Beyond Close Reading: An Empirical Approach for Annotation and Classification of Multimodal Texts

Asen O. Ivanov & Kenzie Burchell

CSDH 2020, June 1-5
Congress 2020

In this talk, I will present a method for the analysis of multimodal texts, Kenzie and I developed. Specifically, I will tell you more about the research context in which we developed the method, the theoretical ideas behind it, and how we plan to use it.

But beyond that, by presenting this work, we are also hoping that we can engage the CSDH community in a dialogue about how to develop robust and generative methods and tools for **the *distant reading* of multimodal texts**.

[slide- 2]

The method I will discuss today was conceived and developed within the broader research framework of Kenzie's project ***Exploring the limits of media power***. In this project, Kenzie comparatively examines media production practices and news coverage of the Syrian war.

In this context, we developed the method to facilitate the multimodal discourse analysis of online political communication, and specifically, the analysis of two genres of news coverage of the Syrian war—namely, online news articles and news items.

By focusing on the comparative analysis of news coverage of the Syrian war our broader theoretical goal was to understand how what we might call the *paralinguistic* features of multimodal texts—such as their layout, composition, image-text relations, and interactive navigational mechanisms—contribute additional layers of meaning, interpterion, and affect to the reporting of news.

In other words, our goal was to use the method in order to describe and analyze the ***strategic narratives*** advanced across news coverage of the Syrian war. And to do so, by paying attention not only to what is being said—i.e., the facts reported on the news—but also on how it is being said through **the orchestration of a range of communicative resources** within multimodal texts**.**

[slide- 3]

To introduce you to this work, in the remainder of this presentation, I will briefly discuss the broader field of **multimodal discourse analysis** and the specific ideas within this literature with which we are working. I will then give you an overview of the method we developed by focusing on three key aspects. Lastly, I will conclude the presentation by identifying some questions we are currently thinking about.

[slide- 4]

Multimodal discourse analysis is an integrative approach for analyzing the **communicative potential** of multimodal texts (i.e., texts that combine a range of linguistic, visual, and design resources—or **modes**, as they are called in this literature).

The overall objective of multimodal discourse analysis is to examine how every empirically observable element, or mode, contributes to the overall communicative potential of multimodal texts, thus, adding additional layers of meaning.

Multimodal discourse analysis has been taken up in a variety of disciplines that routinely deal with the analysis of multimodal texts, including media and communication studies, linguistics and education, to name a few, as well as other fields which deal with the production of multimodal texts such, for example, as human-computer interaction and design.

So in that sense, multimodal discourse analysis is a truly interdisciplinary research project. But the main contributions we are drawing on in our work, come from the fields of *social semiotics* and *corpus linguistics*.

While these two fields are substantively different, they share common roots in that they both build on the epistemological and ontological assumptions of *systemic-functional linguistics*. As such, they see communication as a process constituted by a set of **communicative resources** that are selectively used by people—in our case by both newsmakers and news audiences—situated in a historical, social, and cultural context. The idea, thus, is that by paying attention to how these **communicative resources** are used and articulated, we can gain insights into the nature of communication and meaning-making.

[slide- 5]

More specifically, in developing our method, we consulted both the social semiotics and corpus linguistics literature on multimodal discourse analysis. And while both perspectives influenced our work, the corpus linguistics literature had a much stronger impact on us.

The reason is that the corpus linguistics literature has developed more rigorous techniques that enable the systematic *transcription, annotation, and classification* of multimodal texts—what merging the concepts of John Unsworth and Franco Moretti we propose to call **the scholarly primitives of distant reading**. The idea being that if you want to engage in distant reading, you need means for doing *transcription, annotation, and classification*.

Key contributions we drew on in this literature are the work of John Bateman, a linguist at the University of Bremen, who in a 2008 book introduced the so-called **Genre and Multimodality Model** (aka. The GeM Model). We also drew on the work of Ognyan Seizov, a former student of Bateman who has applied the GeM model to the analysis of political communication.

By virtue of applying the GeM model to the analysis of political communication, the work of Seizov was instrumental to our work, and we drew heavily on it.

[slide- 6]

So next, let me tell you more about the method we developed and how it is meant to work.

The analytical process begins with the transcription of what Bateman calls the **virtual canvas** of a multimodal text. The idea here is that different multimodal texts offer a different range of communicative resources that can be articulated differently. So for example, the canvas of a printed page can offer text, images, and layout as communicative resources, whereas an online 'page' can offer text, images, layout as well as navigation mechanisms (e.g., hyperlinks and thumbnails). Put simply, an online page has more means to communicate than a printed page.

The communicative resources available on a specific virtual canvas vary based on its complexity. The example on the slide identifies the components of the highly complex canvas of classroom interaction.

[slide- 7]

Furthermore, Bateman distinguishes between what he calls the **physical** and the ***virtual canvas*** to highlight the idea that multimodal communication is not constrained only by the material affordances of a given media, but also by the established socio-cultural conventions and genres of articulating different communicative resources in specific ways.

A good example that illustrates this idea is how news screen graphics are used differently in different national news cultures, as shown on the slide for China, Hong Kong, Taiwan, and Japan.

The concepts of the physical and virtual canvas, thus, provide the starting point for the analysis, but importantly they also control for analytical errors in attributing significance to aspects of the multimodal texts that may result simply due to the physical constraints of specific canvas or the social and generic conventions of using it.

[slide- 8]

Drawing on these ideas, this is the template we developed to transcribe the virtual canvas of online news articles.

Simply put, the goal of this template is to provide a structured way to number and describe every communicative resource available on an online page—including title, sub-titles, images, textboxes, and anything else we could potentially find there.
[slide- 9]

The technique we developed to transcribe news items is similar.

The goal here is to create a table that captures the duration of each visual segment in a news item and to then classify those segments based on production conventions—such as, for example, archival material, a segment narrated by a news anchor in the studio, a raw feed, and so on.

The benefit of developing these two types of transcriptions is that you now have a structured data to work with. In other words, rather than the raw data—in our case online articles and news items—you now have a clear outline of the structure and content of the material you are working with, which in turn allows you to begin classifying its features.

[slide- 10]

The classification schema we developed to this end is based on the logic underpinning Bateman's GeM model. This approach recommends that multimodal texts are broken down into individual *layers*, which are then classified separately.

In total, so far, in our method, we have five layers and 23 properties.

On the slide, you can see two of the five layers in our classification, and their associated properties. In turn, each of those properties can take a range of predefined values.

To facilitate the consistent application of this classification, we also developed a **data tables** and **data dictionary**, and **use manuals.** Let me show you how these documents look like.

[My sincere apologies. I had a tech hiccup. At min 13, I'm under the impression that I'm sharing my screen and showing you a few documents. However, I was only sharing my PowerPoint slides, and hence you don't see the documents on the screen. I have recorded a second short video to show you these documents. Please watch the second video when you get to min 13. Then jump to min 16 in my presentation. My apologies again.]

[slide- 11]

Having transcribed and coded the material in this way provides a rich set of structured data that can be analyzed in numerous ways.

We can, for example, start identifying empirical patterns such as that one news outlet illustrates news articles about the Syrian war by publishing photographs of objects (weapons or infrastructure), whereas another news outlet illustrates news articles about the Syrian war by publishing photographs of people (troops or civilians).

Such analytical approach is useful and could be quite revealing. For example, it allows us to conclude that one news outlet uses photographs to humanize the war (by publishing photos of people), and the other news outlet uses photographs to dehumanize the war (by publishing photos of objects).

But beyond such empirical observations, we also wanted to develop a way to analyze the narrative structure of the materials we were work with.

To do this, we adopted a model of narrative developed by Labov during the 1970s. We specifically liked Labov's narrative model as it maps out quite well on the narrative structure of news reporting, as it is taught in journalism schools across the world—as you can see in the table on the slide.

So, by using this narrative model, we can say that elements # 1-Title, 2-Subtitle, 3-Photo1, and 5-Textbox1, fulfil the narrative function of what Labov calls ***Abstract*** *and* ***Orientation***—i.e., collectively, they define the story by providing a summary of the topic and introducing its settings and context.

We can thus conclude that the article defines the story by stating the facts (via the title and subtitle), appealing to emotions (via the photo showing the devastation caused by the war), and providing some preliminary context (via the textbox). Such patterns emerge fairly quickly after you have transcribed and classified a multimodal text.

[slide- 12]

Lastly, as mentioned earlier, our method follows closely the work of Bateman and Seizov from whom we take our key ideas and analytical techniques. But we also integrated social semiotic analytical techniques, particularly to facilitate the analysis of moving images (such as news items).

The corpus linguistic literature offers ideas about that (including a book by Bateman), we found the social semiotics literature to be better suited to the context of out project. Specifically, form social semiotics, we took a range of concepts and techniques for the classification of the relationship between image and text; and the sequencing, rhythm, and transitions of temporal and moving image media.

[slide – 13]

So, this is all, I have prepared for today. But to conclude, I will leave you with a few questions, which I hope we could discuss during the Q&A during the conference.

Specifically, we are at present trying to figure out how we can automate if not all, at least part of the transcription and coding steps our method entails. This research, for now, involves

mostly spending time on websites such as the Programming Historian in an effort to understand how we can use some computational tools to support the use of our method.

Furthermore, as I explained, we created this method for media and communication research, but we are also interested in finding other areas of application. One such area is DH pedagogy. In this regard, we are specifically inspired by work developed at the MIT's HyperStudio for Digital Humanities and specifically their concept of annotation as a pedagogical technique. We have already done a little bit of work in this direction, which I can tell you more about and show you during the conference Q&A.

Lastly, we are also wondering if we should be applying this method to random samples of data, or if we should be adopting principles and techniques for corpora design as developed in corpus linguistics.