global or local
Concealment	The material world represents every category identically, as an undifferentiated structural field, whether the code that fills it is globally accessible or locally negotiated	Finding 2: the distributional divide between category types is invisible in the material artifact itself, and could only be surfaced by filtering categories against the documents and against consultation with people at both organizations

Table 3: *Material-social interactions examined in this section*

The representation gap. The case company's platform, examined in section 4.1, stores every category as a label and a datatype and nothing more. This is a material fact about the schema itself, not an oversight specific to any one category: the platform provides no field in which a definition, a stated purpose, or any account of a category's governing code could be recorded, even for a category whose meaning the case company's own team might have been able to state clearly if asked. The social world's code, whatever form it takes for a given category, therefore has nowhere to be written down inside the material world at all. Finding 1, that no category anywhere in the knowledge base carried an explicit definition, is the direct consequence of this collision: it is not that the customer organization neglected to document its vocabulary, but that the material artifact through which the vocabulary is used has no place for that documentation to live.

Concealment and the filtering process as its empirical trace. The second collision is sharper. Because the material world represents every category identically, an undifferentiated field awaiting a value, nothing in the platform itself distinguishes a category whose code is globally accessible from one whose code is locally negotiated. This is what made the distributional divide invisible until it was investigated directly. The two-stage filtering process described in section 3.4, twenty-four candidates reduced to ten on grounds of information completeness, then further examined for the accessibility of their governing code, is the empirical trace of this concealment being worked through by hand: since the material world offered no marker to distinguish the two kinds of category automatically, distinguishing them required the researcher to investigate the social world directly, consulting case-company and comparing categories against document content, in effect doing manually what the platform's material representation does not do structurally.

Information Classification is the sharpest instance of this. As section 3.4 established, it failed the completeness filter for an ordinary reason, the documents did not consistently record it, shared with other excluded categories such as unlinked sales records or unassigned teams. But unlike those categories, which denote globally intelligible organizational concepts simply left unrecorded, Information Classification denotes a governance function whose meaning is locally negotiated and politically inflected: even a document that recorded its classification explicitly would not have made that value interpretable without independent access to the policy assigning it. Nothing in the material world marks this difference. A field named "Information Classification" and a field named "Document date" are stored identically, as a label and a datatype, and the concealment this section examines is precisely that fact: the material world cannot show, and does not attempt to show, that one of these fields depends on a code available to any sufficiently informed outside reader while the other depends on a code available only to someone inside the organization's own governance practice.

Where the personal world enters. Both collisions trace back to decisions made by the actors identified in section 4.1, not to any inherent property of metadata as such. The case company's team designed a schema that captures a category's datatype but not its code, a design choice, not an oversight forced by the technology. The customer organization's staff, in turn, never had reason to formalize their shared understanding of a category into that schema, since their own practice made the category's meaning clear enough without it. Neither decision reflects any individual's private interpretation of what a category means, consistent with the position in section 3.1; both are practice-level facts about how each organization built and used the system. It is the combination of these

two practice-level choices, one organization's schema design and the other's working habits, that produces the concealment this section has examined, and that made a manual filtering process necessary before any category could be tested at all.

4.3 Generating the hypothesis

Step 7 generates a hypothesis from the interaction examined in section 4.2. That interaction established that the material world represents every category identically regardless of its code, while the social world's actual code divides sharply along the distributional dimension into a globally accessible layer and a locally negotiated layer, and that only the first layer could be brought into this study at all. The hypothesis this generates is not simply that the remaining, globally accessible categories will be classified correctly and the excluded ones would not have been; that much follows directly from Finding 2 and requires no further test. The open question is what happens to agreement within the globally accessible layer once a category is handed to an LLM with nothing but a general, researcher-constructed definition.

Here the temporal dimension, held aside from the personal world in section 3.1 and applied through the material and social worlds instead, becomes directly relevant. A category's code being accessible from outside the organization is not the same as that code being temporally stable. Section 2.4 established, following Hamilton, Leskovec, and Jurafsky (2016), that the words most prone to drifting in meaning over time are precisely the words that are more polysemous, that fold more distinct senses into one form (Hamilton et al., 2016). This suggests that global accessibility and temporal stability are not the same property and do not necessarily travel together: a category can be globally intelligible in the abstract, decodable by any sufficiently informed outside reader, while still being structurally abstract and polysemous enough that its specific, current sense inside a given organization has room to have narrowed or shifted away from whatever a general, unanchored definition would predict. A language model's representation of such a category is a synchronic snapshot in exactly the sense section 2.3 describes, and a general definition constructed without reference to any particular moment of organizational practice cannot by itself collapse that openness into the one sense the organization currently intends.

The first hypothesis follows directly: among the five categories selected for the first test run, agreement between the LLM's output and the human baseline, both operating from the same general definition, will vary systematically with each category's position on the structural dimension, rather than being uniformly high or low across the set. Categories with a syntactically bounded correct answer, where the structural dimension permits only

one form, should show near-complete agreement, since a general definition is sufficient to fix a single answer regardless of how that category's underlying code is distributed. Categories that are semantically open, where the structural dimension permits several senses to be folded into one label, should instead produce multiple distinct but individually plausible outputs, since nothing in a general definition can determine which of those senses the organization currently intends. This hypothesis is tested as Test Run 1, reported in section 4.4.

Consistent with the iterative character of Steps 6 through 8 as applied here, Test Run 1's results (section 4.4) generated a second, more targeted hypothesis rather than concluding the analysis. Variation in the first run was concentrated in the categories with the most structurally open definitions (section 3.4), which raises a sharper question than the first hypothesis alone could answer: is it globally accessible but structurally open categories in general that produce variation, or is variation reducible specifically to structural openness, independent of anything else about a category's distributional status? The second hypothesis isolates this: if structural constraint, rather than distributional status, is what drives agreement, then adding an explicit format pattern to a category's definition should reduce variation for that category regardless of whether its broader organizational meaning remains otherwise undefined. This hypothesis is tested as Test Run 2, also reported in section 4.4.

4.4 Testing the hypotheses: baseline validation

Step 8 validates the LLM's output against the human baseline established in section 3.5, for each of the two hypotheses generated in section 4.3.

Test Run 1. The first hypothesis predicted that agreement between the LLM and the baseline would vary systematically with a category's position on the structural dimension: categories with a syntactically bounded correct answer should show near-complete agreement, while categories left semantically open should produce multiple distinct, individually plausible outputs. Table 4 reports, for each category, the number of distinct output types the LLM produced across the twenty documents and how many documents fell into each type.

Category	Structural status	Distinct types	Occurrences per type
Document date	Syntactically bounded	1	20

Category	Structural status	Distinct types	Occurrences per type
Owner	Syntactically bounded	1	20
Project	Semantically open	2	18, 2
Name	Semantically open	3	18, 1, 1
Customer	Semantically open	6	10, 5, 2, 1, 1, 1

Table 4: *Test Run 1 results (twenty documents)*

The hypothesis holds, though as a gradient rather than a clean binary. Document date and Owner, the two categories whose definitions, though general, admit only one form of answer, a date, a personal name, produced complete agreement. All three semantically open categories produced variation, but of noticeably different magnitudes, which is itself informative: Project's two types differ only in whether a project identifier was appended to the project name, both plausible readings of a definition that specifies only "the project associated with the document." Customer's six types spread far more broadly, across combinations of company name, individual name, and title, matching a definition, "the client or organization for whom the document is created, delivered, or associated," that is open on more than one axis at once: it does not specify whether an organization or a person within it is wanted, nor whether a role title should accompany a name. Customer's variance is the sharpest confirmation of the first hypothesis, since it is also the most open of the three definitions.

Name's variation is worth examining directly rather than only counting, since it complicates a simple story about the LLM introducing semantic drift. Eighteen of twenty documents produced one type; two produced distinct singleton outputs. Inspection of the source documents showed that the two divergent outputs both derived directly from the title already present on the document itself, and that one of those two documents had, at some earlier point, been given a title by a human author that departed from the naming convention used across the rest of the class. The model did not introduce an incon-

sistency here. It reproduced one that already existed in the human-authored record, faithfully extracting a name from a document whose own title had already broken from the pattern. This distinction, between an LLM introducing variance and an LLM surfacing variance already present in organizational data, is not visible in the type count alone and only appears on inspecting the actual case.

Test Run 2. The second hypothesis predicted that adding an explicit format constraint to a category's definition would reduce variation for that category specifically, independent of whether its broader organizational meaning remained otherwise undefined. Table 5 reports the equivalent results for the second test run, against the class of seventeen documents.

Category	Structural status	Distinct types	Occurrences per type
Project id	Newly constrained (format added)	1	17
Order id	Newly constrained (format added)	1	17
Project	Retained, unconstrained	1	17
Owner	Retained, syntactically bounded	1	17
Document date	Retained, syntactically bounded	2	15,2

Table 5: *Test Run 2 results (seventeen documents)*

Project id and Order id, the two categories given an explicit string-pattern constraint, produced complete agreement, directly confirming the second hypothesis: structural constraint eliminated the variation a semantically open definition would otherwise permit, without requiring anything to be resolved about whether the category's meaning was locally or globally anchored. This is worth stating precisely, since it is easy to overstate: the finding is that constraining structure closes off interpretive openness, not that it resolves the distributional status of the underlying category. Order id, for instance, remains

a category whose organizational meaning would still need to be established through practice; the format constraint only fixed the shape its answer could take.

Two results in Table 5 need to be flagged rather than absorbed quietly into the pattern above. First, Project, tested here without any added constraint, produced complete agreement across all seventeen documents, unlike its two-type result in Test Run 1. This is a genuine discrepancy, not confirmation of anything the second hypothesis predicted. The data available here cannot isolate a specific cause: candidate explanations include the smaller size of the second document set (seventeen documents against twenty in the first) and any difference in how uniformly this particular class's documents happened to be structured, but section 3.4 selected both classes under the same consistency criterion, so neither explanation can be confirmed from what was collected. It is noted here as an unresolved boundary of the second test set, rather than pursued further, since its practical weight is limited to a single category's result and does not bear on Findings 3 or 4. Second, Document date, fully consistent in Test Run 1, produced two disagreements in Test Run 2. No clear cause for these two cases could be identified from the documents themselves; they are reported as an open discrepancy rather than explained away.

Together, these two test runs support the following:

Finding 3: General, researcher-constructed definitions produced high agreement for syntactically bounded categories and consistently produced multiple distinct, individually plausible outputs for semantically open categories, with the degree of variation tracking how many axes a definition left open rather than appearing uniformly across all open categories.

Finding 4: Adding an explicit structural constraint to a category's definition eliminated interpretive variation for that category, independent of whether the category's broader organizational meaning remained otherwise undefined, though not every result across the two runs is fully accounted for by this pattern.

4.5 Summary of findings

The four findings established across sections 4.1 to 4.4 follow the sequence of Steps 3 to 8. Findings 1 and 2 emerged from investigating the three worlds and their interaction (Steps 3 to 6), before any classification testing took place. Findings 3 and 4 emerged from testing the hypotheses that interaction generated (Steps 7 and 8). Table 6 collects all four in one place, alongside the specific mechanism that each finding instantiates.

Finding	What was found	Mechanism	Established by
Finding 1	No category in the knowledge base carried an explicit definition	Representation gap: the material world's schema stores a label and datatype only, with no field for a code, purpose, or definition	4.1, 4.2
Finding 2	Organizational vocabulary divides into a globally standardized layer and a locally negotiated layer; only the former could be brought into the study	Concealment: the material world represents every category identically, giving no signal of which layer a given category belongs to	4.2
Finding 3	General definitions produced high agreement for syntactically bounded categories and variable, multi-type agreement for semantically open categories, with variation tracking how many axes a definition left open	Structural dimension: a definition open on more axes permits more distinct, individually plausible interpretations	Test Run 1 (4.4)
Finding 4	Adding an explicit structural constraint eliminated variation for a category, independent of whether its organizational meaning remained otherwise undefined	Structural dimension: a format constraint fixes the shape a valid answer must take, regardless of the category's distributional status	Test Run 2 (4.4)

Table 6: *Summary of empirical findings*

Findings 1 and 2 concern what the study could examine at all: an undocumented vocabulary that, once investigated, divides into an accessible layer and an inaccessible one. Findings 3 and 4 concern what happened once testing was carried out on the accessible layer: agreement tracked structural openness, and constraining that structure closed the

gap it otherwise produced. Taken together, these four findings describe what was observed under the specific conditions of this case. They do not yet establish what these results mean for the theoretical claims of Chapter 2, for the wider literature, or for how human and AI judgment should be divided in practice. That assessment, confirming, revising, or setting aside hypotheses against the literature (Step 9), and sketching an idealized design in which the problems observed here would not arise (Step 10), is the work of Chapter 5.

5. DISCUSSION

This chapter interprets the empirical findings through the thesis's theoretical framework, moving from observation to assessment. Section 5.1 evaluates the four empirical findings against the established literature, confirming the hypotheses regarding structural openness and contextualizing the observed documentation gaps to answer the first two research questions. Section 5.2 then examines the semantic access barrier, reframing the study's inability to test locally negotiated categories not as a methodological limitation, but as a direct empirical manifestation of the core phenomenon. Finally, Section 5.3 sketches an idealized knowledge management design, proposing counterfactual material and social structures to mitigate the observed issues and laying the groundwork for defining the necessary division of labor between human judgment and artificial intelligence.

5.1 Assessing the hypotheses against literature

Step 9 confirms, eliminates, or revises the hypotheses generated in section 4.3 against the theoretical framework developed in Chapter 2. Findings 1 and 2 did not emerge from a stated hypothesis; they emerged from investigating the three worlds and their interaction (Steps 3 to 6), and are assessed here against the general theoretical account of organizational meaning in sections 2.1 to 2.5. Findings 3 and 4 resulted directly from testing the two hypotheses stated in section 4.3, and are assessed against those hypotheses specifically.

Finding 1, that no category in the knowledge base carried an explicit definition, confirms Peirce's pragmatic maxim more directly than the theoretical discussion in section 2.2 could establish on its own: if a concept's meaning is the sum of its practical effects rather than a declarative statement about it (Peirce, 1878), then a category used correctly every day by people who never needed to write down what it meant is exactly what the maxim predicts, not a documentation failure to be corrected. Finding 1 also confirms Tsoukas's account of firms as distributed knowledge systems lacking any single overseeing mind (Tsoukas, 1996): the representation gap identified in section 4.2 is a direct instance of this, a material schema built by one organization's team, holding categories whose meaning was never centralized because it was never held by any single agent in the first place, only distributed across the practice of the people using them.

Finding 2 gives the distributional dimension developed in section 2.4 its first empirical instance. Boisot's diffusion axis, introduced there as a general property along which knowledge moves from undiffused to diffused (Boisot, 1998), is not merely illustrated by the global/local divide reported in Finding 2, it is what that divide is: categories such as Document date sit at the diffused end because their code is available wherever the calendar itself is available, and categories such as Information Classification sit at the undiffused end because their code exists only inside this organization's own governance practice, sustained the way Stamper describes a norm being sustained, as a disposition shared by a community rather than a property of the word itself (Stamper et al., 2000), and the way Wenger describes such norms being produced, through repeated joint engagement rather than through any act of definition (Wenger, 1998).

Together, Findings 1 and 2 answer RQ1: "How do humans and large language models differ in the way they organize and maintain meaning?". Humans maintain organizational meaning through practice that is rarely externalized, sustained by community norms that need not be written down for the community that holds them, and divided, as Finding 2 shows, into a layer whose code happens to coincide with wider convention and a layer whose code does not. A large language model, as a synchronic system trained on externally available text (section 2.3), can only have learned the first layer, not because it performs the second layer poorly, but because the second layer was never externalized into anything a model, or any outside reader, could have learned from. This is a difference in what each kind of agent has structural access to, not a difference in fluency or competence.

Findings 3 and 4 confirm the two hypotheses from section 4.3, though the pattern of confirmation is worth stating precisely rather than as a flat yes. The first hypothesis, that agreement would track a category's position on the structural dimension, held as a gradient: Document date and Owner produced complete or near-complete agreement, while Project, Name, and Customer produced variation whose magnitude scaled with how many axes each definition left open. The one candidate explanation this section can eliminate outright is the possibility that this variation reflects the model introducing semantic drift on its own initiative: the Name inspection in section 4.4 showed the model faithfully reproducing an inconsistency already present in the human-authored record, not manufacturing one. The second hypothesis, that structural constraint reduces variation independent of a category's distributional status, was confirmed directly by Project id and Order id, both reaching complete agreement without either category's underlying organizational meaning being resolved. Document date's two unexplained disagree-

ments in Test Run 2 sit outside this pattern and are not accounted for by either hypothesis; they are carried forward as an open question rather than folded into a confirmation they do not support.

Read together, Findings 3 and 4 answer RQ2: "What challenges arise when large language models are applied to organizational knowledge tasks in practice?". The challenge an LLM faces when applied to organizational knowledge tasks is not primarily one of getting things visibly wrong. Customer's six output types were not six errors; each was, on its own terms, a plausible reading of an underspecified definition. This is the semantic gap from section 2.5 in its most operationally precise form: it does not announce itself, because a semantically open category produces outputs that pass as reasonable rather than outputs that fail. Finding 4 adds a second, more specific challenge: because structural constraint is effective and cheap to apply, it invites being mistaken for a semantic solution when it is a syntactic one. Fixing the shape of an answer, as Project id and Order id show, does nothing to establish what an unconstrained category such as Customer should mean, and an organization that constrains its way to high agreement scores risks concluding a category is well understood when only its format has been resolved.

One further point needs to be stated as a revision rather than a confirmation. The theoretical motivation for the first hypothesis, developed in section 4.3, invoked the temporal dimension: Hamilton et al. (2016) finding that polysemous words drift faster than words with a single sense was the reasoning that connected structural openness to the risk of a synchronic model misrepresenting a category's current meaning. What Test Run 1 actually measured, however, is a single comparison at one point in time between the LLM and a human baseline, not a comparison of the same category's meaning at two different points in time. The finding confirms that structurally open categories produce more interpretive variance than structurally bounded ones in this cross-sectional sense. It does not, on its own, confirm that this variance is caused by temporal drift specifically, as opposed to an ambiguity that would have existed even in a hypothetical organization whose vocabulary never changed at all. The temporal dimension remains the more plausible explanation for why organizational categories become structurally open in the first place, since the process Peirce and Saussure describe is precisely how a term accumulates multiple senses over time, but this thesis's data cannot adjudicate between that explanation and a purely synchronic account of ambiguity. This limitation is taken up again in section 5.2.

What Findings 1 and 2 could not reach is the locally negotiated layer itself, the categories this study's own filtering process excluded because no external definition could be constructed for them at all. Section 5.2 takes up what that exclusion means for what this thesis can and cannot claim to have shown.

5.2 The semantic access barrier as the phenomenon

Section 5.1 closed by noting what Findings 1 and 2 could not reach: the locally negotiated layer of vocabulary, excluded from testing because no external baseline existed against which to evaluate it. Read together with Findings 3 and 4, this exclusion supports an interpretation that goes beyond what any single finding establishes on its own: the study tested definitional matching against the globally standardized layer of organizational vocabulary, not the organizationally produced meaning of the vocabulary as a whole. Test Runs 1 and 2 showed that an LLM can match a definition to information present in a document with reasonable consistency, improving further once definitions carried explicit format constraints. But Finding 2 established that the categories available for this kind of testing were, by construction, drawn from the layer of the vocabulary anchored in a globally standardized code, or in concepts general enough that an outside-constructed definition could plausibly apply. Categories anchored in a locally negotiated code, whose meaning depended on organizational practice rather than any external convention, could not be brought into the study at all, because no baseline existed from which to evaluate them.

This reframes what looked, at first, like a limitation of the research design. The study could not test the theoretically most significant categories because their meaning was constituted entirely in local organizational practice, without a globally shared code to decode against. That absence is better understood as a direct encounter with the phenomenon the theoretical framework describes than as a shortfall to be corrected in a future study. The semantic access barrier the researcher encountered is the same barrier any outside agent encounters when approaching organizationally embedded meaning. The semantic gap does not only manifest as LLM error in production. It manifests as inaccessibility at the point of definition, before the machine is ever involved.

The LLM tested here performed well within the scope of what could be defined from outside, using categories underwritten by a globally standardized or generally intelligible code. That is not a finding about the gap between machine and organizational meaning closing. It is a finding about the narrow domain in which outside agents, human or machine, can operate without encountering that gap. The categories that would have revealed the gap most directly, the locally negotiated ones identified in Finding 2, could not

be brought into the study. This is consistent with the theoretical prediction in section 2.5, that the semantic gap tends to stay hidden: the conditions that make it hidden are the same conditions that made the locally negotiated categories inaccessible here. The study does not demonstrate the gap directly. It demonstrates the conditions under which the gap remains unobservable, and the empirical boundary between the globally standardized and locally negotiated layers of the vocabulary, Finding 2, is itself a form of evidence for the theoretical claim.

Retrieval-augmented generation (RAG) does not overcome this barrier. A natural response to these findings is that grounding the LLM in the organization's own documents at query time would give it access to organizational meaning (Lewis et al., 2020). This response is partially correct and does not resolve the structural problem. Organizational documents are synchronic artifacts: they record what was agreed at a point in time, not the living practice through which meaning is produced and updated. Accessing documents from outside organizational practice does not restore participation in that practice. The researcher had access to the customer organization's documents throughout this study and still encountered the access barrier described in Finding 2. RAG moves the problem upstream, from the model's training data to the knowledge base, without changing its character.

A second, distinct limitation follows the same logic and needs to be stated with the same directness. Section 5.1 noted that Test Run 1's design measured a single comparison at one point in time, the LLM against a human baseline, and could not on its own confirm that structural openness produces variation because of temporal drift specifically, as opposed to an ambiguity that would exist even in a hypothetical organization whose vocabulary never changed. This is not a separate weakness of the study, but the same access barrier applied to a different axis. The locally negotiated layer of vocabulary was inaccessible to this study because testing it would have required participating in organizational practice the researcher stood outside of. Diachronic change is inaccessible to this study for a structurally identical reason: observing it directly would require comparing the same category's meaning at two separated points in organizational time, and a single case study, conducted once, by a researcher positioned outside the organization at a single moment, cannot do this any more than a synchronic language model can. The researcher's own method, in this respect, occupies the same position relative to the phenomenon that the theoretical framework attributes to the LLM under study: an external, single-snapshot vantage point on a process that is by definition ongoing. This does not weaken the temporal dimension's role in the overall argument, since its plausibility rests on independent grounds, Hamilton et al. (2016) longitudinal evidence that polysemous

words drift measurably faster over historical corpora, established in section 2.4 without depending on this thesis's own data. What it does mean is that this study's own empirical contribution to the temporal dimension is indirect, motivating why structurally open categories would be expected to drift rather than directly observing that drift, and this distinction should not be blurred in how the findings are represented.

Read together, this section and the one before it define the actual evidential shape of this thesis's empirical contribution: direct, demonstrated confirmation of the structural and distributional dimensions, developed through Findings 1 through 4, and indirect, theoretically motivated support for the temporal dimension, resting on the literature the study draws on rather than on data the study produced. Section 5.3 takes this shape as its starting point in sketching what a knowledge management design would need to look like for the barriers identified here not to arise.

5.3 Sketching an idealized design

Step 10 asks a different kind of question than Step 9. Rather than confirming or revising a hypothesis against what was observed, it asks what a semiotic world would need to look like for the communication problems identified in Chapter 4 not to arise at all. This section is accordingly speculative rather than empirical: it does not propose that any of the following should be implemented, only sketches what would need to be true of the material and social worlds for each mechanism identified in section 4.2 to be absent. Converting this sketch into an actual recommendation is Step 12's task, taken up in Chapter 6.

Table 7 sets out six such counterfactuals, one for each mechanism identified across Chapters 4 and 5, before the last two are discussed separately, since they behave differently from the rest.

Mechanism	**Problem observed**	**Idealized counterfactual**
Representation gap (Finding 1)	The material schema stores a label and a datatype only, with no field capable of holding a definition or stated purpose	A schema in which a stated purpose is a condition of a category's existence, not an optional addition made after the fact

Mechanism	Problem observed	Idealized counterfactual
Concealment (Finding 2)	The material world represents every category identically, giving no signal of whether its code is globally or locally anchored	A schema that records, for every category at the point it is created, whether its governing code is globally standardized or locally negotiated
Structural openness (Finding 3)	Definitions left open on multiple axes produced multiple plausible, divergent outputs	Definitions that close off ambiguity axis by axis, stating what a category is for and what it deliberately excludes, not only what datatype it takes
Structural constraint's limits (Finding 4)	A format constraint fixed an answer's shape without resolving what the category meant	A schema that distinguishes syntactic constraints from semantic resolution explicitly, so agreement on shape is never recorded as agreement on meaning
Temporal dimension (section 5.2)	A definition, once written, has no mechanism prompting it to be revisited as practice evolves	A definition that carries a creation date and a recurring review trigger, rather than being treated as permanent once written
Semantic access barrier (Section 5.2)	No external definition could be constructed; an outside agent, human or LLM, has no route to the code at all	Categories flagged as locally negotiated are excluded from automated classification by design and routed to someone with current standing in the relevant practice

Table 7: *Idealized counterfactuals for each identified mechanism*

The first four rows describe genuine repairs. Nothing in the theoretical framework of Chapter 2 suggests that a representation gap, a concealment, an underspecified definition, or a conflated syntactic and semantic result are necessary features of organizational knowledge management. Each is a design choice the case company's platform happened to make, not a structural limit on what such a platform could do, and each admits a straightforward counterfactual in which it simply was not made.

The fifth row, the temporal dimension, admits a genuine repair too, but one worth grounding in what Chapter 2 already established rather than inventing fresh. Boisot's own account of the codification and diffusion axes includes what he terms a social learning cycle, the recognition that codified knowledge is not static but continues to be revised as practice develops (Boisot, 1998). An idealized design that attaches a review trigger to a definition is not adding something foreign to this framework; it is building a material mechanism for a process the theory already describes as ongoing, one the case company's actual schema, storing a category as a fixed label and datatype, gave no means of expressing.

The sixth row, the semantic access barrier, is different in kind, and stating why matters more here than in any of the other five. The locally negotiated layer of vocabulary is not locally negotiated because no one has yet written its definition down, in the way Finding 1 showed globally accessible categories to be undocumented. It is locally negotiated because its meaning is constituted by the norms of a specific community's practice (section 2.4), and no act of documentation converts a locally sustained norm into a globally accessible code without changing what it is. An idealized design cannot make Information Classification decodable by an outside agent the way Document date already is; doing so would require either replacing the practice that gives the category its meaning, which is not a documentation problem to solve, or accepting a definition thin enough to lose what made the category useful to the organization in the first place. What an idealized design can do is the more limited thing shown in the table: make the boundary itself visible in the material world, so that the category is correctly flagged as one an outside agent, human or machine, should not attempt alone. This is not a weaker finding than the other five. It is the point the other five would otherwise obscure, that better documentation closes some gaps and cannot close this one.

Read together, these six counterfactuals begin to answer RQ3, what these challenges imply for the division of labor between humans and AI. Five of the six describe repairs to how categories are documented and constrained, work that improves what any outside agent, human or machine, can do with the globally accessible layer of an organization's

vocabulary. The sixth describes something categorically different: not a repair to documentation, but a design that routes certain categories away from automated processing entirely, because no documentation could make them appropriate for it. Chapter 6 offers the practical proposals to resolve RQ3.

6. CONCLUSIONS

This chapter concludes the thesis by synthesizing its core arguments and outlining its broader implications for both research and practice. Section 6.1 restates the central thesis, affirming that the boundary between human and machine interpretation is a structural consequence of an LLM's synchronic architecture rather than a temporary technical limitation. Section 6.2 details the study's theoretical and empirical contributions, clarifying which dimensions of organizational meaning were directly demonstrated and which remain theoretically supported. Moving from theory to application, Section 6.3 translates the idealized counterfactuals of the previous chapter into concrete, actionable proposals for organizational knowledge management design. Section 6.4 transparently acknowledges the limitations of the study, including its single-case scope and the indirect nature of its temporal evidence. Finally, Section 6.5 identifies critical gaps for future research, suggesting pathways for longitudinal and insider studies to further test the phenomena identified in this thesis.

6.1 Central Thesis

The thesis opened with the observation that coordination fails not when people lack information but when they mean different things by the same words, and asked whether an LLM can be deployed to manage organizational knowledge without deepening that problem. The answer this thesis has built toward, across the theoretical synthesis of Chapter 2 and the empirical findings of Chapters 4 and 5, is that an LLM can participate in one layer of organizational vocabulary and cannot participate in another, and that the boundary between the two layers is not a matter of degree to be closed by a better model. It is a structural consequence of what a language model is: a synchronic system, trained on externally available text, applied to a distributional dimension of meaning that divides sharply into a globally accessible layer and a locally negotiated layer (Finding 2). The model tested here performed reliably within the first layer once definitions were made structurally explicit (Findings 3 and 4). It had no route to the second layer at all, not because the model performed poorly there, but because that layer was never externalized into anything an outside agent, human or machine, could learn from (section 5.2). Organizations that deploy such a model without this boundary in view are not making a technical error correctable by a newer model. They are asking a synchronic system to do diachronic, community-held work it has no structural access to.

The significance of this boundary lies less in its existence than in how it behaves once organizations act on it without knowing it is there. Section 5.2 showed that grounding a model in an organization's own documents does not close the boundary, since documents are themselves synchronic artifacts, records of what was agreed at a point in time rather than the living practice through which meaning continues to be negotiated. Finding 3 showed something sharper still: the boundary does not announce itself through visible failure. A semantically open category produced several distinct outputs, each individually plausible, none flagged by the model as uncertain. An organization watching only for errors would have seen none. This is the condition section 2.5 predicted before any data were collected, that the gap tends to stay hidden precisely because fluent output and correct output are not the same thing, and it is the condition this thesis's empirical chapters found directly rather than merely anticipated.

6.2 Theoretical and Empirical Contribution

Step 11 asks that a study's results be disseminated in a way that states plainly what they correct or add to existing understanding. This section takes up that task for the theoretical and empirical contribution directly; section 6.5 takes up the further research gaps Step 11 also calls for.

The theoretical contribution of this thesis is the synthesis developed in section 2.4: organizational meaning can be described along three dimensions, temporal, structural, and distributional, each with its own literature and its own mechanism for why a synchronic system runs into difficulty in a diachronic environment. This synthesis is not merely descriptive. Section 3.1 showed it doing methodological work, providing a principled basis for adapting Mingers & Willcocks (2017) three worlds to a case their own producer-consumer model does not foreground in the same way, and section 4.1 showed it doing analytical work, giving the empirical material a vocabulary sharper than "categories vary" or "categories are undocumented".

The empirical contribution follows the shape section 5.2 made explicit. Two dimensions received direct, demonstrated support. The distributional dimension is confirmed by Finding 2: organizational vocabulary divides into a globally standardized layer and a locally negotiated layer, and only the former could be brought into this study at all, a boundary visible in the category data itself before any classification testing began. The structural dimension is confirmed by Findings 3 and 4: agreement between the LLM and a human baseline tracked how many axes a definition left open, and constraining that structure eliminated variation directly, independent of whether a category's broader organizational meaning remained otherwise undefined. The temporal dimension received

a different kind of support. Its role in the argument, why organizationally embedded categories become structurally open and prone to drift in the first place, rests on Hamilton et al. (2016) longitudinal evidence from computational linguistics, not on data this thesis produced; this study's own design, a single case examined at one point in time, could motivate that role theoretically but could not test it directly. Stating this precisely, rather than allowing the temporal dimension to appear as fully demonstrated alongside the other two, is itself part of what this thesis contributes: an account of which claims a case study of this kind can support directly and which it can only make plausible.

6.3 Practical Proposals

Step 12 converts the idealized counterfactuals sketched in Table 7 into concrete proposals for organizational knowledge management practice. Table 8 maps each counterfactual onto a specific, implementable design element.

Counterfactual (Table 7)	**Practical proposal**
A schema in which a stated purpose is a condition of a category's existence	Require a purpose statement, why the category exists and what decision it supports, before a category can be created in the platform, not as an optional field added later
A schema that records whether a category's code is global or local	Require a distributional flag at creation: globally standardized or locally negotiated. This flag is not decorative; it is the trigger for the two proposals that follow
Definitions that close off ambiguity axis by axis	For categories flagged globally standardized, require the purpose statement to specify what the category includes and what it excludes, not only its datatype, closing the axes Finding 3 showed a bare name-and-datatype pair leaves open
A schema that distinguishes syntactic constraint from semantic resolution	Record format constraints and purpose statements as separate fields, and report classification agreement for each separately, so a high agreement score on a format-constrained category is never displayed as though it also confirmed the category's meaning

Counterfactual (Table 7)	Practical proposal
A definition with a creation date and a review trigger	Attach a creation date to every category and trigger review not on a fixed schedule but when a coordination failure, a downstream decision made on a misclassified or contested category, is reported against it
Categories flagged locally negotiated are routed to a person with current standing in the practice	For categories flagged locally negotiated, disable automated classification by default and route the task to a nominated person, someone whose current, active involvement in the relevant practice is documented, rather than to whichever person is available

Table 8: *From counterfactual design to practical proposals*

The distinction between the first five rows and the sixth, already established in section 5.3, carries through directly into what is being proposed. The first five are documentation and workflow changes: better metadata design closes the representation gap, the concealment, the structural openness, and the conflation of syntax with semantics, and none of them depend on resolving what any locally negotiated category actually means. The sixth is not a documentation fix at all. It is a routing decision that accepts a category's meaning may never be made accessible to an outside agent, and designs around that fact rather than against it. Implementing the first five without the sixth would improve consistency within the globally accessible layer while leaving the locally negotiated layer's risk, an LLM producing a fluent, plausible, wrong classification for a category no external definition could have been constructed for in the first place, exactly where Finding 2 found it.

6.4 Limitations

Several limitations qualify what this thesis can claim, beyond those already discussed as they arose.

The study examines a single case: one case company's classification product, one customer organization, two document classes, and ten categories tested across two runs. Findings 1 through 4 describe what was observed in this case and are not established as general properties of organizational knowledge bases, though the theoretical mechanisms Chapter 2 develops give reason to expect the same pattern elsewhere.

Confidentiality prevented the customer organization's full category set from being disclosed. The remaining category candidates and the excluded majority of the knowledge base's several hundred classes cannot be listed. This limits how fully another researcher could audit the category-selection process described in section 3.4 against the underlying data, even though the selection logic itself has been reported in full.

The human baseline against which LLM output was evaluated (section 3.5) was constructed by the researcher alone. No second rater and no inter-rater reliability measure were used. For syntactically bounded categories this is a minor concern, since agreement or disagreement follows directly from an unambiguous pattern. For semantically open categories, where the evaluation asked whether an output was a plausible interpretation of a definition rather than whether it matched a single correct answer, a single rater's judgment of plausibility is a real, if partially mitigated, limitation. Section 3.5 addressed this by recording the baseline before reviewing the corresponding LLM output, which rules out anchoring, but does not substitute for independent validation.

As established in section 5.2, this study's support for the temporal dimension is indirect. The case study design, a single comparison at one point in time, cannot on its own distinguish variation caused by temporal drift from variation that would exist even in a hypothetical organization whose vocabulary never changed. Two further data points remain unexplained rather than resolved: Document date's two disagreements in Test Run 2, and Project's shift from two output types in Test Run 1 to complete agreement in Test Run 2 (section 4.4). Both are reported as open rather than folded into an interpretation the data does not support.

Finally, the adaptation of Mingers & Willcocks (2017) framework, defended on methodological grounds in section 3.1, remains one principled reinterpretation among others that could have been built on the same justification. A different researcher, starting from the same three dimensions, might have mapped them onto the three worlds differently, or weighted the temporal dimension's relationship to the personal world differently than section 3.1 argues for here. This does not weaken the justification given; it is simply worth stating that a defended choice is not thereby the only one available.

6.5 Future Research

Step 11's call to disseminate results includes identifying further research gaps, which this section takes up directly.

Several extensions follow directly from where this thesis's evidence runs out. The most direct is a longitudinal study of the same or comparable categories, examined at two or

more separated points in organizational time, which would allow the temporal dimension to be tested against this thesis's own data rather than supported only by Hamilton et al.'s (2016) findings from general language. Second, a study conducted from inside an organization, with the standing to define and evaluate locally negotiated categories directly, would close the access barrier identified in section 5.2 and allow the theoretically most significant categories, the ones this study could not bring into testing at all, to be examined for the first time.

Third, section 2.5 proposed a compounding-dynamic hypothesis: that an LLM's output, once accepted without review, may enter organizational circulation and become embedded in the organization's official record. Later users would then encounter that record as authoritative, with no way of knowing it diverged from organizational intent at the point of production. Crossan et al.'s (1999) account of institutionalization and Walsh & Ungson's (1991) account of organizational memory, both drawn on in section 2.5, describe the general mechanisms by which individual outputs become organizational fact over time, though neither was developed with LLM-produced classifications specifically in mind. Testing whether an LLM's output follows this path, and whether it does so differently than an equivalent human classification error would, requires longitudinal access to an organization, which this thesis's single-point design does not provide.

Fourth, the practical proposal in section 6.3 has not been implemented or evaluated. Piloting the distributional flag and the routing mechanism it triggers, in an organization willing to test whether locally negotiated categories are correctly identified and correctly withheld from automated classification, would test the proposal's central claim directly: that the risk Finding 2 identifies can be designed around rather than only documented.

Finally, replication across organizations in other industries would test whether the global/local division found here, in the case company's knowledge base, is a general property of organizational vocabulary or a feature more specific to this case's particular history and sector.

REFERENCES

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations And Research Issues1,2. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Allen, R., & Choudhury, P. (Raj). (2022). Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion. *Organization Science*, *33*(1), 149–169. https://doi.org/10.1287/orsc.2021.1554

Atkin, A. (2023). Peirce's Theory of Signs. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/peirce-semiotics/

Austin, J. L. (1962). *How to Do Things with Words* (M. Sbisá & J. O. Urmson, Eds.). Clarendon Press.

Boisot, M. H. (1998). *Knowledge Assets: Securing Competitive Advantage in the Information Economy*. OUP Oxford.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Chandler, D. (2021). *Semiotics for Beginners*. http://visual-memory.co.uk/daniel/Documents/S4B/

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An Organizational Learning Framework: From Intuition to Institution. *The Academy of Management Review*, *24*(3), 522. https://doi.org/10.2307/259140

Gubelmann, R. (2024). Large Language Models, Agency, and Why Speech Acts are Beyond Them (For Now) – A Kantian-Cum-Pragmatist Case. *Philosophy & Technology*, *37*(1), 32. https://doi.org/10.1007/s13347-024-00696-1

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*

(pp. 1489–1501). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1141

Harris, Z. S. (1954). Distributional Structure. WORD, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hayek, F. A. (1945). The Use of Knowledge in Society. In *Modern Understandings of Liberty and Property*. Routledge.

Karakurt, E., & Akbulut, A. (2026). Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) for Enterprise Knowledge Management and Document Automation: A Systematic Literature Review. *Applied Sciences*, *16*(1), 368. https://doi.org/10.3390/app16010368

Kellogg, K. C., Lifshitz, H., Randazzo, S., Mollick, E., Dell'Acqua, F., McFowland, E., Candelon, F., & Lakhani, K. R. (2025). Novice risk work: How juniors coaching seniors on emerging technologies such as generative AI can lead to learning failures. *Information and Organization*, *35*(1), 100559. https://doi.org/10.1016/j.infoandorg.2025.100559

Klein, H. K., & Myers, M. D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems1. *MIS Quarterly*, *23*(1), 67–93. https://doi.org/10.2307/249410

Kripke, S. A. (1982). *Wittgenstein on rules and private language: An elementary exposition* (2. ed). Harvard Univ. Pr.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2005.11401

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. https://doi.org/10.1287/isre.12.3.240.9709

Mingers, J., & Willcocks, L. (2017). An integrative semiotic methodology for IS research. *Information and Organization*, *27*(1), 17–36. https://doi.org/10.1016/j.infoandorg.2016.12.001

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 Technical Report* (Version 6). arXiv. https://doi.org/10.48550/ARXIV.2303.08774

Orlikowski, W. J. (2002). Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science*, *13*(3), 249–273. https://doi.org/10.1287/orsc.13.3.249.2776

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Peirce, C. S. (1877). *The Fixation of Belief*. https://www.peirce.org/writings/p107.html

Peirce, C. S. (1878). *How to Make Our Ideas Clear*. https://www.peirce.org/writings/p119.html

Polanyi, M. (1967). *The Tacit Dimension*. Anchor Books.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*.

Saussure, F. de. (2011). *Course in General Linguistics*. Columbia University Press. (Original work published 1916)

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. (2009). Chinese room argument. *Scholarpedia*, *4*(8), 3100. https://doi.org/10.4249/scholarpedia.3100

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.

Singh, A., Fry, A., Perelman, A., Tart, A., Ganesh, A., El-Kishky, A., McLaughlin, A., Low, A., Ostrow, A. J., Ananthram, A., Nathan, A., Luo, A., Helyar, A., Madry, A., Efremov, A., Spyra, A., Baker-Whitcomb, A., Beutel, A., Karpenko, A., … Wang, Z. (2026). *OpenAI GPT-5 System Card*. arXiv.Org. https://arxiv.org/abs/2601.03267v2

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the roles of signs and norms in organizations—A semiotic approach to information systems design. *Behaviour & Information Technology*, *19*(1), 15–27. https://doi.org/10.1080/014492900118768

Star, S. L. (1989). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In *Distributed Artificial Intelligence* (pp. 37–54). Elsevier. https://doi.org/10.1016/B978-1-55860-092-8.50006-X

Tsoukas, H. (1996). The firm as a distributed knowledge system: A constructionist approach. *Strategic Management Journal*, *17*(S2), 11–25. https://doi.org/10.1002/smj.4250171104

Walsh, J. P., & Ungson, G. R. (1991). Organizational Memory. *The Academy of Management Review*, *16*(1), 57. https://doi.org/10.2307/258607

Wenger, E. (1998, July 28). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press. Cambridge Aspire Website. https://doi.org/10.1017/CBO9780511803932

Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, Ed.). Wiley-Blackwell.

APPENDIX A

Selected categories	Category Definition	Task	Instruction
Name	A category specifying the document's title or identifier used for naming and retrieval.	Create a name for the document.	Category definition: A category specifying the document's title or identifier used for naming and retrieval. Task: Create a name for the document.
Project	A category identifying the project associated with the document.	Identify the project associated with the document.	Category definition: A category identifying the project associated with the document. Task: Identify the project associated with the document.
Owner	A category identifying the individual responsible for the document's content, maintenance, and lifecycle management.	Identify the owner of the document.	Category definition: A category identifying the individual responsible for the document's content, maintenance, and lifecycle management. Task: Identify the owner of the document.
Document date	A category specifying the date associated with the document, typically indicating its creation, approval, or official issue date.	Identify the date of the document.	Category definition: A category specifying the date associated with the document, typically indicating its creation, approval, or official issue date. Task: Identify the date of the document.
Customer	A category identifying the client or organization for whom the document is created, delivered, or associated.	Identify the customer associated with the document.	Category definition: A category identifying the client or organization for whom the document is created, delivered, or associated. Task: Identify the customer associated with the document.

APPENDIX B

Selected categories	Category Definition	Task	Instruction
Project	A category identifying the project associated with the document.	Identify the project associated with the document.	Category definition: A category identifying the project associated with the document. Task: Identify the project associated with the document.
Project id	A category identifying the project id associated with the document. Project id format is "P/*" where "*" represents some amount of numbers.	Identify the project id associated with the document.	Category definition: A category identifying the project id associated with the document. Project id format is "P/*" where "*" represents some amount of numbers. Task: Identify the project id associated with the document.
Owner	A category identifying the individual responsible for the document's content, maintenance, and lifecycle management.	Identify the owner of the document.	Category definition: A category identifying the individual responsible for the document's content, maintenance, and lifecycle management. Task: Identify the owner of the document.
Document date	A category specifying the date associated with the document, typically indicating its creation, approval, or official issue date.	Identify the date of the document.	Category definition: A category specifying the date associated with the document, typically indicating its creation, approval, or official issue date. Task: Identify the date of the document.
Order id	A category identifying the order id associated with the document. Order id format is "Ref. No: *" where "*" represents the order id.	Identify the order id associated with the document.	Category definition: A category identifying the order id associated with the document. Order id format is "Ref. No: *" where "*" represents the order id. Task: Identify the order id associated with the document.