The capacity for language is a quintessentially human one, though its ubiquity means that it is sometimes difficult to theorize. We can never move beyond language in order to study it from outside itself, but meta-linguistic study helps us uncover the ways in which the most dubious of human traits have been embedded in this, our most important adaptation. A series of theoretical building blocks, from Wittgenstein to Saussure, from Derrida to Foucault, help us look at language as it is built and as it functions in society. Pragmatically, we know that meaning is stable in some way, for we are able to use language to communicate with one another, but by conceptualizing the relational nature of language—that being the way that meaning is generated through the syntagmatic and paradigmatic relationship between signs—we can see that meaning is more elusive than grammatical arrangement suggests.

Language is an emulation of the natural world rather than a straightforward referent to it, and examining the significance of particular terms throughout history demonstrates how meaning shifts over time as certain forms of telling become (un)sanctioned. This brief exercise in deconstruction is essential to lay the foundation for critical discourse analysis (CDA). At the base of CDA, we recognize that “knowledge [and its linguistic expression] is not a matter of getting an accurate picture of reality, but of learning how to contend with the world in the pursuit of our various purposes” (Barker and Galasiński, 2001: 3). By analyzing
language in use, we can see how some groups are more successful in their pursuits than others based on the ability to participate in desired social constructions. Individual utterances either contribute to or resist structures of power that are constantly being built, reproduced, and defended, even as they are dismantled, re-mixed, and attacked. CdA seeks out these sites of contention and examines the solid ground asserted when accumulated power is able to fix meaning in a constructed (and often exploitative) truth. From these temporary centers, analysts can “hope to transcend our acculturation” by finding linguistic “splits which supply toe-holds for new initiatives” (Rorty quoted in Barker and Galasiñski, 2001: 20). For CdA researchers, language holds the potential to both restrict and liberate its users.

Historically CdA research has been limited by modes of publication, with texts being produced within (and sometimes in service of) the exploitative systems that CdA hopes to challenge. The advent of the World Wide Web has created an expanding corpus of living language that speaks in the voices of the corporation and the factory worker, the voice of the bureaucrat, the voice of the power broker, offering the CdA researcher more economical access to the voice of the every-person than has been available in the past. Oddly, CdA has been slow to apply itself to this collection for a variety of reasons. Following Rebecca Rogers et al.’s (2005) explication of the terms “critical” “discourse” “analysis,” I problematize CdA’s engagement with the digital world, and add to arguments first put forth by Gerlinde Mautner (2005) calling for consideration of digital texts. Finally, I point to contemporary developments that are ripe for input from critical discourse analysts, as the power structure of the web begins to shift from emancipation to indoctrination.

**DISCOURSE**

At the heart of critical discourse analysis, we have discourse. Broadly defined we can think of discourse as a “communicative event” (van Dijk, 2001: 98), that is, an interaction meant to share information or ideas. With this definition we are able to analyze “conversational interaction, written text, as well as associated gestures, facework, typographical layout, images, and any other ‘semiotic’ or multimedia dimension of signification” (van Dijk, 2001: 98). Any instance where one entity is using a sign or symbol to communicate with another we can claim a form of discourse is taking place. In order to create a manageable research subject, CdA researchers identify varying levels of discourse to uncover multiple layers of meaning.
One of the significant CDA scholars, James Gee, breaks discourse into two factions, what he calls “little d’ discourse and big D’ Discourse” (Gee, 2004: 10). Little d discourse refers to the bits of language, signs, or symbols that make up a communicative event. It is “the syntax of sentences and formal relations between clauses or sentences in sequences: ordering, primacy, pronominal relations, active-passive voice, nominalizations and a host of other formal properties of sentences and sequences” (van Dijk, 2001: 107). Despite the name, little d discourse is a complex set of relations within the text that give keys to understanding. Little d discourse pays close attention to the detailed structure of the language and how that structure functions in the text. As this is the foundational level of language, CDA is not complete without this type of analysis.

Teun van Dijk begins an analysis of a text by focusing first on what he calls the “global meaning” (van Dijk, 2001: 102) of little d discourse. This is the semantic macrostructure of the text, that is, what the text is about. It is the aspect of the text that is easiest to recall, and can be outlined by looking at topic sentences or general impressions from specific sections of a text. From there van Dijk moves on to what he calls the “local meaning” (van Dijk, 2001: 103) of the text. This is when we begin to look at specific word choices in order to understand what is being explicitly said about the topic. A couple of important aspects to consider are pronoun use—how the text constructs “us” and “them”—and hyperboles: the degree to which bias is evident in the way both sides are described. Both of these sites reveal the extent to which there is positive presentation of the self and negative presentation of the other. Van Dijk also asks what is missing from the text. Examining cases of active/passive voice may show missing or hidden agents, lexical choices and metaphors emphasize polarization and downplay exclusions or eliminate negative aspects. Throughout the investigation of little d discourse, the researcher examines the text in its own unity—as an entity in and of itself—and seeks to uncover the secrets hidden in its linguistic structure.

Little d discourse becomes particularly problematic in the digital world, where publishing platforms and text are dynamic rather than static entities. For example, the website where a text is published determines the extent to which there are language bits that are not applicable to the text under study. Websites that feature external advertising will have ads that change not only with each visit to the site but often within a specific time frame. While this type of outside encroachment on the text may be easy to dismiss, other websites feature user-generated content, including links to previous posts, personally endorsed ads, and links to whole texts via recommended reading and blog lists. Looking
more closely at the text in question, we approach the challenge of hypertext: bits of language that the author purposefully imbues with content that lies outside of the specific utterance. In some cases, the author is making familiar moves, such as citing or referencing direct quotations as well as directing the audience to relevant resources that contribute to the theoretical foundation. In this instance, the hypertext can assist in interpretation but it is not required. However, in other instances, authors will use hypertext to link to salient aspects of discourse that have been published in other locations. Without the information contained in the hypertext, the audience cannot fully understand the text at hand. This unique aspect of digital text confirms that while little d discourse is important, analysts must also examine context. This brings us to what Gee calls “big D” Discourse.

Big D Discourse situates discourse within its myriad relationships. It takes into account not only what is being said and how, but evaluates who is speaking to whom, when the discourse is taking place, and the purpose. Big D Discourse argues that any text “need[s] a historical, cultural, socio-economic, philosophical, logical, or neurological approach, depending on what one wants to know” (van Dijk, 2001: 97). Van Dijk’s global and local distinction is also useful in specifying the parameters of big D Discourse.

Global contexts of big D Discourse situate the text historically, politically, socially, and culturally. When we investigate the global context of an utterance, we are identifying circumstances that both the transmitter and receiver take for granted and are not necessarily mentioned in the text. Having a context allows language users to make selections of pertinent information and to omit superfluous or inapplicable details. This does more than just situate the text within its historical, cultural, socioeconomic, and philosophical context. When we identify the global context, and begin to reveal the positions of the transmitter and receiver of the text within larger society, we can begin to make inroads into local context, where the mental models of the participants begin to play a role.

The local context considers the interactional situation. Where is this exchange taking place and in what form? Is the text itself a command, argument, request, or action (say, a piece of legislation)? Who is participating and what social role does each individual play in relation to the topic and each other? What knowledge, intention, goals, norms, and beliefs do the participants have? All of these factors contribute to the ways in which a text is constructed, delivered, and received, and so inform meaning. For example, Gee discusses the example of the sentence “the cat broke” (2004: 21) in order to demonstrate how
the signifier “cat” can have multiple meanings based on the context. In this case, the specific contextual meaning is a statue of a cat.

This local level of context is the crux of critical discourse analysis, for this is where the questions of agency and hegemony arise—two of the central concerns of the term “critical.” Van Dijk argues that there is no direct link between discourse and society, and so there must be a necessary detour through the mental models of the participants. These models contain markers of socially situated identity, including ways of thinking and feeling, beliefs and solidarity with particular social groups, ways of acting and interacting, and value systems that influence how the recipient will understand what is being related (van Dijk, 2001). Gee says,

> how people say (or write) things (i.e., form) helps constitute what they are doing (i.e., function). In turn, what they are saying (or writing) helps constitute who they are being at a given time and place within a given set of social practices (i.e., their socially-situated identities). Finally, who they are being at a given time and place within a given set of social practices produces and reproduces, moment by moment, our social, political, cultural and institutional worlds. (Gee, 2004: 48)

This observation is key to understanding discourse. While on the surface any language is, in its own terms, as linguistically sound as any other, over time we come to understand that vernacular choices can yield very different results. Each time we assert a certain identity through discourse we are re-enacting social norms that have provided our blueprint for the desired identity. In this way, “discourse moves back and forth between reflecting and constructing the social world [. . . and] cannot be considered neutral” (Rogers et al., 2005: 369).

In traditional texts, multiple levels of editing, including assistance from others, are used to help us make sure we are properly negotiating the language in order to achieve the desired results. Even in oral presentations we are careful to consider our audience in order to adopt the appropriate discourse, including dress and facial expressions. This care and attention will be compounded if we are being recorded. As we prepare to present ourselves in the digital world, however, we come up against the disorienting feeling of what Michael Wesch calls “context collapse” (2009: 22). Multiple layers of access mean we are not always sure who will be a part of our audience. We must choose an appropriate discourse to present ourselves to the “generalized generalized other” (24)—a discourse for all possible contexts. Further, this discourse is direct in that it is
usually unmediated by editors, and the speed and ease of publication means that we are sometimes able to bypass some of our own internal editors and publicize an utterance charged by emotion or out of character. On the one hand, this question of context collapse makes it difficult for the CDA researcher to distinguish global and local context and account for the mental models that may be pertinent to understanding. On the other hand, it can provide a more intimate view of the speaker or author, a view that has been emphasized in CDA in relation to oral situations, such as research in classrooms.

We can see that CDA is not simply concerned with the grammatical structure and word choice of a text, nor is it limited to the social, cultural, and political context an utterance speaks of and from. CDA research also examines the ways in which individuals understand, appropriate, and reuse discourse in order to produce, maintain, or enact identity.

ANALYSIS

With such a broad scope, the analysis of discourse requires practitioners to account carefully for the aspects of the context, both local and global, which are pertinent to an examination of the communicative event in question. This is known as the “frame problem” (Gee, 2004: 32) and is one of the main sites of criticism in CDA. There is no explicit theory of context, and in order to counter accusations of mining data for examples supporting their arguments, CDA researchers must be explicit in both their analytical frames as well as their justifications for data choice. Another prominent scholar in the CDA field, Norman Fairclough (1995), has developed one of the more comprehensive frameworks for analysis. His method grounds analysis first in text, then in interaction or discourse practice, and finally in society or sociocultural practice (Sheyholislami, 2001).

In this method, the first level of analysis is textual; that is, we are looking at little d discourse, and from it we are gleaning linguistic evidence that helps us construct a frame of context. This linguistic evidence includes the ideas being discussed, the construction of identity of the writer and reader, and the way this relationship is constructed. These factors aid in understanding the second level, the interaction level, which is slightly more interpretive. Similar to van Dijk’s discussion of mental models, Fairclough positions interaction as the dynamic space between text and society (Sheyholislami, 2001). This asks how participants receive and interpret texts and subsequently either reproduce or transform those texts. In this space we can begin to analyze traces of intertextuality,
or how the text appropriates “snatches of other texts, which may be explicitly demarcated or merged in, and which the text may assimilate, contradict, ironically echo, and so forth” (Sheyholislami, 2001: 8).

Interactional elements become embedded in text, either through direct quotation or by mirroring conventions of discourse, providing linguistic attributes that point analysts toward viable frames at the social level. Finally, at the social level a multidisciplinary approach is used to explore the complex structures of power present in a particular discourse. Fairclough looks at the media type and evaluates access. In traditional media, there is often an overrepresentation of powerful groups, leaving many voices unheard. Considering the politics of the media context uncovers the relationship to the state, an elite class, a major corporation, or other vested interest. The cultural aspect analyzes social values as well as the norms of production and consumption. Finally, “the economics of an institution is an important determinant of its practices and its texts” (Fairclough quoted in Sheyholislami, 2001: 10), with content selection and bias toward economic interests being obvious consequences.

All these levels of analysis means that the CDA researcher must not only account for the choice of text, but must defend the various frames of meaning that establish the context. The frame problem is a possible reason that CDA researchers have been slow to engage with the Internet; “principled criteria for choosing what should go into the corpus need to be developed and applied” (Mautner, 2005: 815), but “there is no such thing as an explicit theory of context” (van Dijk, 2001: 108). Indeed, web-based utterances even remove markers of corpus that have been previously used in data selection. In traditional forms of media, “demarcation lines between texts are easily drawn and provide clear guidance for corpus compilation. Most hypertext, by contrast, is ‘borderless,’ not only with respect to beginnings and ends but also, through clickable links, at the ‘sides’” (Mautner, 2005: 819). Further, as mentioned in the discussion of Big D Discourse, the issue of context collapse makes the evaluation of the interactional and even societal level of Internet discourse problematic. As Gee states, the socially situated identities of individuals are of importance to CDA researchers, and, unless we consider the web as a world of its own rather than a reflection of our analogue existence (which would require a further level of inquiry), “the given time and place” and “given set of social practices” (Gee, 2004: 48) that frame a digital discourse are not always evident.

Considering the frame problem in the digital world also hints at a deeper tension within CDA in general. As researchers must carefully account for these
levels of analysis, it is recommended that web-based researchers freeze content in order to maintain a hold on their research subject. This process denies the emergent nature of knowledge and removes authorial agency, as the researcher moves the text from an unstable to a fixed form in order to use it for tests of verifiability. These analytic practices encourage researchers to exercise power over the text and the author, a position inconsistent with the overarching stance of CDA: the critical perspective.

**Critical Perspective**

When CDA freezes content, it is demonstrating the ways language selection produces structures of authority. Concern for structures of power and authority is what distinguishes CDA from other forms of linguistic analysis. As van Dijk notes, “unlike much other scholarship, CDA does not deny but explicitly defines and defends its own sociopolitical position. That is, CDA is biased—and proud of it” (van Dijk, 2001: 96). CDA is not interested in simply describing the linguistic aspects of the text; it seeks to explain them in terms of social relations and social structures. Further, it aligns itself with critical theory, following from the Frankfurt School, and positions analysis in the service of uncovering conditions of inequality. Critical theory is not a unified approach, as such, but a common thread is its overarching concern with power and justice. From under this umbrella, critical theory shows great diversity in approaches to critiques of power. Considering racial, neocolonial, feminist, and queer theories, among others, critical theorists seek to identify the many forms power assumes: “ideological, physical, linguistic, material, psychological, [and] cultural” (Rogers et al., 2005: 368). By deconstructing or problematizing power structures, critical theorists are able to argue that facts are not neutral and are cultivated within and in service to established hierarchies. One feature that crosses most approaches to critical theory is Antonio Gramsci’s notion of internalized hegemony, or the tendency of oppressed groups to willfully adopt their subjugation through a process of coercion and consent.

Explorations of present trends in the online world beg for input from CDA scholars in order to keep the Internet “considerably less prejudiced in favour of élites” (Mautner, 2005: 816). In 2005, Mautner described one of the difficulties in doing CDA online:
in the absence of gate-keepers, who structure and vet content in the traditional media, the onus falls on the researcher to establish the nature of the data that search engines have laid before him or her, and to select those sources that will be useful in answering specific research questions. (Mautner, 2005: 817)

While the frame question became more problematic on the web, from a critical perspective, the openness and accessibility of the Internet meant that an entire host of voices that were hitherto difficult or impossible to include in discourse analysis could be accessed in a “huge repository of authentic data” (Mautner, 2005: 809). Over time, however, the Internet has lost some of its impartial flavour, as argued by Eli Pariser (2011) in his 2011 TED talk. “There’s this kind of shift in how information is flowing online, and it’s invisible,” he says, pointing to the growing personalization of the web by large service providers such as Google, Facebook, Yahoo, The Huffington Post, The New York Times and others. Almost in conversation with Mautner, Pariser outlines how

in a broadcast society there were these gatekeepers, the editors, and they controlled the flows of information. And along came the Internet and it swept them out of the way and it allowed all of us to connect together and it was awesome. But that’s not actually what’s happening right now. What we’re seeing is a passing of the torch from human gatekeepers to algorithmic ones [. . . and they] don’t have the embedded ethics the editors did. (Pariser, 2011: n.p.)

Using data frames that could be found in a CDA paper—namely, geographic origin, time of publication, and mode of access (if you are not logged in to your search engine account), and authorship, gender, age, and detailed patterns of consumption (if you are logged in to your search engine account)—search engines now produce customized results to any query. It is “very hard for people to watch or consume something that has not, in some sense, been tailored for them” (Eric Schmidt of Google quoted in Pariser, 2011: n.p.). Of course the first question that a critical analyst asks is, “Tailored by whom?” With the search filter Pariser describes, individuals “don’t actually see what gets edited out” (Pariser, 2011: n.p.) of their searches. This allows for systems of power to determine access to web-based discourses based on algorithms that are not publicly scrutinized. The potential for continued patterns of exploitation are ripe if the “Internet is showing us what it thinks we want to see” (Pariser, 2011: n.p., my emphasis).
Pariser’s sentiments have led me to question my own experience using the search page Blackle, a low-energy alternative powered by Google. While I have a Google account, I am not usually signed in, limiting the personal information Google can access to flesh out issues of relevance. Using the most basic search page, I still find curious manipulations of text that produce spurious search results. Van Dijk outlines useful rhetorical devices to recognize and consider overt manipulative tactics promoting particular mental models in the audience (van Dijk, 2006). I will employ his model as I consider power discrepancies promoted by Google’s search queries.

At the level of text, I have noticed while searching my own digital and musical identity, onepercentyellow (one word), that Blackle will automatically “correct” my search criteria, modifying my search to one percent yellow (three words). This yields dramatically different results, as my unique web presence is subordinated to the grammatical markers of spaces and separate words. Below the search box a question appears: “Did you mean onepercentyellow?” In addition, the page titles and snippets generated for a search result may also be manipulated. According to Google’s Webmaster Central Blog, sometimes “a single title might not be the best one to show for all queries, and so [Google has] algorithms that generate alternative titles to make it easier for . . . users to recognize relevant pages” (Far, 2012: n.p.). Between tailoring of results and manipulation of user and content text, Google may be presenting a search result that is “incomplete or lack[s] relevant knowledge—so that no counter arguments can be formulated against false, incomplete, or biased assertions” (van Dijk, 2006: 375).

At the level of interaction, personalized results can mean that past constructions of the self determine the world that will be reflected back to the user. This means that rather than exploring the varying horizons of existence, we are interacting with ourselves more and more, increasing a sense of comfort in a world that is as we would predict. It is difficult to comprehend the multiple ways that Google gathers information that could be used to generate personalized results, and many are unaware that this is occurring at all. This covert manipulation of access to information means that “fundamental norms, values and ideologies” that go into the algorithms tailoring results “cannot be denied or ignored” by users (van Dijk, 2006: 375).

At a social level, access to search tailoring is restricted to Google itself (or whatever site you are considering); questions of economic and political affiliation are valid concerns. The longstanding existence of the most popular search engines, along with the highly specialized knowledge required to understand
the process, “induce[s] people into tending to accept the discourses and arguments of elite persons and organizations” (van Dijk, 2006: 375). These elite persons control access to the largest repository of knowledge in human history.

Addressing van Dijk’s final aspect of manipulative rhetoric, focusing on the eliciting of emotional response, I was struck by what I will call Google’s performance rating on each result. What is the purpose of telling me that my search delivered 2.5 million results in .22 seconds? By demonstrating the vastness of the Web at every search, “emotions . . . that make people vulnerable” (van Dijk, 2006: 375) are drawn into each encounter with the digital world. If my search alone has generated more than a million results, how could I negotiate the vastness of the entire web without the assistance of Google or another search engine?

POSITIVITY

Of course there are many aspects of the web that contribute significantly to CDA. While CDA researchers acknowledge the inequity in access to technology, the advent of Web 2.0 or the Social Web has reduced the technical skills required to participate in digital discourse. This “access is egalitarian [. . . and] Web content is not subject to the ordering and standardizing influence of institutions and the professionals active within them” (Mautner, 2005: 817). Take, for example, Sheryl Prentice’s (2010) examination of Scottish independence. By combining historical discourse, corpus linguistic technique, and CDA, Prentice conducted a keyness comparison between British newspapers and a pro-independence website in order to determine correlations and divergences between the language of popular Scottish opinion and sanctioned British voices. Before the spread of web-based textual discussion forums, a study of this kind, which boasts access to over three million words from the every-person, could have never taken place without extensive and costly interviews or questionnaires.

The “size of the Web creates an embarras de richesses” (Mautner, 2005: 815), but extraction, tagging, and computer-assisted analysis software has become user-friendly, addressing the technical requirements that made the first versions “beyond the majority of CDA researchers” (Mautner, 2005: 816). Prentice used automated semantic tagging software in place of by-hand tagging done by the researcher, making the analysis of the corpus possible. Prentice argues, “automated semantic tagging can be used to lend reliability to the tagging process, which in turn lends reliability to one’s findings” (Prentice, 2010: 431). As
reliability and verifiability have been disputed in CDA, research in the field is bolstered by appeal to a more quantitative gathering and tagging of data.

In contrast with traditional print media, the web allows a current glimpse into any research subject. As “the medium is so dynamic and flexible, it reacts with unprecedented speed and precision to social change” (Mautner, 2005: 821), allowing for web-based CDA to examine minute shifts in social power relations that may be omitted from other media. In another study, Nelya Koteyko found the Web to be a “useful resource of words and phrases that are too rare to appear in any standard purposefully-built corpora” (2010: 655). In this study, more than 80,000 RSS feeds were examined in order to conduct a concordance examination of compounds involving “carbon.” Not only did this provide a diverse list of word compounds, the “variety of data representing different social domains” (Koteyko, 2010: 658) points to the web as a space to gather a more representative sample than could be created under analog circumstances.

CONCLUSION

CDA puts under a microscope the most important of all human innovations: language. It aspires to do this in the explicit service of uncovering and combating systems of exploitation that are present in our linguistic structures and reflective of larger social and global power systems. CDA examines how discourse moves from a local, personal communicative event into constituting social and global structures of power. The link is not direct, and instead occurs first through cognition—the individual gets it (from somewhere) and then interaction—the individual spreads it to someone else. If we are to understand the dimensions of control and authority in the present time, we are remiss if we fail to look to the Internet as a crucial space for CDA engagement. From the use of the web as a repository of authentic data to the examination of the structures emerging on the web, CDA has a major role to play in this developing medium. Further, the ephemeral quality of the web makes it a site on which we can revise our understanding of knowledge. If we can see that “knowing is a matter of being able to participate centrally in practice and learning is a matter of changing patterns of participation (with concomitant changes in identity)” (Rogers, 2004: 12), we may be able to use CDA’s understanding of authority to address the continued development of exploitative power structures in the digital world.
REFERENCES


