

Theine, Hendrik; Verita, Carlotta

Working Paper

Beyond text-as-data: Integrating structural topic modeling and critical discourse analysis for contextualized transformations research

ICAE Working Paper Series, No. 174

Provided in Cooperation with:

Johannes Kepler University Linz, Institute for Comprehensive Analysis of the Economy (ICAE)

Suggested Citation: Theine, Hendrik; Verita, Carlotta (2026) : Beyond text-as-data: Integrating structural topic modeling and critical discourse analysis for contextualized transformations research, ICAE Working Paper Series, No. 174, Johannes Kepler University Linz, Institute for Comprehensive Analysis of the Economy (ICAE), Linz

This Version is available at:

<https://hdl.handle.net/10419/339370>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Beyond Text-as-Data: Integrating Structural Topic Modeling and Critical Discourse Analysis for Contextualized Transformations Research

Hendrik Theine and Carlotta Verita

February 2026

Beyond Text-as-Data: Integrating Structural Topic Modeling and Critical Discourse Analysis for Contextualized Transformations Research

Hendrik Theine^{a,b,c}, Carlotta Verità^{a,b}

^a Socio-Ecological Transformation Lab, Johannes Kepler University Linz, Austria

^b Institute for Comprehensive Analysis of the Economy, Johannes Kepler University Linz, Austria

^c Darmstadt University of Applied Sciences

hendrik.theine@jku.at; carlotta.verita@jku.at

Abstract¹

Socio-ecological transformations are long-term reconfigurations of socio-technical and socio-economic provisioning systems, contested through struggles over pathways, timelines, and distributive consequences. Because these struggles are fought through public sense-making as well as policy and investment, transformation research increasingly turns to large text corpora. Yet a dilemma persists: computational text analysis scales but often remains descriptive, while qualitative, interpretative methods explain power and meaning in context but is typically limited to small samples. This paper develops a critical realist mixed-methods strategy that integrates Structural Topic Modeling (STM) and Critical Discourse Analysis (CDA). Drawing on critical methodological pluralism, STM is treated as an extensive mapping device that identifies candidate demi-regularities without reifying them as causal laws. CDA then supplies the intensive interpretation needed to situate these patterns, reconstruct their argumentative and ideological work, and support retroductive reasoning about generative mechanisms in open, stratified systems. Thus, by integrating those methods, we enhance both the explanatory power and critical depth of transformation research.

Keywords: Socio-Ecological Transformation, Critical Realism, Mixed Methods, Structural Topic Modelling, Critical Discourse Analysis

¹ Acknowledgements. Many thanks to Clive Spash for an instructive meeting and for generous, collegial advice and encouragement, and to Ernest Aigner for equally valuable creative brainstorming and encouragement. Thanks to Matthias Aistleitner for detailed comments and probing questions on an earlier draft, and to the SET Lab and ICAE colleagues at Johannes Kepler University Linz for ongoing discussions and support. This work benefited from feedback at the European Association for Evolutionary Political Economy conference (Athens, 2025) and at International Sustainability Transitions conference (Lisbon, 2025), as well as from presentations at the Center for Climate Change Communication, George Mason University (2024) and Leuphana University Lüneburg (2024). This research was supported by FWF (project: Sustainable Socio-economic Transition and Economic Reasoning, SETER), the Klima- und Energiefonds (project: Socio-Ecological Transformation – assessing and co-creating Media Imaginaries for a just and sustainable future, SET-MI), and 1st LIFT_C call (project: The Feasibility of Socio-Ecological Transformation: Exploring Political and Economic Constraints, F-SET).

1. Introduction

Socio-ecological transformations and processes of change are long-term, multi-dimensional reconfigurations of socio-technical, socio-economic provisioning systems (i.e. the social and technological organisation of the production and distribution of resources, goods and services) that unfold through contestation over feasible pathways, legitimate timelines, and distributive consequences (Geels, 2002; Köhler et al., 2019; Meadowcroft, 2009). These struggles are fought not only through policy, investment, and technological experimentation, but also through discourses, narratives, and imaginaries that orient expectations, define what counts as “feasible,” and help coordinate actor coalitions (Hajer, 1995; Jasanoff & Kim, 2015; Stirling, 2014). For transformation research, this makes public communication a central empirical site: what becomes salient, normalised, or unsayable in discourse is part of how socio-ecological futures are governed.

Methodologically, however, the study of transformation discourse is caught in a persistent scaling dilemma. Computational approaches can map thematic patterns across large corpora and track change across time and institutions, yet they often remain at the level of description. This carries the thread of treating textual regularities as self-explanatory and inviting “correlation-as-explanation” shortcuts. Conversely, qualitative, interpretive approaches are well suited to reconstruct meaning, power, legitimation, and silence in context, but they tend to operate on limited samples, which can make it difficult to assess broader regularities, institutional differentiation, or temporal dynamics in large, heterogeneous discursive fields.

This paper develops an integrated, critical realist mixed-methods strategy that identifies large-scale demi-regularities in the visibility, dominance, and silencing of transformation imaginaries across time and institutional contexts, and then interrogates these patterns through contextual, interpretive analysis of their ideational work and underlying conditions. From this perspective, discursive patterns are best treated as empirical traces of deeper generative mechanisms that condition which imaginaries become publicly visible and politically actionable. Building on the logic of critical methodological pluralism (Danermark et al., 2002), we propose an integrated design that combines Structural Topic Modeling (STM) and Critical Discourse Analysis (CDA). STM is used to map discursive structures and identify candidate demi-regularities across contexts (including via metadata), while CDA provides the intensive, power-sensitive interpretive work needed to interrogate those patterns, reconstruct their ideological and argumentative operations, and inform retroductive claims about underlying mechanisms within open, stratified systems.

To do so, the paper proceeds in four steps. Section 2 develops the case for contextualised transformation research by foregrounding sense-making processes and imaginaries as constitutive of socio-ecological change. Section 3 outlines the critical realist metatheoretical commitments that motivate our methodological choices. Section 4 then specifies a mixed-methods research design that integrates STM and CDA to identify discursive demi-regularities at scale and to interrogate them in context. Section 5 concludes by summarising the methodological payoffs and sketching implications for transformation scholarship.

2. Contextualized Research on Transformations: Sense-making and imaginaries

A key implication of sustainability transitions research is that socio-ecological transformation, transition, and related processes of sustainability-oriented change (hereafter: SET) are best treated as broad and contested umbrella terms with multiple meanings, rather than as single, clearly defined concepts. SET is routinely conceptualised as a plurality of pathways, each carrying distinct normative assumptions and plethora of implications. In this view, SETs are not simply technical substitution problems, but multi-dimensional political, economic, and social projects that reconfigure power relations, redistribute costs and benefits, and (aims to) reshape what today and the future looks like (Meadowcroft, 2009; Patterson et al., 2017).

A growing strand of transitions and transformation scholarship therefore stresses that SET are inherently contested and cannot be understood adequately without explicit attention to politics and power. Governance for transitions takes place under conditions of uncertainty and conflict, where actors disagree on problem definitions, causal responsibility, feasibility, urgency, and fairness (Meadowcroft, 2009; Stirling, 2014). These conflicts are not incidental but constitutive: incumbents may defend existing socio-technical regimes, challengers seek to expand alternative pathways, and states oscillate between enabling innovation and protecting competitiveness or stability. As a result, SET pathways are shaped by struggles over which visions of the future become credible, which risks are prioritised, and which solutions are rendered “realistic” or “extreme” (Köhler et al., 2019; Patterson et al., 2017).

This has a direct consequence for how SET should be studied: attention must extend beyond policy outputs and technology indicators to the interplay of material power relations and meaning-making. Put differently, future visions do not operate merely as “ideas” floating above the material world; they are stabilised and contested through discursive practices that help orient actors, justify interventions, and coordinate coalitions. From a cultural political economy perspective (CPE), semiosis, i.e. meaning-making through language, symbols, and interpretive frames, plays a constitutive role in the formation of social objects (e.g., “markets,” “innovation,” “competitiveness”) and social subjects (e.g., “responsible consumers,” “green entrepreneurs”), while remaining inseparable from material and institutional conditions and historically sedimented power relations (Jessop, 2010; Sum & Jessop, 2013). Using this perspective avoids two symmetrical reductionisms: on the one hand, reifying economic objects as natural givens and on the other, reducing social reality to actors’ interpretations alone. Instead, SET visions become transformative when semiotic practices (variation, framing, coalition-building) align with, and help consolidate, extra-semiotic structures (selection, institutionalisation, retention) that enable some pathways while constraining others (Jessop, 2004, 2010; Sum & Jessop, 2013).

Drawing on social philosophy (Taylor, 2004), cultural political economy (Sum & Jessop, 2013), science and technology studies (Jasanoff & Kim, 2015), environmental politics (Hajer, 1995), and socio-economics (Beckert & Bronk, 2018), we conceptualise these future-oriented meaning-

making systems of socio-ecological change as imaginaries, discourses, and narratives—partly distinct but overlapping lenses (in the remainder of this paper: imaginaries as shorthand)². Collectively, these approaches help capture how shared interpretations make certain futures thinkable, desirable, and “realistic,” while rendering others implausible or illegitimate. Across these perspectives, such meaning systems foreground particular social, political, and economic dimensions while backgrounding others, shaping not only preferred endpoints but also which pathways appear feasible, urgent, and governable (Pühringer & Bänderle, 2025). They are continuously produced and revised through collective sensemaking and political struggles that structure whose interpretations travel, stick, and become institutionalised (Sum & Jessop, 2013). They stabilise what counts as common sense, normalising some solution repertoires while marginalising others, coordinate, build legitimacy, and shape policy agendas (Stirling, 2014; Hajer, 1995).

Concrete examples of SET imaginaries include those that conceptualise different roles for growth and social provisioning, such as green growth, steady-state/a-growth, and degrowth; but they are not limited to this axis (Brand, 2016; Buch-Hansen & Carstensen, 2021; Kallis, 2015). Other imaginaries foreground commoning and challenges to private ownership and wealth (Federici, 2011, 2018), or climate-social policy perspectives that centre poverty, wealth, and redistribution within decarbonisation pathways (Armutskonferenz et al., 2021; Hirvilammi et al., 2023); imaginaries can also be restrictive or regressive, for instance through authoritarian visions of “order” under climate crisis or narratives that downplay climate risks and normalise delay (Machin, 2022; Lamb et al., 2020). The plurality of SET meaning-making becomes particularly visible in typological work that reconstructs competing narratives within specific domains: Hausknost et al. (2017) map bioeconomy visions into a techno-political option space (agroecology vs. industrial biotechnology; sufficiency vs. capitalist expansion), showing how policy discourse tends to cluster around growth-compatible, technology-forward narratives while sufficiency-oriented and planning-intensive alternatives remain marginal (see also Wilgosh et al., 2022 for related work on just transition pathways).

3. A Critical Realist Perspective on Methods

The features of SET scholarship discussed in the previous section sharpen two intertwined methodological challenges for transitions and transformation research. First, because imaginaries are implicated in power, coalition-building, and the institutionalisation of pathways, they cannot be treated as self-explanatory evidence; from a critical realist perspective they are better understood as empirical traces of deeper generative mechanisms, such as institutional capacities,

² Pühringer and Bänderle (2025) make these distinctions between the different concepts explicit by “stacking” key meaning-making concepts within their SET research framework. They differentiate between social imaginaries as broader, often implicit horizons that orient practice; narratives as more concrete story-form articulations that structure identity, desirability, and action; and discourses as wider patterned ensembles of statements and practices through which narratives circulate, stabilize, and can become hegemonic.

incumbent power, policy paradigms, and socio-technical dependencies, that condition which meanings become dominant and actionable (Danermark et al., 2002; Jessop, 2010). Second, transformation imaginaries are simultaneously large-scale and context-dependent: they unfold across long time horizons, crises, and multiple arenas (media, policy, science, business), so explanation requires both corpus-wide mapping of stable discourses as well as a close interpretation of how specific argumentative moves, legitimations, and silences work in situated settings (Köhler et al., 2019; Stirling, 2014).

In order to further explicate this argument, section 3 proceeds in two steps. Section 3.1 introduces critical realism's depth ontology and its critique of the epistemic fallacy, clarifying why observable patterns are not reducible to underlying reality and why reality is not reducible to what can be observed in (text) patterns. Section 3.2 then develops the methodological implication, namely critical methodological pluralism, as a rationale for combining methods so that computational mapping and interpretive analysis can be mobilised toward explanation rather than mere description.

3.1 Depth Ontology and the Epistemic Fallacy

Given the rising popularity of methodological approaches rooted in a critical realist perspective (Maisuria & Banfield, 2022), it is worth elaborating on methodological implications that arise from the particular ontological and epistemological premises of critical realism. Understanding these will be important to relate the strength of a mixed method approach using STM and CDA at alleviating the critiques and tensions critical realism points out.

Critical realism, in essence, assumes that a good explanation in research refers back to causality. Here, causality is not a linear, deterministic, or easily quantifiable quality whose effects can be generalised and extrapolated from particular observations of regularity to fit large data. Critical realism, rather, tells us to focus on the structures and mechanisms that give rise to certain events which we are interested in studying. This is inherently tied to the depth-ontology which characterises critical realism, in which the recognition that our society and any particular phenomena of interest, are embedded in a stratified, open system: for example, specific capitalist economic relations as embedded in economic provisioning systems and institutions, in turn embedded in our society, in turn embedded in the biological realm, in turn in the physical, and so on (Sayer, 1992; Bhaskar, 1978).

The implications of this understanding are threefold: i) the theory-laden character of observation, ii) fallibility of knowledge and practical adequacy as 'measure' and iii) the recognition of potentiality, or emergence. The recognition that neutrality or objectivity in the sciences, especially in social sciences, is a faulty assumption is a powerful break from certain paradigmatic approaches so far. From a critical realist perspective, observation inherently relies on certain pre-developed concepts: we implicitly relate what we are studying to things we already know, for example

through differentiation (Sayer asks us to consider how we could describe night without reference to day). Hence, observation relies on certain pre-conceived notions, and is therefore ‘conceptually saturated’. Sayer (1992) writes:

“Social scientists who treat 'data' literally as 'given things' (often those who feel most confident about the objectivity of their knowledge and the 'hardness' of their facts) therefore unknowingly take on board and reproduce the interpretations implicit in the data: they think with these hidden concepts but not about them.” (p. 52)

By treating observation as neutral, hegemony is unwittingly reproduced: it ignores the structures and power dynamics that lay behind observed phenomena, and in doing so allow these structures to continue unquestioned. In social sciences, critical realists identify the problem of the ‘double hermeneutic’, where the objects of study also have their own conceptual (context-dependent) understanding of the world around them, adding an additional layer to the conceptual saturation of the researcher. This is precisely what a political economy understanding emphasises! Naturally, the question follows: how can we generate knowledge, or assess the quality of it? Critical realists argue that there can never be a claim to have an absolute truth: knowledge is fallible, but not all knowledge is equally fallible. Practical adequacy, as Sayer (1992) defines it, has to do with the ability of knowledge to generate expectations in the world. The level of practical adequacy of knowledge will change over time, and is again context-dependent.

Closely tied to this point is the recognition of potentiality and emergence, which calls back to the depth and stratification discussed above. Knowledge concerned with the ‘why’ behind certain phenomena looks for a causal explanation: however, as already said, causality is not so easy to determine. Critical realists tie causality back to the embedded nature of all things: while individual objects of study have their own agency, this agency alone is not enough to explain why something happens (or does not happen). Essentially, all ‘objects’ have certain causal powers, which are closely related to the internal nature of these objects. However, the embeddedness of these objects in larger social relations and structures implies “that there are powers, capacities, or identifiable features of given specifiable entities that cannot be reduced to the powers and capacities of their parts” (Morgan, 2016: p. 20). The interaction of different causal properties means that some generative mechanisms, or even counter-mechanisms, will be activated under specific circumstances, thus giving rise to the phenomena in question. The contingency of these different mechanisms on each other give rise to emergent properties which are highly context-dependent. Thus, room must be left for *potential* occurrences given the interaction of these generative mechanisms and structures: in open systems, it is even more difficult to ascertain what interactions go on. In this sense, empirical regularity is not enough to draw a causal connection between observation A and B: this ignores the structural conditions that give rise to this, which may mean that under different circumstances, observation A is more regularly seen with C.

Bhaskar (1978) points out that, in ignoring these considerations, we reduce the question of what something is to the question of what we can know about it. Critical realism has so far critiqued

approaches to social science research that fall into this reductionist trap, pointing out the methodological fallacies that follow from this one-dimensional collapse of ontological and epistemological questions. In particular, critical realism criticises the notion that empirical data is neutral and the reliance on quantitative approaches to make generalisable results. This over-reliance on uncritical quantitative methods has been left unchallenged in mainstream economics in particular.

The problems with the use of quantitative methods from a critical realist perspective can be grouped according to Sayer (1992), namely: quantification, mathematical modelling, and inferential statistics. For one, it is only possible to quantify objects/phenomena that are *qualitatively invariant* (ibid., 177), meaning they can be discretely understood in different contexts without changing their essence. This is particularly difficult in the social sciences, given the context-dependent character of inquiry. Second, mathematical modelling cannot causally explain phenomena. It is unable to distinguish between causal relations and accidental ones, because it cannot account for open-system dynamics. It can represent and calculate different effects of specific actions, but cannot identify “what it is about social objects that produce certain changes” (ibid., 180). For example, nominal differences in representative variables occlude interactions between socially important characteristics: intersectionality is difficult to capture, e.g. the relationship between gender, age, class, race, etc. Thus, how can one separate what is a cause rather than a correlation? Finally, statistical modelling relies on probabilistic elements, which again says nothing about the causes of the occurrence of future events. Applying a theory to the inference of statistics must be able to say something regarding these causal mechanisms, more than just describing the relationship between variation in the dependent and independent variables, if it is to be useful.

By contrast, it would seem that qualitative methods offer a much more appropriate approach to uncovering the particularities and context-dependent contingencies that define causality in a critical realist ontology. The openness of qualitative methods leaves space for the understanding of the position of the researcher, i.e. recognition of subjectivity. They are also more adept at analysing particular case-studies in-depth, leading to more contextually grounded explanations than quantitative abstractions. However, qualitative methods applied with a different metatheoretical base can still fall into the same epistemic fallacy as their quantitative friends. For one, they are not free from the issue of generalisation. If explanations derived from the study of a specific case study are applied to larger social phenomena, this is the same reductionism and over-extrapolation that ignores the limits of what knowledge can be generated. Sayer (1992; 240) argues that while this is well understood, it has not been acknowledged in methodological applications. Furthermore, there has been criticism of hermeneutics and phenomenology as metatheoretical starting points for qualitative research. Both these positions still focus on experiences as the ‘primary object of knowledge’ (Danermark et al., 2002; Bhaskar, 1989): the aim is to understand phenomena rather than explain them, relying on perception and interpretation, which they argue is still (partially) positivistic.

3.2 Critical Methodological Pluralism

The previous critiques of quantitative research should not be taken to mean that there is no room for such methods in the social sciences. What underscores the above critiques is the lack of relation the methods have to addressing the relevance of mechanisms and structures to the empirical observation of a phenomena or event. Beyond a description of frequency or surface-level generalisability, the application of supposedly ‘objective’ quantitative approaches can never answer *why* said phenomena occurred. Instead, critical realism argues, we must approach retroductively: in order for this to have occurred, what mechanisms/structures need to have been activated (or not)? This ‘transcendental’ reasoning is the basis for defining methodological approaches that fit into a critical realist tradition; choosing methods should be based on metatheoretical considerations and reflect the ontological depth discussed above. The question needs to be, what can different approaches tell us about generative mechanisms behind the phenomena of study?

This is the backbone of what Danermark et al. (2002) refer to as critical methodological pluralism: a pragmatic combination of methods rather than a dichotomous understanding of quantitative vs qualitative approaches. They see research design as being a combination of intensive and extensive approaches: the former referring to substantive data collection and qualitative analysis, and the latter to quantitative data collection and statistics. Given the context dependence and depth needed to study causal powers of social phenomena, intensive design is very difficult to apply to large numbers of cases. However, understanding how common certain phenomena are (in different contexts) is also important to know. Drawing on Lawson (1997), Danermark et al. (2002) use the idea of demi-regularities to describe phenomena that are relatively stable over a longer time span. In critical realist terms, this implies a kind of dominance of certain generative mechanisms that are able to produce these empirical regularities in different contexts: the aim of social scientific research is to identify these and try to find explanations for them.

This, they argue, is not possible without a combination of extensive and intensive approaches. The identification of demi-regularities necessitates the use of an extensive empirical working method, especially the ability to see the same phenomena manifest under different circumstances: this allows researchers to contrast systematic differences, which provide initial ideas for generative mechanisms (ibid., 167). Simultaneously, arriving at an explanation for these is impossible without an intensive approach: without the ability to experiment as one could in a closed system, retroductive approaches from in-depth cases are fundamental. The important consideration when combining these is to understand the limits of each approach. They write:

“When, in our endeavour to explain a demi-regularity, we have reached an explanation we believe is valid, it must also be empirically grounded in demi-regularities. This is not the same thing as saying the latter part of the process is decisive for whether the explanation is valid or not. But it is a very plausible assumption that those mechanisms we

assume generate the phenomena under investigation have a certain duration in space and time. However, one must be aware of the fact that empirical manifestations are just that, demi-regularities.” (ibid., 167)

This combinatory approach is especially relevant for the study of SETs, particularly from a critical political economy perspective, as it embraces the complexity, context-dependency and emergent nature of social change. SETs are very complex, involving the actions and dynamics of heterogeneous actors, institutions and large uncertainties and path-dependencies. They are not simply questions of material shifts: they are inherently tied to discourse and discursive contestations, which are embedded in the institutional and material reality of society. Transformation is thus not only about visible shifts but also about how dominant ideologies and hegemonies are reproduced or challenged through material and discursive practices. A pluralist, critically realist orientation is therefore essential, not only for identifying patterns, but for explaining them within their broader political-economic and institutional configurations. Critical methodological pluralism allows researchers to account for the interplay between stability and change: certain generative mechanisms may produce relatively stable patterns, i.e. demi-regularities, over time, while also leaving space for potentiality and the empirically unseen manifestation of mechanisms. It also demands a reflexive awareness of the limits and possibilities of mixed-method research, ensuring that complexity is not oversimplified but interrogated through multiple, complementary lenses.

4. A mixed methods methodology to study text: Combining Structural Topic Modeling with Critical Discourse Analysis

Section 4 translates these metatheoretical commitments from section 3 into a concrete research design for studying text. We outline a mixed-method methodology that combines STM and CDA to link corpus-wide mapping with contextualized interpretation and critique.

Section 4.1 introduces STM as a scalable, transparent way to identify latent thematic structures in large corpora and to model how these structures vary with metadata and context. Section 4.2 then presents CDA as an interpretive framework for reconstructing how meaning, legitimation, and power operate within texts and discursive formations. Finally, Section 4.3 argues that integrating STM and CDA yields a complementary workflow: STM helps locate discursive regularities and shifts across contexts, while CDA provides the conceptual and analytical resources to explain how those regularities are produced, stabilised, and contested.

4.1 Structural Topic Modeling (STM)

Structural Topic Modeling (STM) is a probabilistic, unsupervised machine learning method designed to discover latent semantic patterns in large text corpora. At its core, STM builds on

Latent Dirichlet Allocation (LDA) by identifying topics as distributions of co-occurring words and assigning them in varying proportions across documents. However, STM advances this approach by incorporating document-level metadata—such as publication dates, outlets, or political affiliations—as covariates that can influence either topic prevalence (how much a topic appears in a document) or topic content (how a topic is articulated). This makes STM particularly valuable for researchers aiming to understand how thematic patterns in discourse vary across social, political, or temporal dimensions (Roberts et al., 2019).

Unlike dictionary-based or keyword-driven methods common in computational linguistics, STM is inductive, meaning that the structure of topics emerges from the data itself rather than being imposed by the researcher. It accounts for important linguistic phenomena like polysemy (the same word having different meanings depending on context) and heteroglossia (the presence of multiple voices or themes in one document) by allowing both words to belong to multiple topics and documents to contain several topics (DiMaggio et al., 2013). Moreover, STM models are evaluated using statistical indicators such as semantic coherence and exclusivity (Roberts et al., 2019). In theory, the evaluation of STM models could also benefit from qualitative validation of topics (Brookes & McEnery, 2019; Romero et al., 2024).

STM's ability to model large corpora in a structured yet flexible way has made it increasingly popular across fields such as media studies, organisational research, and political communication. However, its use is not without limitations. When applied as a standalone technique, STM can result in decontextualised or overly abstract topic clusters that risk reinforcing researchers' unexamined assumptions - particularly if the choice of covariates is not critically interrogated. The very power of STM to incorporate metadata can turn into a weakness if those variables are treated as neutral or self-explanatory, rather than socially constructed and theoretically loaded. As Aranda et al. (2021) and Schmiedel et al. (2019) caution, STM models can remain empty or indirected without critical guidance. Therefore, as we will argue in section 4.3, combining STM with interpretive approaches such as CDA is essential to explicitly surface underlying preconceptions, scrutinise covariate effects, and meaningfully connect computational results to broader ideological and discursive dynamics.

4.2 Critical Discourse Analysis (CDA)

CDA is a qualitative, theory-informed approach to studying the ways in which language is used to enact, legitimise, or challenge power relations in society. Drawing from applied linguistics, CDA encompasses a range of interconnected approaches rooted in Marxist theory, (post-)structuralism, and Critical Theory and being shaped by intellectual traditions from the Frankfurt School and the Birmingham Centre for Contemporary Cultural Studies (Forchtner & Wodak, 2017; Reynolds, 2019). Despite their differences, these approaches share a foundational view of discourse as both socially conditioned and constitutive - language not only reflects social reality, but actively

constructs and sustains it. CDA thus focuses on how ideologies, identities, and social hierarchies are embedded in and propagated through language (Fairclough, 2007; Jäger, 2015).

CDA involves both text-immanent critique (analysing the structure, lexicon, and argumentation of specific texts) and socio-diagnostic critique (linking these texts to their broader sociopolitical contexts and power structures). It operates abductively, moving between empirical detail and theoretical abstraction, and emphasises interdiscursivity and intertextuality; that is, the ways in which discourses draw on and reconfigure existing discursive formations (Fairclough, 2003). The goal is not merely description, but critique: to make visible the taken-for-granted assumptions, silences, and normative claims that structure dominant narratives. This critical dimension involves an explicit ethical and political stance on the part of the researcher, often aligned with goals of social justice and democratic transformation (Rieder & Theine, 2023; Fairclough, 2013; Pickard, 2013).

In recent years, CDA has gained prominence across disciplines, including organisational, management, media and communication studies, where it has been used to analyse legitimacy struggles, power asymmetries, and identity construction (Samdanis & Lee, 2019a, 2019b; Vaara, 2014). However, the increased digitisation and mediatisation of discourse pose new methodological challenges: scholars are now confronted with massive, complex corpora that exceed the practical scope of traditional close-reading methods. This is where computational techniques like STM become useful: not as replacements for CDA, but as tools to scale up discourse analysis while remaining grounded in critical interpretive frameworks.

Additionally, CDA is not without its own limitations. Precisely because it is grounded in the researcher's critical and often normative standpoint, it risks becoming overly selective or self-confirming. Without anchoring textual interpretation in broader patterns or systematically sampling across corpora, CDA may fall short of demonstrating how discourses function at scale. It may miss important structural variables or sociolinguistic dynamics that influence discourse formation, such as political alignment, ownership, geography, or temporality. In this sense, CDA benefits from integration with quantitative tools like STM, which can systematically surface such patterns and make the researcher's interpretive choices more transparent and accountable. Together, STM and CDA offer a complementary approach: STM structures the empirical landscape, while CDA provides the critical lens to interrogate meaning, ideology, and power.

4.3 STM + CDA: The Best of Both Worlds?

The combination of STM with CDA fits nicely to the critical methodological pluralism outlined by Danermark et al. (2002), where the strengths of each approach complement the other in a manner fitting to the depth ontology and causal-orientation of critical realism. In particular, this combination is useful for the identification and explanation of demi-regularities as explained above, for three reasons: i) both are compatible on an epistemological level due to their

understanding of power and structures, and knowledge generation; ii) their ‘emptiness’ allows for the needed exploratory breadth; iii) the limits of each approach are mitigated by the other.

At first glance, the epistemological compatibility of STM and CDA seems rather questionable. CDA is grounded in a position that recognises the importance of the social and material reality which exists outside the researcher’s interpretations while also appreciating the social construction of empirical phenomena (Fariclough, 2005). STM, rather, reflects a positivist tradition rooted in machine learning and data sciences. However, both approaches are compatible with the recognition of embeddedness and stratification of reality, and the design of STM fits the analytical process of doing discourse analysis. Of course, this compatibility hinges on the ability of a researcher to understand the limits of knowledge generation each of these has, as discussed.

While it may seem obvious, the structural element in STM is indicative of a different understanding about reality when compared with other quantitative modelling methods. It is inherently tied to the idea that empirical observations, in this case text data, are embedded into larger structures which affect their expression in different contexts. Recognising that different metadata may have an impact on the content of a corpus means recognising the power potential different generative mechanisms (e.g., ownership, social events, structure of media production) may have in reality. Assigning covariates comes with all the shortcomings outlined in critical realism’s critique of quantitative methods, and is a theory-laden selection process. Of course, the danger here would be to conflate different covariates in the metadata automatically with causal powers: however, this is not something prescribed implicitly by the method but rather depends on the reflection of the researcher. Here in particular, CDA adds an especially useful perspective to STM: the ‘critical’ perspective of the researcher, which is a fundamental part of CDA, may help circumvent the notion of objectivity often tied to quantitative modelling. This identification of the importance of structures and powers more generally fits nicely to the aim of CDA, namely to unearth how these powers embed themselves into language and continue to be reproduced. STM’s use of metadata comes from a perspective that understands text output is not neutral: the implicit potentiality of generative mechanisms manifests in different contexts, and is thus a good starting point to identify demi-regularities, as they manifest with different characteristics, and begin to ask questions about these.

Furthermore, the design of STM fits well to the analytical process of discourse analysis. CDA considers the meaning of words to be relational and open: meaning arises from the context a word is employed in, and is not an inherent feature of the word itself. STM corresponds well with this understanding of language and meaning. In STM, a ‘topic’ is “defined as a mixture over words where each word has a probability of belonging to a topic” (Roberts et al., 2013, p. 2), basically a cluster of often co-occurring words. Words have a different probability of being associated with different topics, therefore its meaning varies based on its connection to other terms. Topic modelling explicitly models ‘polysemy’ and ‘heteroglossia’ (DiMaggio et al., 2013), the idea that words can obtain multiple meanings depending on the context they are used in, and multiple topics can be contained in a single document. In a sense, it traces the multiplicity of contexts of every

word in the corpus, independent of a pre-supposed meaning associated with it outside. Interpreting the meaning of topics, then, has to do with how the frequent terms are ranked within a topic's probability distribution, how this relates to the same terms appearing in other topics, and how frequently topics occur together throughout the corpus. This also allows words to change meanings over time, as surrounding words are taken up in the analysis: in fact, Brinkmann (2019) writes that "STM was specifically developed with the idea in mind that the contents of texts (or the underlying discourse) changes over time" (ibid, p.91).

Exploring and validating different topics in another area of fruitful balancing quantitative and qualitative logics (see for a recent application to the media discourse on the climate crisis – economic inequality nexus: Verità & Theine, 2026). Within the STM package, there are different ways to validate topics in regards to semantic coherence and exclusivity, however they are insufficient without a qualitative exploration into the topics. Importantly, there are features of the modelling package that allow researchers to look into the representative texts for each topic, allowing for a deeper, contextualised, qualitative understanding of the topics (Gillies et al., 2022). In fact, Gillings and Hardie (2023) that a qualitative, close-reading approach of topics is an effective step in the overall topic modelling analysis. This shows that STM aligns well with discourse analysis' assumptions about relationality and openness, and in particular CDA's ideas about intertextuality and interdiscursivity (Reisigl & Wodak, 2009).

Importantly, STM deals with three levels of probability: one, the probability of terms to a topic; two, the probability of topics to a document; and three, the expected probability of a topic in relation to different metadata. However, without qualitative interpretation these probabilities remain indistinguishable. They are not enough to make causal statements, but can showcase relationships that are worth looking into qualitatively: the critical interpretation behind topics and metadata covariates is needed to give it explanatory power. And, in a first step, the choosing of metadata variables of interest should be rooted in a retroductive approach to studying the phenomena in question. STM is thus extremely adept at acting as an exploratory approach in a larger landscape of text data than would be available with CDA alone, allowing a researcher to see how certain characteristics differ across multiple contexts. Identification of regularity is thus more reliable than in case-specific studies. With the combination of CDA, this exploratory approach can be guided and built upon to make these explorations more than just a descriptive standardisation of the data. In particular, this is extremely useful in understanding the role of significant silences, i.e. meaningful omissions in a text that nevertheless shapes interpretation and can do ideological work (Huckin, 2002, 2009), something STM alone is incapable of doing. Thus, an intensive approach, grounded also in a deep understanding of the particular question being studied, is fundamental to properly assessing modelling output. This fits precisely to the benefits Danermark et al. (2002) describe regarding the use of both intensive and extensive approaches to studying demi-regularities. Rouhana (2023) showcases how CDA guided topic-modelling can serve to analyse a large data set of text data, and allow for a more granular approach by re-running the method on smaller samples of data by year.

This role of interpretation unites these despite their distinctive traditions: both STM and CDA are ‘empty’ methods without the direction given by researchers. As already established, STM is an unsupervised, inductive modelling approach that requires interpretation at different steps (Hannigan et al., 2019): the agency of the researcher is intertwined with the agency of the automated process, meaning that choices made by the researcher will invariably affect the outcome of the modelling, as will the choices of automation process (Aranda et al., 2021). CDA similarly relies on the critical perspective the researcher takes, it does not start off with a fixed theoretical stance (Rouhana, 2023): again, this could lead to issues regarding over-extrapolation of results if case studies are not understood as specific manifestations of a phenomena rather than a whole picture. Understanding these approaches through the perspective of critical methodological pluralism showcases how they complement each other in the search for explanations behind demi-regularities. In the context of discourse research (discursive phenomena), this allows topics to be understood as ‘semantic macropositions’ (Brinkmann, 2019) that require further contextualisation and analysis to be understood historically: the advantage of modelling is to highlight topical structures that are interesting to look into on a deeper level.

5. Conclusion: Toward a Contextualized Understanding of Socio-Ecological Transformations

This paper has proposed and substantiated the value of combining Structural Topic Modeling (STM) and Critical Discourse Analysis (CDA) within a critical realist research framework. While both methods originate from different epistemological traditions, we have shown that they are epistemologically and methodologically compatible when viewed through the lens of critical methodological pluralism (Danermark et al., 2002). STM offers a systematic way to identify semantic patterns and thematic structures across large corpora, including the ability to model the influence of social and institutional metadata. CDA, in turn, ensures that such patterns are not merely described but interrogated, situated, and understood in terms of their relation to broader systems of power, ideology, and material embeddedness. Their integration therefore helps to avoid both the descriptive empiricism often associated with computational methods and the limited generalisability that can result from qualitative case-specific research.

Critically, this methodological combination responds to the urgent conceptual and empirical challenges of transformation research. Socio-ecological transformation are not linear, deterministic, or purely material shifts. They are deeply imaginary and discursive processes, shaped by hegemonic narratives, contested framings, and the (re)production of power relations across scales. Analysing transformation than requires tools that can identify regularities in the visibility, dominance, and silencing of different imaginaries over time and across institutions, while also remaining attentive to their contextual contingencies and ideational underpinnings. STM is uniquely suited to trace such demi-regularities—empirical patterns that indicate the relative stability or transformation of underlying generative mechanisms. CDA, in turn, makes it

possible to explore whether these patterns reflect continuity, rupture, appropriation, or ideological re-articulation. As the challenges of climate crisis, inequality, and systemic reorganisation intensify, so too must our methods be capable of grasping the complexity of these transitions - not only in what they are, but in how they are narrated, legitimised, and resisted.

Literature

- Aranda, A. M., Sele, K., Etchanchu, H., Guyt, J. Y., & Vaara, E. (2021). From Big Data to Rich Theory: Integrating Critical Discourse Analysis with Structural Topic Modeling. *European Management Review*, 18(3), 197–214. <https://doi.org/10.1111/emre.12452>
- Armutskonferenz, Attac, & BEIGEWUM (Eds.). (2021). *Klimasoziale Politik: Eine gerechte und emissionsfreie Gesellschaft gestalten*. Bahoe Books.
- Beckert, J., & Bronk, R. (Eds.). (2018). *Uncertain futures: Imaginaries, narratives, and calculation in the economy*. Oxford University Press.
- Bhaskar, R. (1989). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. Hassocks: Harvester Press.
- Bhaskar, R. (1978). *A realist theory of science*. Hassocks, Sussex : Harvester Press ; Atlantic Highlands, N.J. : Humanities Press.
- Brand, U. (2016). Green economy, green capitalism and the imperial mode of living: Limits to a prominent strategy, contours of a possible new capitalist formation. *Fudan Journal of the Humanities and Social Sciences*, 9(1), 107–121. <https://doi.org/10.1007/s40647-015-0095-6>
- Brinkmann, F. (2019). Topical Discourse Structures: Using topic modeling in discourse analysis approaches. *Human IT: Journal for Information Technology Studies as a Human Science*, 14(3), Article 3.
- Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 21(1), 3–21. <https://doi.org/10.1177/1461445618814032>
- Buch-Hansen, H., & Carstensen, M. B. (2021). Paradigms and the political economy of ecopolitical projects: Green growth and degrowth compared. *Competition & Change*, 25(3–4), 308–327. <https://doi.org/10.1177/1024529420987528>
- Chen, Y., Peng ,Zhao, Kim ,Sei-Hill, & and Choi, C. W. (2023). What We Can Do and Cannot Do with Topic Modeling: A Systematic Review. *Communication Methods and Measures*, 17(2), 111–130. <https://doi.org/10.1080/19312458.2023.2167965>
- Danermark, B., Ekström, M., & Karlsson, J. Ch. (2002). *Explaining Society: Critical Realism in the Social Sciences*. Routledge. <https://doi.org/10.4324/9781351017831>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41(6), 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge.
- Fairclough, N. (2007). Critical discourse analysis as a method in social scientific research. In R. Wodak & M. Meyer (Eds.), *Methods in critical discourse analysis* (pp. 121–138). Sage.
- Fairclough, N. (2013). Critical discourse analysis. In M. Handford & J. P. Gee (Eds.), *The Routledge handbook of discourse analysis* (pp. 9–20). Routledge.

- Federici, S. (2011). Women, land struggles, and the reconstruction of the commons. *WorkingUSA: The Journal of Labor and Society*, 14(1), 41–56. <https://doi.org/10.1111/j.1743-4580.2010.00319.x>
- Federici, S. (2018). *Re-enchanting the world: Feminism and the politics of the commons*. PM Press.
- Forchtner, B., & Wodak, R. (2017). Critical discourse studies: A critical approach to the study of language and communication. In *The Routledge handbook of language and politics* (pp. 135–150). Routledge.
- Gillies, M., Murthy, D., Brenton, H., & Olaniyan, R. (2022). *Theme and Topic: How Qualitative Research and Topic Modeling Can Be Brought Together* (arXiv:2210.00707). arXiv. <https://doi.org/10.48550/arXiv.2210.00707>
- Gillings, M., & Hardie, A. (2023). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities*, 38(2), 530–543. <https://doi.org/10.1093/lc/fqac075>
- Hajer, M. A. (1995). *The politics of environmental discourse: Ecological modernization and the policy process*. Oxford University Press.
- Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic Modeling in Management Research: Rendering New Theory from Textual Data. *Academy of Management Annals*, 13(2), 586–632. <https://doi.org/10.5465/annals.2017.0099>
- Hausknost, D., Schriebl, E., Lauk, C., & Kalt, G. (2017). A transition to which bioeconomy? An exploration of diverging techno-political choices. *Sustainability*, 9(4), 669. <https://doi.org/10.3390/su9040669>
- Hirvilammi, T., Häikiö, L., Johansson, H., Koch, M., & Perkiö, J. (2023). Social policy in a climate emergency context: Towards an ecosocial research agenda. *Journal of Social Policy*. Advance online publication. <https://doi.org/10.1017/S0047279422000721>
- Huckin, T. (2002). Textual silence and the discourse of homelessness. *Discourse & Society*, 13(3), 347-372.
- Huckin, T. (2009). On textual silences, large and small. In Bazerman, C., Krut, R., Lunsford, K., McLeod, S., Null, S., Rogers, P., & Stansell, A. (Eds.), *Traditions of writing research* (pp. 431-443). Routledge.
- Jäger, S. (2015). *Kritische Diskursanalyse: Eine Einführung* (6th ed.). Unrast Verlag.
- Jessop, B. (2004). Critical semiotic analysis and cultural political economy. *Critical Discourse Studies*, 1(2), 159–174. <https://doi.org/10.1080/17405900410001674506>
- Jessop, B. (2010). Cultural political economy and critical policy studies. *Critical Policy Studies*, 3(3–4), 336–356. <https://doi.org/10.1080/19460171003619741>
- Kallis, G., Demaria, F., & D’Alisa, G. (2015). Degrowth. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 24–30). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.91041-9>

- Köhler, J., Geels, F. W., Kern, F., Markard, J., Wieczorek, A., Alkemade, F., Avelino, F., Bergek, A., Boons, F., Fünfschilling, L., Hess, D., Holtz, G., Hyysalo, S., Jenkins, K., Kivimaa, P., Martiskainen, M., McMeekin, A., Mühlemeier, M. S., ... Wells, P. (2019). An agenda for sustainability transitions research: State of the art and future directions. *Environmental Innovation and Societal Transitions*, 31, 1–32. <https://doi.org/10.1016/j.eist.2019.01.004>
- Lamb, W. F., et al. (2020). *Discourses of climate delay*. *Global Sustainability*, 3, e17. <https://doi.org/10.1017/sus.2020.13>
- Lawson, T. (2004). *Economics and reality* (Reprinted). Routledge.
- Machin, A. (2022). *Climates of democracy: Skeptical, rational, and radical imaginaries*. *WIREs Climate Change*, 13, e774. <https://doi.org/10.1002/wcc.774>
- Maisuria, A., & Banfield, G. (Eds.). (2022). *Working with critical realism: Stories of methodological encounters*. Taylor & Francis.
- Meadowcroft, J. (2009). What about the politics? Sustainable development, transition management, and long term energy transitions. *Policy Sciences*, 42(4), 323–340. <https://doi.org/10.1007/s11077-009-9097-z>
- Morgan, J. (2016). Critical realism as a social ontology for economics. In *Handbook of Research Methods and Applications in Heterodox Economics* (pp. 15–34). Edward Elgar Publishing. <https://www.elgaronline.com/edcollchap/edcoll/9781782548454/9781782548454.00007.xml>
- Patterson, J., Schulz, K., Vervoort, J., et al. (2015). *Transformations towards sustainability: Emerging approaches, critical reflections, and a research agenda* (Earth System Governance Working Paper No. 33). Earth System Governance Project.
- Pickard, V. (2013). Being Critical: Contesting Power within the Misinformation Society. *Communication and Critical/Cultural Studies*, 10(2–3), 306–311. <https://doi.org/10.1080/14791420.2013.812590>
- Pühringer, S., & Bäuerle, L. (2025). *The economics of Socio-Ecological Transformations: A conceptual framework* (No. 171). ICAE Working Paper Series.
- Reisigl, M., & Wodak, R. (2009). The discourse-historical approach. In R. Wodak & M. Meyer (Eds.), *Methods of Critical Discourse Analysis* (pp. 87–121). Sage Publications Ltd.
- Reynolds, C. (2019). Building theory from media ideology: Coding for power in journalistic discourse. *Journal of Communication Inquiry*, 43(1), 47–69.
- Rieder, M., & Theine, H. (2023). Breaking Down the Discourse, Exposing Power in Economic Journalism: Critical Discourse Analysis. In Silke, H., Quinn, F., & Rieder, M. (Eds.), *How to Read Economic News* (pp. 192–225). Routledge.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). **stm**: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2). <https://doi.org/10.18637/jss.v091.i02>

- Romero, J. D., Feijoo-Garcia, M. A., Nanda, G., Newell, B., & Magana, A. J. (2024). Evaluating the Performance of Topic Modeling Techniques with Human Validation to Support Qualitative Analysis. *Big Data and Cognitive Computing*, 8(10), 132. <https://doi.org/10.3390/bdcc8100132>
- Rouhana, T. (2023). Critical discourse analysis guided topic modeling: The case of Al-Jazeera Arabic. *Information, Communication & Society*, 26(5), 904–922. <https://doi.org/10.1080/1369118X.2023.2166364>
- Samdanis, M., & Lee, S. H. (2019a). Access Inequalities in the Artistic Labour Market in the UK: A Critical Discourse Analysis of Precariousness, Entrepreneurialism and Voluntarism. *European Management Review*, 16(4), 887–907. <https://doi.org/10.1111/emre.12154>
- Samdanis, M., & Lee, S. H. (2019b). Uncertainty, strategic sensemaking and organisational failure in the art market: What went wrong with LVMH's investment in Phillips auctioneers? *Journal of Business Research*, 98, 475–488. <https://doi.org/10.1016/j.jbusres.2018.08.030>
- Sayer, A. (1992). *Method in Social Science: Revised 2nd Edition*. Routledge.
- Schmiedel, T., Müller, O., & Vom Brocke, J. (2019). Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture. *Organizational Research Methods*, 22(4), 941–968. <https://doi.org/10.1177/1094428118773858>
- Stirling, A. (2014). Transforming power: Social science and the politics of energy choices. *Energy Research & Social Science*, 1, 83–95. <https://doi.org/10.1016/j.erss.2014.02.001>
- Sum, N.-L., & Jessop, B. (2013). *Towards a cultural political economy: Putting culture in its place in political economy*. Edward Elgar Publishing.
- Taylor, C. (2004). *Modern social imaginaries*. Duke University Press.
- Vaara, E. (2014). Struggles Over Legitimacy in the Eurozone Crisis: Discursive legitimation Strategies and Their Ideological Underpinnings. *Discourse & Society*, 25(4), 500–518.
- Verità, C., & Theine, H. (2026). Exploring the Climate Crisis: Economic Inequality Nexus in the US American Media Landscape. *The Political Economy of Communication*, 12(1), 20-50.
- Wilgosh, B., Sorman, A. H., & Bàrcena, I. (2022). When two movements collide: Learning from labour and environmental struggles for future just transitions. *Global Environmental Change*, 73, 102452. <https://doi.org/10.1016/j.gloenvcha.2022.102452>