Multilevel Discourse Analysis
A Structured Approach to Analyzing Longitudinal Data

Steven J. Kahl and Stine Grodal

Introduction

In order to address the growing theoretical interest in temporally unfolding organizational phenomena like the emergence of new organizational forms, the practice and process of organizing, formation of occupational identities, and organizational change, we need better methods for studying longitudinal data. Qualitative analysis of organizational phenomena has historically relied on analysis of interviews and ethnographic observations. Due to the necessity of the researcher to be present during the time of data collection, the temporal periods studied using these methods are restricted.

Recently, a theoretical push to understand how events unfold over time, coupled with the availability of large online repositories of texts and images, has made qualitative analysis of longitudinal data more prevalent (Ventresca and Mohr, 2002). These data sources include documents that people in and around organizations produce as part of their tasks and communication, such as meeting notes, email exchanges, formalized reports, interviews, and verbal exchanges.

Collectively, data from interviews, ethnographic observations, and archival data is called discourse. Discourse is part of the social fabric within and around organizations in which individuals sense-making, organizations act, processes evolve, concepts develop, and identities form. Discourse can be produced with intervention by the researcher (as in interviews and ethnographic observations) or without such intervention (as in archival research). The primary unit of analysis for analyzing discourse is texts. Within discourse analysis, the term text is used broadly to refer "not just to written transcriptions but to any kind of symbolic expression requiring a physical medium and permitting of permanent storage" like objects and images (Phillip, Lawrence, and Hardy, 2004, p. 636).

However, analyzing large quantities of longitudinal data sources challenges our existing methods regarding both data selection and data analysis. First, studying longitudinal data often necessitates drawing on diverse data sources, which challenges our traditional approaches to data analysis that have tended to rely on linguistic texts only. Second, when analyzing texts longitudinally, it becomes important not just to understand the meaning of the text itself, but also the temporal relationship among texts. For example, Heracleous and Barrett (2001) analyze the relationship between different enunciatives (that is arguments in use) in order to study how the implementation of a new
trading platform in the London Insurance Market enabled organizational change. Third, because the institutional context in which texts are produced is likely to change over time, it also becomes more pertinent to have a theory that is able to capture such contextual changes. Last, when tracking discourse over time we need a method that allows us to record even subtle changes in meaning between texts.

To measure changes in meaning over time requires a structured approach to coding texts and images. This is a challenge because traditionally textual coding has used interpretive approaches that lack precision (Franzosi, 2010). Indeed, if we are not able to address these challenges, the analysis that we engage in is likely to be distorted because it does not account for the temporal patterns that are at the core of conducting longitudinal analysis.

In this chapter, we argue that to address the mentioned challenges, we need a structured contextualized approach to analyzing longitudinal data. In particular, we advocate for the use of multilevel discourse analysis, which decomposes text into its semantic relationship and uncovers textual relationships. At its core, discourse analysis involves analysis of collections of texts, the ways they are made meaningful through their links to other texts, the ways in which they draw on different discourses, how and to whom they are disseminated, the methods of their production, and the manner in which they are received and consumed. (Phillips, Lawrence, and Hardy, 2004, p. 636)

Discourse analysis treats text as a communicative act such that language does not just express what people say, but also helps create their understandings, beliefs, and processes (Phillips and Hardy, 2002; Phillips and Oswick, 2012). Discourse scholars use the term texts broadly as communicative expressions that come in different forms, including written documents, verbal expressions, visual representation, and physical design (Hargadon and Douglas, 2001; Phillips and Hardy, 2002).

The first (and often challenging) step in doing multilevel discourse analysis of longitudinal data is to generate a nonbiased sample. To develop such a sample, we advocate using the historical method (see Kipping, Wadhwa, and Bucheli, 2014) to select texts, organize them temporally, and develop an initial understanding of the domain area. We call this first step historical reconstruction. The next steps include a structured contextualized approach to analyze the texts in the sample. We build upon Fairclough (1992; 1995) and argue that texts ought to be analyzed at three different levels: intratextual, intertextual, and contextual. At the intratextual level, researchers code both the content and linguistic structure of the text, such as not only the noun phrases of the text, but also linguistic structures like semantic clauses (subject/verb/object combinations) and sentence types. For example, Bingham and Kahl (2013) coded noun/verb combinations within insurance texts to analyze the evolution of different analogies used for the computer. To bring structure to the intratextual coding—which helps in detailed tracking of meaning over time in longitudinal data—we build upon Franzosi (2010) and advocate a coding process based upon grammatical and semantic rules. This approach contrasts to traditional coding methods, which rely on interpretive techniques, such as grounded theory (see Glaser and Strauss, 1967).

Whereas the intratextual level of analysis focuses on the language used within the text, the intertextual level addresses the purpose and process of producing the text, the relationship between texts, as well as the audiences’ interpretations and understandings of the text. This analysis examines how the languages of producers and audiences mutually react to each other and the context of the exchange that influences its production and reception (Fairclough, 1992). Intertextual relations include both the content that links texts together and how those relations between texts are created. The intertextual level stresses the importance of understanding the social processes that created these linguistic expressions in the first place by capturing the production and consumption of texts. These processes can be traced through the direct dialogue or references between texts.
Last, discourse itself takes place within a broader institutional and cultural context that may extend beyond the organization. The contextual level, therefore, investigates how discourse is embedded within the institutional and social practices. Part of this analysis entails measuring the uptake of conventional norms within the focal discourse (see Khaire and Wadhwa, 2010). This level entails understanding the significance of the discourse in terms of the patterns of exchange within the broader institutional and cultural context.

Instead of coding all levels simultaneously (see Alvesson and Karreman, 2000; Phillips and Oswick, 2012), we propose to follow Barry, Carroll, and Hansen (2006), who emphasize that each level should initially be coded separately and then be integrated as a last step in the analysis, which we call iteration and theory development. This allows researchers to use the different levels to triangulate between the data, identify inconsistencies, and use the process to link the analysis back to the original theoretical question.

Figure 32.1 provides an overview of the five steps in what we term multilevel discourse analysis: (1) historical reconstruction; (2) intratextual analysis, (3) intertextual analysis, (4) contextual analysis, and (5) iteration and theory development (see Kahl and Grodal, forthcoming) for more detail). We elaborate each step next.

**Step 1: Historical Reconstruction**

We propose that the first step in the multilevel discourse analysis should use historical methods to reconstruct the broader discourse in which the phenomenon of interest exists (see Wodak, 2001; Khaire and Wadhwa, 2010; and Kipping, Wadhwa, and Bucheli, 2014, for similar approaches). Because many organizational phenomena take place in heterogeneous environments, historical reconstruction facilitates capturing texts from different participants, organizing them in the proper temporal order, and addressing potential biases.
Based on the researcher's general theoretical question, they should use the historical method to identify texts, who produced them, the location of production and consumption, as well as the contemporaneous cultural themes that were occurring at the time that the texts were produced. These nontextual factors help develop the broader context in which each text is embedded. Moreover, the historical method yields a more comprehensive data set because it does not privilege a single point of view. For example, if a report has several authors, the historical method encourages tracing prior texts from each of the authors as a means to understand the focal text. Last, the historical method guards against potential sample bias by questioning whether the gathered text is representative (Golder, 2000). That is not to say that unique texts are not included in the sample, but this method encourages identifying other texts to corroborate the focal text. In fact, validating representativeness helps identify those texts and points of view that are less prevalent and developed.

This process culminates with a historical timeline of the texts, people involved, events, and locations. Finally, as a last element in historical reconstruction, researchers should read through the texts as they unfolded. This open-ended reading helps develop a sense of how the discourse progressed over time, as well as identify aspects of the discourse that warrant more structured coding. This last part of structural coding, thus, bears resemblance to the "open coding" used in grounded theory building (see Glaser and Strauss, 1967).

**Steps 2 Through 4: Structured Analyses at Multiple Levels**

Whereas historical reconstruction focuses on organizing the data by participants, audiences, time, and interactions, structured analysis focuses on the detailed and systematic coding of the collected texts. The goal of these analyses is thus to track the phenomena of interest over time with a great degree of precision.

A general issue identified with textual coding is to systematically capture the same kind of data across different modes of text. For example, data may be comprised of an image as well as written text (see Chapter 23 in this volume regarding the use of visual data). Because an adequate identification of the phenomena of interest necessitates understanding the relation between different forms of text, the researcher must take care in making sure that images and text are coded in such a way that allows comparability. Consider an image of the organization chart in a text. We may treat this image holistically and code its broad image. Features of the image—in this case reporting relationships—can be captured at different levels of coding. Finally, structured analysis requires detailed analysis of texts, which can be cumbersome; however, computers can help automate this process through the use of sophisticated linguistic analysis algorithms (Phillips and Hardy, 2002). For example, Kennedy, Chok, and Liu (2012) use automated linguistic analysis to track the evolution of what it means to be "green" over time. The advantage of structured analysis of texts at the multiple levels is three-fold: (1) to generate tables, graphs, and other visuals, which can help communicate and illustrate the phenomena of interest to the reader; (2) accountability in the form of a precise tool to understand the temporal unfolding; and (3) the generation of new insights and challenges to existing understandings. As we detail below, structured analysis requires coding at each intratextual, intertextual, and contextual level of analysis.

**Step 2: Intratextual Analysis**

At the intratextual level, researchers code each text's discursive elements (e.g., analogies, imperative sentences, or certain semantic relationships between nouns and verbs) in a systematic way. Currently, researchers have used a wide variety of discursive approaches to code text (Van Dijk, 1997), including rhetoric (Hermans, 2006; Silbince, Jarzabkowski, and Shaw, 2012; Suddaby and Greenwood, 2005), narrative analysis (Lounsbury and Glynn, 2001; Wry et al., 2011), hermeneutics (Hermans, 2006; Khair and Wadhawani, 2010; Phillips and Brown, 1993), and metaphorical analysis (Bingham and like Heracleous more grounded).

One issue with discursive structures is that they are often abstract and require detailed analysis. For example, researchers might analyze how discourse is constructed and how it relates to other discourses. This can be done using multiple ways of expressing discursive forms.

Another issue is that these discursive structures often occur in a context that is, however, meaningful of the text.

To address grammatical and syntactic structures, researchers can use methods such as the modula

verb/Object
as an example of the declarative: 
the sequence or

It is important to note that these approaches are

interpretative understandings and level textual level can be

approach in the more specific levels of how
torical analogy, for example, involves e:

(1) content

- (2) context

- (3) form
historical method to
m, as well as the con-
verge produced. These
added. Moreover, the
lack a single point of
vestment. Methodological
apers (Golder;
this method encour-
representativeness helps

ed, events, and loc-
ad through the texts
discourse progression
t coding. This fits
in grounded theory

audiences, time, and
the collected texts,
with a great degree

same kind of data
d well as written text
ite identification of
k forms of text, the
y that allows commu-
image holistically—
can be captured
texts, which can
use of sophisticated
ok, and Liu (2012)
“green” over time.
to generate tables,
t, text (Van Dijck,
12; Suddaby and
1), hermeneutics
phorical analysis
(Bingham and Kahl, 2013; Etzioni and Ferraro, 2010). Some studies combine discursive approaches, like Heracleous’s (2006) combination of hermeneutics with rhetorical analysis. Other studies use more grounded theory approaches to broadly coding the content of text (Glaser and Strauss, 1967).

One issue with selecting a specific discursive approach is that it only codes one element of the text’s discursive structure. For example, while Bingham and Kahl’s (2013) metaphorical analysis identified noun/verb combinations, it did not capture the text’s rhetoric, such as the use of tropes or argument structures. However, many research questions often require using multiple textual analysis techniques. For example, studies of how an organization’s identity evolves over time necessitate analyzing texts
using multiple different discursive approaches because organizational members might use a variety of ways of expressing their views on the organization’s identity whether in objects, images, or different discursive forms.

Another issue is that texts are comprised of different layers (e.g., clauses, sentences, paragraphs) and that these discursive elements occur at different layers within the text. For instance, think of a manage-
gment journal article. It includes clauses, sentences, paragraphs, and sections. Argument structures
often occur at higher levels of the text, which involve combinations of sentences and paragraphs,
whereas others, such as metaphors, operate within sentence clauses. Therefore, a more systematic
coding of the texts requires addressing how different discursive expressions occur at these multiple layers.

To address this problem, we advocate for a more flexible and structured coding approach based
on grammatical and semantic coding. At the clause level, Franzosi (2010) suggests coding semantic
triplets or, more generally, the subject/verb/object combinations. For example, take the prepositional
sentence, Franzosi is the subject, suggests is the verb, and coding semantic triplet subjects/verb/object
combinations are the objects. In addition, the researcher can code aspects of the triplet, such
as the modality of the verb or its type (see Halliday, 1994, for one kind of classification of verbs),
pronouns and their referents, and modifiers. At the sentence level, researchers can code its type, such
as declarative versus imperative, as well as the complexity (clause combinations). At the paragraph level,
the sequence of sentences and their relations, for example, the use of conjunctions such as therefore
or however can be coded.

It is important to recognize the differences between this grammatical/semantic approach and the
discursive approaches like rhetorical or hermeneutic analysis that are currently in use within organi-
zational research. As Franzosi (2010) points out, starting with a particular discursive approach requires
interpretation of the text on behalf of the researcher. In contrast, our structured approach requires an
understanding of the grammatical and semantic rules. As such, it is much easier to validate coding pro-
cedures and minimize error. As an additional validation step, one should be able to reconstruct higher
level textual layers based on the coding of the lower layers; for example, the coding of clauses at a lower
level can be recombined to identify the argument structure at a higher level. In addition, this coding
approach captures the core semantic structure of the texts and is flexible enough to provide data for
the more specific discursive approaches. For example, coding the semantic relations of a sentence’s core
clauses also isolates words and phrases, which could be used in hermeneutical analysis. Or, the coding
pronouns, types of verbs, their modality, types of sentences, and their sequence can all be used in rhe-
torical analysis. While we recognize that coding schemes cannot be exhaustive and may vary depend-
ing upon the research question, we believe that coding texts following grammatical rules and semantic
relations provides a more structured way to code the texts that is also more flexible for analysis.

Step 3: Intertextual Analysis

As mentioned earlier, one of the core challenges when engaging in longitudinal analysis is to under-
stand how different texts are related in their production and consumption. Intertextual analysis
involves capturing these relations between texts. Texts can related to each other in at least three ways:
(1) conceptually; (2) personally; or (3) spatially.
Analyzing conceptual relations entails identifying whether texts share the same language and structure (Khaire & Wadhwani, 2010). These relations can be studied by comparing the intratextual coding of each text and identifying the occurrence of similar themes across texts. However, links between texts can also be created because the same people produced or consumed them. For example, if two texts are authored by the same person, these two texts will have an intertextual tie. Texts can also be linked because they were produced in the same location. For example, several texts might be produced at the same team meeting, so these texts have a spatial link. Moreover, since we are measuring the discourse over time, the directionality of relationships is important. Finally, not all texts agree with each other; some may respond negatively or even reject the points of previously made texts. Therefore, it is also important to capture the valence of the relationship.

We have found that network-mapping tools greatly facilitate developing and analyzing the intertextual relations. This requires formulating what constitutes a tie between texts and building out the relational matrix that assists in analyzing the structure and relations between the texts.

**Step 4: Contextual Analysis**

Contextual analysis involves identifying the presences of broader cultural themes within the text. Empirically, this first requires identifying themes that are present in the broader culture at the time that the texts were produced. Sometimes this will entail collecting an additional data set from which these broader themes can be identified. Second, after cultural themes have been identified, intertextual analysis can be used to identify connections (conceptual, personal, or spatial) between the broader corpus of texts from the general culture and the corpus of texts of interest. Recognizing that cultural themes can enter through the mode of production and consumption of the texts aids in identifying dissenting and less dominant viewpoints. Situating the texts of interest within the broader culture is important when conducting longitudinal analysis because the cultural context differs in various historical periods and thus exerts differentiated influences at different points in time.

**Step 5: Iteration and Theory Development**

The last step in multilevel discourse analysis is iteration and juxtaposition between the various steps in the analysis, which culminates in new theory development. An important part of this process is to identify inconsistencies in the coding at all the different levels. For example, by engaging in intertextual analysis, researchers might realize that they have overlooked some important texts during historical reconstruction. Or contextual analysis might reveal the importance of coding for other linguistic elements during the intratextual coding. This comparative process can also resolve coding errors because it helps transmute coding between the different levels. Consequently, iteration between the different levels can reopen analysis at each of the steps and serve to resolve inconsistencies in the coding between the levels.

Another important part of iteration is to begin the process of theory development. During the iterative process, researchers should circle back to their original research question to use the empirical material to develop new insights about the phenomena in question. Iteration also serves to test alternative explanations to the emerging theory. Indeed, during steps 1 through 4, researchers ought to create alternative explanations to emerging interpretations that can be compared to other parts of the data. Consequently, the structured contextualized approach to multilevel discourse analysis is an iterative process meant to manage potential interpretive bias, deepen the analysis, and develop novel theory.

**Application of Multilevel Discourse Analysis**

Few current studies go through all of the five steps outlined in Figure 32.1. However, each of the studies listed in Table 32.1 incorporate some of the elements of the process and serve as examples
same language and paring the intratextual texts. However, consumed them. For e an intertextual tie, sample, several texts. Moreover, since we portant. Finally, not points of previously hip, analyzing the intertext building out the texts.

<p>| Table 32.1 Examples of Articles Using Elements of Multilevel Discourse Analysis |
|------------------------|-------------------------------|------------------------|------------------------|</p>
<table>
<thead>
<tr>
<th>Reference</th>
<th>Research Context</th>
<th>How the Innovation Was Used</th>
<th>Outcomes/Results of Innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barley (1986)</td>
<td>The introduction of CT scanners into two hospitals</td>
<td>Detailed structural coding of ethnographic data collected from June 1982 to June 1983</td>
<td>Ability to track the interaction between structure and agency over time</td>
</tr>
<tr>
<td>Bingham &amp; Kahl (2013)</td>
<td>The emergence of the computer; used archival historical data</td>
<td>Tracked the emergence of the schema for the computer from 1945-1975</td>
<td>Ability to identify the specific mechanisms through which the schema emerges over time</td>
</tr>
<tr>
<td>Khaire &amp; Wadhwani (2010)</td>
<td>Creating a market for modern Indian art</td>
<td>Used discourse analysis to track the changing meaning of modern Indian art from 1995 to 2007</td>
<td>Allowed the authors to compare the relationship between the creation of meaning around modern Indian art and the evaluation of this work</td>
</tr>
<tr>
<td>Navis &amp; Glynn (2010)</td>
<td>The emerging satellite radio industry</td>
<td>Tracked the change in reference categories from mid-1990s to 2005</td>
<td>Allowed the authors to identify a shift during industry emergence from focusing on legitimating the category to differentiating among competitors</td>
</tr>
<tr>
<td>Dunn &amp; Jones (2010)</td>
<td>Medical education</td>
<td>Used a structured approach to identify changes in the care and science logic from 1910–2005</td>
<td>Authors were able to associate changes in the care and science logic to the interest of different communities</td>
</tr>
<tr>
<td>Maguire &amp; Hardy (2009)</td>
<td>Environmentalism and the use of DDT</td>
<td>Tracked discourses related to DDT from 1962–1972</td>
<td>Identified how participants engaged in defensive institutional work by authoring texts</td>
</tr>
<tr>
<td>Nelson &amp; Irwin (forthcoming)</td>
<td>Librarians occupational identity and relationship to Internet search</td>
<td>Investigated the role of librarians’ attitude toward search in their professional journal over the 22-year period 1980 through 2010</td>
<td>Show how occupational identity conditions the interpretations of technology, while also showing how these interpretations might change over time</td>
</tr>
</tbody>
</table>

Various steps in the process is to identify intertextual analysis, storical reconstructive elements during use it helps triangulate levels can reopen the levels. During the use the empirical uses to text alternatives ought to create parts of the data analysis is an iterative xp novel theory.

of how multilevel discourse analysis might be carried out in practice. For example, Khaire and Wadhwani (2010) used historically informed processes to identify the distinct stages of the institutionalization of Indian art, hermeneutics to examine the influence of the context and broader themes on this process, and tracked the development of key concepts through intra- and intertextual analysis.

Conclusion

In this chapter, we highlight a structured contextualized approach to multilevel discourse analysis as a novel method for analyzing large volumes of temporally unfolding data. In particular, we suggest that such data need to be analyzed through a five-step process: (1) historical reconstruction, (2) intratextual analysis, (3) intertextual analysis, (4) contextual analysis, and (5) iteration and theory development. Multilevel discourse analysis improves upon existing research methods in its ability to both ground the analysis within an evolving context and systematically track the phenomenon of interest over time. In so doing, multilevel discourse analysis allows for the creation of graphs and
visuals, which illustrates the progression of the phenomenon of interest over time. It also strikes a balance between the rich contextualization of traditional qualitative research and quantitative analysis.

Note

1 This method is based on work done in Kahl and Grodal forthcoming.

References


