

1. The Scale of Work in Technical Communication

“Are we there yet?” is a common refrain heard during car trips and discussions about the state of technical communication. Technical communication is no different from fields around the academy that must consistently evaluate their state of the art and state of the practice. Yet the unique contours of technical communication’s history and practice make the exercise a fraught one. The establishment of the practice as a part of engineering work, the subsuming of the academic enterprise into English department hierarchies, and the often-complicated relationship between employers and technical communicators collide to make those in the field ponder: do we have things to call our own? Can we delineate what is *ours* and what is *theirs*? Have we clearly articulated the value that this field delivers to the world? Are we a mature field yet? In short: are we there yet?

The answer to “Are we there yet?” is almost necessarily no; the need to ask the question suggests that the asker *knows* we are not there yet, but *cannot believe* that we are not actually there yet. Yet we argue that technical communication is a *maturing* field. By some accounts, the practice of communicating technical knowledge is centuries old (see Durack, 1997; Malone, 2007) but the field of technical communication has been recognized for more than 100 years. This history shows continued advancement in the practice, pedagogy, administration, and research of the activity that we call technical communication. To illustrate, let us give a quick historical sketch of the advancement of technical communication.

Technical communication concerns the delivery of specialized information about and via technology. This activity has been taught and practiced from as long as technology has been created. Aristotle is often mentioned as a forerunner of technical communication, given his systematic informational communication practices. Technical communication began more formally as a trained component of professional engineering in the 1890s (Kynell, 1999). Technology advanced rapidly after the turn of the twentieth century, and technical communication advanced along with it: The Society for Technical Communication notes that “the professional field was firmly established during the First World War, growing out of the need for technology-based documentation in the military, manufacturing, electronic, and aerospace industries” (Society for Technical Communication, n.d.).

The growth of technologies led to a need for teaching technical communication: technical communication pedagogy inside English and communication departments began in the 1950s (Connors, 1982), with the first technical communication program appearing at Rensselaer Polytechnic Institute in 1953 (Melonçon, 2012). Academic journals in technical communication started to appear in the 1980s, while standalone technical communication departments began in the 1990s.

The mass digitization of technical communication in the latter part of the twentieth century greatly increased demand for the skills of technical communicators (even if the job title “technical communicator” has never been prominent). Technical communicators currently fill roles as disparate as grant writer, medical writer, social media specialist, and content strategist (Brumberger & Lauer, 2015), with many technical communicators plying the trade in cutting-edge technology spaces. Programs continue to serve students in their own majors and in other majors, both in in-person and online modes. Thus, the academic enterprise of technical communication has been training professional information delivery specialists for over 130 years, and many industries’ desires for those skill sets have grown over that time.

The history of technical communication practice makes a clear case about the advancement of the field. Throughout this advancement, technical communicators have routinely taken stock of technical communication practice and pedagogy, assessing what the field needs to do to mature into its own freestanding, fully developed concern. Often, factors have been found to be lacking, such as a lack of established practices, a lack of professional certification/gatekeeping, or a central topic of the field. The various concerns shift over time, as the field moves to address the previous concerns.

Where advancement of technical communication practice resulted in the development of a range of practices and an overall expansion of the field to meet a variety of ends, questions of maturation have turned our attention inward, leading to conversations about best practices, professional certification, central topics, and core methodologies. These are questions rooted in disciplinary identity, which will surprise no one who follows such conversations. It will also surprise no one that these questions are still difficult to answer after all of these years. The growing professional record of technical communication practice promises that these questions will never become easier to answer. Yet we are at a spot where some cornerstones have been established and new progress can be made.

Technical communication has certainly reached a point where we have engaged in enough professional and teaching practice that we are no longer figuring out the principles of what works. We have done enough now that it is profitable for us to look back on all that has been accomplished to see what we have learned and, from that, to project a path forward. This inward look is a move that results in better self-awareness, moving technical communication from being more of an inductive field (focused on doing, trial and error, and invention) to being equal parts inductive and deductive (reflecting, learning from successes and failures, theorizing, and framing future practices).

This motion toward field maturation is already present and we are certainly not the first to point it out. Our niche is to suggest an approach that enables critical reflection on what the field has accomplished without oversimplifying the scope of the field. The need we identify in this chapter and respond to throughout this book is one of adopting methods of scale that are capable of accounting for

the scale and diversity of our field's practices. We will show how corpus analysis' technical capabilities and techniques can help address research questions about the field at large and individual research areas.

To establish the value of a corpus analytic methodology for field maturation, we first need to sketch a picture of the challenge posed by the scale of technical communication practice. It is scale that makes it difficult to obtain an overview of the field and observe the many currents that comprise it. The development of technical communication into a large-scale enterprise began more than fifty years ago: As society digitized in the 1970s, technical communication began a multi-decade shift from predominantly print products toward a mix of print and digital products (Carradini, 2022). The practices that produce digital technical communication products have been advancing and maturing over the last 50 years. One outcome of this shift to digital is a new ability to create and store of massive amounts of texts. The primary work of technical communication is now created, iterated, and delivered online. Online work encourages the proliferation of texts because the production and delivery constraints of online content are far fewer than the constraints of print content. Old genres have gone online, and in many cases they have stayed available online for extremely long amounts of time.

In addition to moving old genres online, technical communicators have created new genres in emerging online spaces. The genre of the forum post has become an integral part of technical communication work, as users relay technical concerns to organizations in idiosyncratic ways (Swarts, 2018). The crowdfunding proposal is a new genre that transforms the social and technical aspects of grant writing to a great extent: proposals on websites like Kickstarter and IndieGoGo allow writers to make an appeal to a public audience of potential funders (Carradini & Fleischmann, 2023). Podcasts and their attendant transcripts offer new ways to deliver oral and written information. These new genres reflect significant areas of development in technical communication practice and research, which add to the proliferation of texts.

Finally, the standard workplace communication practices of technical communication are now digital. Text is created as a byproduct of standard communicative interactions of organizations when essential operations move online, and much of that (such as email or Slack chat threads) is stored indefinitely. Thus, the number of texts in technical communication has proliferated rapidly as online practice has developed. The now-commonplace nature of online activity that gave rise to this amount of text suggests that the proliferation is likely to continue.

Given these three drivers of text proliferation in technical communication organizations (old genres going online, new genres developing, standard digital communications being stored), many organizations involved in technical communication have amassed huge amounts of digital and digitized texts. Some companies have produced decades of digital work stored in online content management systems. These CMSs can include huge numbers of policies, reports,

product-related content, social media posts, user comments, interview notes, and video transcripts fill practitioners' content management systems.

Technical communicators don't just write copious amounts of texts, however. They also do things with those texts. Doing things with texts, and especially managing texts, requires a reflective understanding of what texts do, how they are used, and what values and motives they represent. Yet management, categorization, delivery, and storage are not the only practices affected (Hackos, 2002; Halvorson & Rach, 2012). Technical communicators must consider how to communicate with customers, attract funding, shape user experience, communicate identity, and share knowledge amid the troves of content held by their organizations.

Academic technical communicators also have varied needs within this deluge of texts. Digital and digitized research articles, conference presentations, syllabi, web content, student papers, and class materials have accumulated over more than half a century in academic technical communication. An ever-growing amount of academic research identifies opportunities to further aid practitioners, engages with underrepresented topics and voices, develops research practices, and considers new ethical concerns. Researchers must consult this consistently growing body of literature to understand the state of research topics and develop further projects. Beyond research, teachers and administrators assess the vast amount of text that classes collectively create to evaluate students and appraise how well teaching practices prepare students for an ever-changing workplace. With the amount of text in each of these categories growing every day, academic technical communicators need methods and tools to help make sense of this ocean of texts.

Ultimately, we argue in this book that corpus analysis is a method that can help technical communicators of all types respond productively to the immense amount of text created by the various arms of the field of technical communication. This method can aid in reflective study of technical communication to help further develop our maturing professional practice and academic field.

Corpus analysis offers a way to approach the work of technical communication at the source material's level of scale by allowing analysis of more texts than an individual or team could read alone. Researchers can then draw out insights that hold across large numbers of texts and apply those insights to the concerns at hand. In this book, we explain concepts, describe techniques, give examples, and outline potential applications of corpus analysis for technical communication practice, research, teaching, and administration. We offer emerging technical communication scholars, established faculty, and practitioners a way to further develop and maintain awareness of their work at scale.

■ A Brief Sketch of Corpus Analysis

As a brief introduction, the method of corpus analysis helps researchers study collections of texts larger than an individual could analyze alone. A corpus (singular)

or corpora (plural) must be organized around unifying characteristics (such as topic, professional organization, chronological window, or all of these together) and converted to machine-readable text. Researchers use various analytic techniques to quantitatively identify patterns of words, phrases, and other discourse objects from a corpus that can support analysis of their use patterns.

For example, a researcher could analyze a corpus of 50,000 comments from a user help forum on a piece of software. To build the corpus, the researcher could download all the comments from the content management system and convert them to .txt files. To conduct analysis, the next step would be to upload the corpus to an analysis tool such as Lancsbox or AntConc, then generate analyses of word frequency in the corpus (the baseline first step of many corpus analysis efforts). After the initial step of word frequency, scholars can conduct analyses that build on those results. One analysis could identify words that appear with unexpected frequency in the corpus (as compared to a different corpus for reference), words that are unexpectedly absent in relation to the reference corpus, and words that commonly appear together. Further investigation with these analysis methods can help the researcher discover in those forum comments topics that users frequently need help with, errors or challenges they often experience, or issues that are changing in frequency over time. These insights can help identify areas of documentation needs that could be too cumbersome to do through manual inspection of the forum comments or too prone to bias depending on how individual users might be queried for the same information.

Thus, corpus analysis allows the field to identify and find evidence of its practices in text while also allowing assessment of those practices at a scale that allows us to reflect on what we know and what we do. Much corpus analysis has already been conducted in technical communication (Boettger & Ishizaki, 2018; Orr, 2006). Scholars have investigated how specialized terminology regarding search engine optimization is translated into Spanish (Laursen et al., 2014); what rhetorical strategies are included in writing templates for professional letters (Kaufer & Ishizaki, 2006); “how corpora can help copy editors adopt a rhetorical view of prescriptive usage rules” (Smith, 2022, p. 194); how passive style is used in civil engineering practitioner documents, with the goal of teaching writing more effectively (Conrad, 2017); and more over the past 30 years. Even though technical communication is not new to corpus analysis, we argue that the practitioner and scholarly concerns that come with the maturation of the field call for more corpus analysis.

Although technical communication has an established body of corpus analysis work, even more corpus analysis has been conducted outside technical communication, in writing studies more broadly. The comparison to writing studies more generally is important because researchers in that field have already reached the same point of historical development and maturation as technical communication, and they have already been using reflective corpus analysis to theorize, to establish best practices, to guide future experience. We draw on these and other

studies to demonstrate the capabilities of corpus analysis and suggest how they could be turned toward technical communication's goals and ends.

The first step in the research method is to pick a technique of quantitative analysis. Frequency is a common first place. Identifying the most frequent words in a corpus has many applications for research. Some research questions can be answered through frequency alone, such as: “what are the most common adjectives and adverbs in a set of user help tickets?” This information can help identify areas of text that indicate users' emotional experiences regarding a software. Frequency allows us to answer some of the more basic reflective questions: what do we write about and how often?

Another technique is keyness, which uses statistical analysis of the frequency of terms in two corpora to determine words that are more “key” to one corpus than the other. This technique can be used to answer questions such as “What words and phrases differ between two sets of reports published ten years apart on climate change mitigation?” This information could show change or lack thereof in an organization's attempts to help mitigate the effects of climate change. Here too, keyness allows a kind of reflection. In addition to revealing what the field writes about, keyness tells us what our body of work is about, what is important, and what differentiates the field from other fields.

A third technique is collocation analysis, which shows researchers what words often occur near each other. This technique can help answer questions such as “what nouns or pronouns appear nearby conditional words (if, might, could, would) in our content management system,” where this information could be used to assess accessible language in healthcare documents.

These three techniques each point toward words, phrases and other discourse objects that can be relevant to answer research questions or repay further quantitative or qualitative study (Archer, 2009a). As with frequency and keyness, collocation facilitates reflection by elucidating how keywords combine into larger conceptual units. It allows us to name what complex topics have preoccupied us as practitioners and academics. These techniques reveal emergent patterns in texts that represent the outcomes of our professional and academic practices. These patterns are a form of evidence that supports reflective analysis of what those practices are and perhaps how they have changed over time. We will discuss these techniques and their application in Chapters 2, 3, and 5.

■ The Possibilities of Corpus Analysis in Technical Communication

In the sections that follow, we show how corpus analysis can be specifically valuable for practitioners, researchers, teachers, and administrators of technical communication. Corpus analysis can assist researchers of technical communication in conducting research on existing topics and emerging topics. Teachers

and administrators of technical communication can use corpus analysis to assess and enhance teaching and programmatic outcomes in a variety of ways. Corpus analysis can also affect many different types and areas of work for the practitioner, such as in handling user feedback, tackling content management, and conducting large-scale technical editing tasks. Gaining reflective insight about our practices can help us be more deliberate and intentional in those practices as well as critical of those practices, as warranted.

To be clear, the specific topics that follow are illustrations of corpus analytic techniques that can allow practitioners and scholars alike to reflect on what the field has accomplished, what those accomplishments may mean, and project from them to additional questions. It is not our aim to set an agenda for field research; given the many strands of technical communication research, many agendas should be set. We will argue in the final chapter that agenda setting may be a necessary future step, considering the effort required to build and maintain corpora. For now, we hope to point toward areas where corpus analysis could productively aid ongoing research efforts, acknowledging and expecting that each area's researchers will find more ways forward for each topic.

■ Research

Corpus analysis can be used on corpora of texts representing professional activity. Technical communication researchers can use corpus analysis' techniques of identification and re-contextualization to reflect on what we have learned through and across research studies. To begin, we will discuss existing and emerging areas of interest for technical communication.

■ Existing Area of Interest: Genre

Corpus analysis can extend and support work in well-established areas of technical communication research. Topics such as genre and medical communication are two of many areas that could have open questions further analyzed by large-scale, document-based research.

Technical communication scholars have been interested in genres of technical communication for 40+ years. And genre questions tend to be big questions, the answers to which are intended to give us insight into entire genre types, genre systems, and historical eras of genre development. Genre gains power as an explanatory concept when conventions can be displayed as common across many instances. S. Scott Graham, et al. (2015) note of a big-data approach to genre that uses statistics: "Characteristics that may be invisible at the level of a single text may become visible in a statistical representation that takes into account an enormous number of texts" (p. 92) This statement is true of corpus analysis as well. Because corpus analysis can reveal generic conventions at scale and show in detail the elements of a trend that are present across a large number of documents, the

results contribute to our ability to explicitly discuss what is customary or routine. By the same reasoning, corpus analysis can help falsify genre claims by revealing patterns that may seem interesting but are isolated and idiosyncratic: “Genres can be defined . . . with more precision (i.e., which features are actually typical across the genre, not just in the particular text one may have analyzed?)” (Graham et al., 2015, p. 92). Thus, corpus analysis can help confirm, support, and extend the findings of qualitative genre research by investigating large corpora of genred text.

Corpus analysis also offers ways to reflect, deliberately, on what we know as a field and our own genres (e.g., consider Dryer, 2019); in this specific case, it allows researchers to develop and extend genre research by considering these open questions with more examples across an ever-widening range of genres.

Genre scholars already conduct comparisons across methods (Campbell et al., 2020; Miller et al., 2018;) or on larger data sets (Robles, 2018) as ways of confirming findings, but many open questions in genre can benefit from large-scale analysis. Genre scholars are interested in understanding how new genres work, particularly new genres on the internet (Mehlenbacher, 2019; Robles, 2018); how genre operates in multilingual and multinational settings (Hodges & Seawright, 2019); how emotions interact with genres (Miller et al., 2018; Weedon, 2020); and the evergreen concern of how to teach genres (Kim & Olson, 2020; Tardy et al., 2020).

Each of these expansions in the study of genre builds on existing research. As that body of research grows, our capacity to gain an overview of those genre practices, and to examine large scale patterns and changes over time, grows. So too does the challenge of engaging in such investigations grow. Corpus analysis can be used for meta-research: research on the research. Meta-research makes connections across large bodies of research to assess trends or patterns in the research. Technical communication scholars can use corpus analysis to reflect on what we cover in our research and how we have covered it. This reflective practice can help identify points where we can steer the field’s research in new directions.

Technical communication scholars can and do use corpus analysis and related types of large data analysis for meta-research. Researchers frequently mine corpora of technical communication research to identify disciplinary issues. Ryan K. Boettger and Erin Friess (2016) investigated “the content alignment (or lack thereof) among academics and practitioners” as exemplified in work published in academic and practitioner outlets. They found little content alignment via their quantitative content analysis of 1,048 articles, suggesting that the field is fragmented in its research interests. This comparative analysis once again demonstrates the value of contrasting corpora as a technique.

Kate White et al. (2015) conducted a quantitative content analysis of nine textbooks and 1,073 articles from five technical communication journals using keyword searches to identify “the treatment of gender and feminism in technical, business, and workplace writing studies” (p. 27; also, the title of the article). After reviewing content associated with the terms “female, feminist, gender, gendered,

cross-gender, gender-neutral, sex/saxes, sexual, sexism, sexist, and woman/women” (p. 34), the authors argue that “the discourse seems to paint a false picture of the workplace as neutral and nongendered” (p. 49). This article used a quantitative approach (checking books and articles for the existence of terms) to drive a qualitative analysis of what the limited number of book sections and articles about gendered issues meant for the field. This combination of quantitative and qualitative approaches is a productive one that we will discuss further in Chapter 2.

In another meta-study, Heather Noel Turner (2022) conducted a corpus analysis to compare the topics of the ATTW conference presentations against topics found in the *Technical Communication Quarterly* journal, finding ways that the journal topics and conference topics support and diverge from each other. This type of comparative corpus analysis allows for clear differentiation between corpora. Turner used keyness as a way to determine degrees of difference in terminology use between the ATTW corpus and the TCQ corpus. Building on Turner’s work, conference content could be further mined. The text of technical communication books (building on the work of Rude, 2009) and textbook content (following White et al., 2015) are two more of the unintentional repositories of data waiting to be activated as a corpus and to deliver field-level insights.

Corpus-based meta-research can also be conducted to build theories, as Julie A. Corrigan and David Slomp (2021) do. They conducted a “critical review of writing scholarship from the past 50 years” to “synthesize the significant scholarship in the field in order to advance theory” (p. 143). Their content analysis of “109 texts revealed that the following writing knowledge domains have predominated the literature: metacognitive, critical discourse, discourse, rhetorical aim, genre, communication task process, and substantive knowledge” (p. 143), which they used to build a new theory about “the knowledge domains that constitute expertise in writing” in a digital age (p. 167). Technical communication scholars can also use corpus analysis for theory-building meta-research by identifying terms or phrases from a corpus of literature to examine further. Areas such as social media, user experience, and other areas with many contributing theories from varied fields could benefit from this type of journal article meta-analysis. Integrative literature reviews, which painstakingly synthesize journal articles on topics (Andersen & Batova, 2015a; Lauren & Schreiber, 2018) could benefit from the identification aspects of corpus analysis.

Another area where corpus analysis can be effective is medical communication, which has been a part of technical communication since before 2000 (Connor, 1993). Research on medical communication demonstrates that corpus analysis can be used on transcripts of oral communication as well as written documents. For example, Ellen Barton (2004) studied the oral communication of oncologists by drawing on transcripts of 12 “front stage” conversations inside clinical rooms and 33 “backstage” conversations between medical professionals and the researcher. Barton found that “the oral genre of treatment discussion in oncology encounters is organized to allow practitioners to do, appear to do, or

avoid doing difficult work like presenting a prognosis” (p. 67). Barton discovered this finding by analyzing the structure of the oral presentations that oncologists gave to patients and family, as well as comparing the content of the message to the patients with the content of the messages spoken outside the clinic room to other professionals and the researcher. This comparative method allowed Barton to develop a critical awareness of differences between the two, in order to ask why those differences might be important.

Corpus analysis also allows cross-referencing of large amounts of texts against other data. For example, Graham et al. (2015) conducted “statistical genre analysis” on a large corpus of transcripts and metadata from Federal Drug Administration drug advisory committee meetings. They found that “the use of efficacy data seems to lower the chance of approval, whereas a greater presence of conflict of interest increases the probability of approval” (p. 89), which “indicate[s] the need for changes to FDA conflict-of-interest policies” (p. 70). The ability to cross-reference the content of texts with metadata (in this case, metadata being the outcome of voting on the approval of a drug as a result of the meeting) led to an insight on how the content of the meetings may have affected the outcome of voting. Researchers can conduct this type of cross-referencing outside medical documentation for proposed policy documents concerning issues of technical communication interest that may have a range of outcomes (passed, tabled, returned to committee, rejected), as well as emerging proposal genres such as crowdfunding campaigns (Ishizaki, 2016) that have largely binary outcomes.

■ Emerging Areas of Interest: Social Justice

A reflection on what a maturing field has done can also present the opportunity to recognize what has been unaddressed and what has yet to be done. Reflection can be agenda-setting. One example of this outcome for reflection is the growing focus on social justice over the last 20 years. The term “social justice” did not appear in the abstracts of five technical communication journals from the period 2000–2005, but appeared in journal abstracts thereafter (Carradini, 2022). Social justice work in technical communication seeks to be productively critical of and to intervene in the ways that writing, discourse, and actions based on discourses can systematically exclude or marginalize particular readers and reader experiences.

Research methods that call us to reconsider “established” knowledge, make textual problems visible, catalog the scope of problems, and illumine starting points for interventions can be an aid to social justice research. Corpus analysis is one such research method (among others). Thus, corpus analysis can fit in with the work that scholars of social justice do and are calling for.

First, the conceptual basis of corpus analysis can aid the overall goals of social justice work. In Emily January Petersen and Rebecca Walton’s 2018 call for critical, feminist analysis in addition to critical action we identify a space for corpus analysis: “We agree that action is needed to redress inequities, but we

also see a potential danger in the field's shift toward critical action if that shift is not carefully informed by critical analysis" (p. 418). To that end, corpus analysis can be used in critical ways to support and encourage critical action. Researchers can produce studies that identify issues in texts that need addressing, review field-level practices to make sure social justice practices are achieving their desired outcomes (Itchuaqiyag & Matheson, 2021), and make connections between topics in corpora that appear comparatively infrequently and actions that could be taken to redress those textual practices. In short, corpus analysis can illumine potential ways forward for social justice efforts through critical reflection on large amounts of texts.

Corpus analysis also calls us to reconsider "established" knowledge through critical reflection through all parts of the research. The process of thinking through a corpus analytic study creates moments for critical reflection before the data is even collected. Considering what constitutes a representative corpus of content for study (see Chapter 4) requires researchers to have an educated sense of how to build a corpus that represents both the range and diversity of the field and its practices. Daniela Agostinho et al. (2019) remind us that any large collection of data is a form of archive, and all archives can have serious limitations and exclusions. Historically, archives have "overlooked the experiences of women and queers" while archives related to slavery and colonialism expose "both the capture and exclusion of people of colour in and from archives and the kind of knowledge that can be gleaned from the archives of the ruling classes, archives that dehumanise those under colonial rule" (p. 424). While data gathered without careful attention to what is going into the corpus can reproduce these sorts of inequities, gathering data that effectively represents the range of content in a situation can produce corpora that help lead to research that helps identify and, ideally, correct injustices of this type.

This type of pre-collection reflection is necessary because concerns of bias in corpus analysis are legitimate: if bias goes into the data, then bias can come out in the findings (O'Neil, 2016). This concern features prominently when professionals use "big data" for controversial ends, such as training artificial intelligence to skim pools of job applicants' resumes (Miller, 2019), analyzing loan applications (Lane, 2017), handing down sentencing suggestions in courts (Tashea, 2017), and ever-more invasive iterations on this theme (Stephens, 2018). Whatever biases exist in training data will be reproduced in the results the algorithm produces. Constructing corpora carefully (Chapter 4) and conducting work that identifies bias can work against these trends.

Next, corpus analysis can help make social justice concerns visible in large amounts of texts. Work identifying systemic bias and discrimination, systemic racism, systemic misogyny, systemic homophobia, systemic classism, systemic ableism, and more can build on qualitative work, extending and supporting these concerns to develop a wider picture of the problem. This kind of language is said and printed in public and private spaces, which perpetuates casual discrimination

and bias of all sorts. This sort of language can persist unless we first make an effort to locate it, which could be accomplished by investigating the possibility of overrepresentation bias via an analysis of keywords. Underrepresentation bias could also be sought through an analytic method called “negative keywords,” or evaluating a corpus of text against a prepared corpus that reflects an expected dispersion of a term or phrase tested against a study corpus for comparative over/under representation. Understanding the trends of bias, overrepresentation, and underrepresentation in texts can support claims that certain types of text include bias against certain types of people in specific ways. From there, interventions can be designed on the local or individual scale to address the issue.

Beyond aiding researchers in identifying certain types of problems in text, corpus analysis can help researchers catalog the scope of problems by quantitatively displaying the scale of those types of problems. Qualitative analysis can identify findings in a small-to-medium amount of data, and corpus analytic techniques can help researchers test to identify if those practices are present in large amounts of data. Corpus analysis is very well suited to establishing mathematical relationships between words, such as identifying if one word is much more frequent than another or if one word is unusually absent in a corpus (in relation to a reference corpus). These tracking and confirmation efforts can help establish the scope of a problem or interest area over a large set of data. This results in research that can determine if those relationships appear across a large amount of data or are idiosyncratic to a specific text, condition, or situation.

This type of work already exists: Godwin Y. Agboka (2021) used a quantitative content analysis to point out the scope of the problem of using the word “subject” to describe the human participants of technical communication research. Likewise, Barton et al. (2018) used content analysis to identify the circumscribed ways that community members contributed to research ethics discussions concerning their neighborhood. While these analyses were not corpus analytic, they were large-scale approaches to text analysis that can reveal patterns of activity in text (e.g., portrayal and participation) that point toward topics of interest, findings, and suggestions for action.

Corpus analysis findings can also help illumine starting points for interventions in social justice concerns. After analyzing 450,000 online comments from *New York Times* articles, for example, John R. Gallagher et al. (2020) point out social justice interventions that could take place in the space of online content moderation. While the authors acknowledge the complexities and difficulties of localizing their ideas to individual websites (pp. 167-168), their findings present a starting point for more local interventions to develop and grow in relation to this concern. The authors do not explicitly use corpus analysis as a method, but they do demonstrate the value of a large-scale analysis of text (similar to corpus approaches).

These steps of reconsidering “established” knowledge, making social justice concerns visible in texts, cataloging the scope of the problem, and illumining

starting points for local interventions are each demonstrated by Cana Uluak Itchuaqiyaq and Breeanne Matheson (2021). The authors began by reconsidering “established” knowledge, as they “used corpus analysis techniques to investigate the field’s working definition of ‘decolonial’ as it relates to methods and methodologies” (p. 21). The reason to reconsider this “established” knowledge, they argue, is that “TPC scholarship designed using decolonial frameworks lacks a clear, centralized definition and may overgeneralize and/or marginalize Indigenous concerns” (p. 20).

Itchuaqiyaq and Matheson (2021) employed corpus analysis to make visible a concern they had about texts: “we already suspected that many of the texts would use decolonial as a euphemism for social justice or humanitarian work because of our previous exposure to this particular critique coming from scholars Tuck and Yang (2012)” (p. 21). Their corpus analysis cataloged the scope of the problem, identifying that this concept did hold across a range of texts (p. 24). This finding prompted the authors to suggest starting points for local interventions, as they “propose a centralized definition of ‘decolonial’ that focuses on rematriation of Indigenous land and knowledges” (p. 20). Thus, the process of corpus analysis in this article moved from a reconsideration of knowledge via an initial concern about a concept, through making the concept visible via analytic techniques and a subsequent confirmation of the concept in a variety of texts, to suggestions on how to redress the issues raised as a result of the analysis. This exemplar shows that corpus analytic work can help social justice researchers be productively critical of and intervene in the ways that writing, discourse, and actions based on discourses can systematically exclude or marginalize particular readers and reader experiences.

■ Emerging Areas of Interest: User Experience

Corpus analysis can aid other areas of technical communication practice as well, such as user experience (UX). User experience currently has little published corpus analysis work conducted about it. Yet the nature of UX suggests that researchers may use forms of large-scale analysis (such as content analysis) to work with the texts representing many user experience tests. Assessing many tests at once could reveal holistic insights about users. User experience research takes many forms, with some of the more prominent being researcher-guided speak-aloud protocols tests. Technical communicators often record these complex tests for analysis purposes. If the test is of a computer-based item, then the user’s screen, the user’s voice, and the researcher’s voice may be recorded separately or together. The oral recordings can be transcribed (automatically by a machine or by a human) and used as the basis of corpus analysis. Florentina Armaseleu (2022) analyzed a corpus of recorded user experience tests regarding a software for viewing historical documents, identifying four different categories of users as a result of their transcribed oral responses to the software. This type of analysis can be implemented in a wide variety of user experience test transcriptions, regardless of the number of tests.

Corpus size may be small in user experience research, as user experience tests for a specific piece of software or website often include fewer than 100 tests (although the number can range into the hundreds or thousands). However, corpus analysis of the transcripts of many studies (which could range into the hundreds or thousands) may be able to tell a researcher about the guidance habits of researchers in think-aloud protocols. Alternatively, a corpus of usability test transcripts regarding many versions of a software could be structured chronologically to investigate how users' difficulties or successes in the software changed over time. Similarly, a chronological corpus of that type could be investigated for changing habits of researchers' guidance over time. Chronological analyses can develop over time to become more meaningful as the researcher collects more data, because in many cases a long chronological window can demonstrate a phenomenon's persistence, growth, or decline more meaningfully than a small window.

User experience research can also be conducted in other ways. Phillip Brooker et al. (2016) use corpus analysis to identify "user experiences of epinephrine auto-injectors ('epipens')" from a small corpus of posts on Twitter (around 4,000 tweets over 68 days) (p. 8). They argue that the corpus approach:

allowed us to explore a broad topic of interest—epipens—without relying on simple term frequency to point us in any particular direction. Navigating around the cluster map in this way, analysts can sift their data for "needles in haystacks"—here, this provided insight into user experiences with epipens unlikely to be uncovered with more formal search terms (i.e., "weight" and "size"). (p. 9)

Thus, user experience information can be gleaned from social media sites, help forums, website-hosted email forms, and other areas of user-generated content via corpus analysis.

Corpus analysis may also help provide perspective on emerging trends in professional technical communication. For example, entrepreneurship is another topic of interest associated with volumes of text. One type of analysis of entrepreneurship has focused on the genres of entrepreneurial activity (Spartz & Weber, 2015; Spinuzzi et al., 2014). As with the study of genres mentioned above, corpus analysis could reveal aspects of entrepreneurial communication and activity across a wide set of examples. Identifying distinctive aspects of successful or unsuccessful entrepreneurial activities or communication habits could be instructive for entrepreneurs and for those seeking to teach entrepreneurs.

Transcripts of meetings between entrepreneurs and funders that take place after funding cycles could help researchers identify ways that entrepreneurs signal success or make efforts to repair relationships amid difficulties. From a pedagogical perspective, instructions on how to be entrepreneurial abound; using a corpus analysis to research what consistent claims, ideas, or patterns are present across many different forms of pedagogy (popular books, textbooks, websites, transcribed online videos, etc.) could develop categories or meta-categories

(categories of categories) for the types of concerns that successful or unsuccessful entrepreneurs have. This point about meta-research leads us to the next category of research where corpus analysis can be particularly valuable.

■ Education

Academic technical communication includes research, pedagogy, and administration. Corpus analysis can help technical communicators reflect on our accumulated pedagogical and administrative practices. Many technical communication programs have been active for long enough that we can use the accumulation of texts over time to assess the efficacy of our teaching practices. In fact, many studies of first-year writing in the composition sub-field of writing analytics and its attendant *Journal of Writing Analytics* have already shown that corpus analysis can productively aid the teaching of student writing (e.g., Aull, 2017; Holcomb & Buell, 2018). Administration of pedagogy can also be aided by corpus analysis, as the texts students produce (Peele, 2018), syllabi, or other artifacts of teaching can be analyzed to help programs best fulfill their remit to educate emerging technical communication students.

■ Pedagogy

Teachers can use corpus analysis on large amounts of text to aid student development in a variety of ways. Lexical elements of student writing can be made meaningful on their own or in comparison to the corpora of work of more advanced writers. Teaching individual argumentative strategies, citation moves, or practical elements of various genres scratch the surface of the possible findings (and attendant teaching outcomes) that corpus analysis can contribute to pedagogy.

Decades of pedagogical development and implementation in technical communication has produced a potentially immense amount of data on student progress and student development. Technical communication teachers can consult this data to understand what our pedagogy has focused on and how students have used the skills they have been taught. This work requires corpora, and technical communication scholars have begun to collect, curate, and investigate student work at scale to create corpora. Ryan K. Boettger and Stephanie Wulff created a corpus of technical communication student writing that allows investigation of how students respond to prompts, use language, invoke topics, and more (Boettger & Wulff, 2022). Similarly, Bradley Dilger, Michelle McMullin, and others have developed CROW (Corpus & Repository of Writing) to “create a web-based archive for research and professional development in applied linguistics and rhetoric & composition” (Staples et al., 2021), while USF Writes includes a large corpus of student technical and professional writing that the Department of English at the University of South Florida uses for “continual assessment processes, and programmatic and pedagogical improvement” (University of South Florida, 2023). Researchers initially developed the Stanford Study of Writing

(SSW) corpus for qualitative analysis, but Noah Arthurs (2018) used quantitative analysis to investigate a subset of the corpus. The author found that corpus analytic techniques revealed stances students took toward their topics, the topics of their writing, and elements of sentence complexity. Given these features and metadata associated with the texts about the students' characteristics, the study "characterize[d] the development of the SSW participants across four years of undergraduate study, specifically gaining insight into the different trajectories of humanities, social science, and STEM students" (p. 138). These findings provide insight into the process of writing development and can be useful for curricular development and course design.

Analysis of student work outside established corpora can also directly help develop pedagogical outcomes. Individual instructors can conduct analysis of students' papers over an individual assignment, class section, or semester's worth of classes to evaluate elements of writing practice. (These are sometimes called ad-hoc corpora in relation to permanent corpora, but ad-hoc corpora seem to be more common than permanent corpora, due to the complexity of corpus creation in large-scale, permanent corpora; see Anne Lise Laursen et al. [2014].) Some grammar elements of writing, such as nominalizations and conjunction use, are readily identifiable and can be assessed in relation to desired pedagogical outcomes. More complex analysis is possible as well, focusing on words frequently appearing together or words frequently appearing in the beginning of the assignment that may allow the instructor to understand how students are taking up the class information into their own writing processes.

Barton (1993) demonstrated how analysis of an ad-hoc corpora of student argumentation from a university writing proficiency requirement could help identify differences between the approaches of writers who vary in experience. Comparing student writing with that of writers writing in the *Chronicle of Higher Education*, Barton (1993) focused on the writers' use of evidentials: "words and phrases that express attitudes toward knowledge," such as *must*, *should*, and *I believe that* (p. 745). Through a discourse analysis comparing 100 student papers to 100 *Chronicle* opinion articles, Barton demonstrates that experienced academic writers "adopt an epistemological stance that privileges knowledge defined as a product of contrast" (p.754) (as demonstrated by use of phrases like *as a result* and *undeniable*) while student writers "more consistently assume an epistemological stance that privileges knowledge defined as a product of shared social agreement" (p.765) (as demonstrated by use of phrases like *today in America* or *most will agree that*). While this finding demonstrates an area where young writers can be taught conventions of academic discourse, Barton identifies an opportunity for pedagogical reflection, suggesting that "we may wish to ask ourselves why we seem to be rewarding our student writers primarily for reproducing our own contrastive and competitive epistemological stance" (p. 766). Barton's study is successful because it relies on analysis of two contrasting corpora to identify, classify, and illustrate (with examples) instances of evidentials. These findings can relate

directly to pedagogical efforts by individual teachers in technical communication classrooms. Other similar studies could produce findings concerning writing in technical communication: instruction sets, regulatory writing, grant writing, social media writing, and more.

A similar example concerns how corpus analysis can support pedagogical choices related to genre knowledge and genre writing skill in first-year writing. Laura Aull (2017) compared two different types of student writing: argumentative vs explanatory. Aull identified “generalized, interpersonal, and persuasive discourse in argumentative essays versus more specified, informational, and elaborated discourse in explanatory writing, regardless of course or task” (p.2). This type of finding relates both to genre analysis and pedagogy, as work of this type can be used to identify specific types of arguments or moves in genres to teach them to students.

Along the same lines, Steven Walczak (2017) developed tools to distinguish between the prose of different types of genres, creating exercises for students to be able to develop information literacy by distinguishing text from different genres (newspapers, magazines, and journal articles). Walczak’s work demonstrates how student use of corpus analysis can directly relate to genre learning.

Also adopting a pedagogical focus, Ian G. Anson et al. (2019) used corpus analytic tools and custom-built tools to study text recycling in published academic research: a researcher’s use of their own previously published sections of text in new academic work. Their custom tools allowed them to identify close matches or subtle changes in sentences (instead of exact copying) that would reflect different types of text recycling for different purposes. Understanding the purposes and contexts of text recycling could help student writers recognize conventions of different discourse communities in regards to the practice of text recycling.

Similarly, Ryan Omizo and William Hart-Davidson (2016a) created tools to identify the “hedginess” of published academic research writing, identifying one goal of the work as: “For learners, tools like the Hedge-O-Matic might make explicit the kinds of patterns that are expected by scientific discourse communities” (n. p.).

Corpus analysis has also been used to suggest effective types of mentorship for advanced students. Omizo and Hart-Davidson (2016b) “explore[d] the possibilities of using computational methods to create an assistive environment for advisor-advisee mentoring in academic writing” (p. 487). They identified “lexical patterns and rhetorical uses of the in-text citations” to create categories of citation moves (Extraction, Grouping, and Author[s] as Actant[s]), then compared these moves between three dyads of advisor-advisee writing. They sought “to automate the discovery of a generic baseline for citational moves among academic mentoring relationships” (p. 507). Corpus analysis can replicate this process to determine the progress of advanced technical communication students. This process can also compare students’ failed grant campaigns to successful ones for evidence of stylistic differences that may have been hindering the grant. Similarly,

comparing multiple versions of theses to identify areas of and types of significant development over time can aid the thesis-writing process.

■ Program Administration

Corpus analysis of student work can help with writing program administration, as Danielle Wetzel et al. (2021) note: “Those of us who lead writing programs continue to press toward using writing analytics to better understand how to design, deliver, and assess instruction” (p. 292). Corpus analysis of pedagogical and departmental materials can help administrators analyze, assess, evaluate, and improve pedagogy consistently and continuously (Sonnenberg et al., 2022).

Corpus-assisted studies about trends in student writing can tell us something about the changing nature of student work and the kinds of pedagogical practices that can effectively reach students. We can also learn something about our values in the process, by discovering how corpora of student writing or our own teaching materials tell us something about how our academic programs are oriented to particular outcomes. For example, Dylan Dryer’s corpus analytic work studying scoring rubrics (2013) reveals insights about how the instruments that instructors develop for assessing writing shift attention to qualities taken to be inherent in the writing and the writers, rather than situationally derived qualities. In other words, corpus analysis can help us better understand the instruments and analytics that we use to gain perspective on programmatic pedagogical choices.

Corpus analysis is an ideal tool for large-scale assessment of student work emerging from a program, and findings from that assessment can lead to insights into how to design and deliver pedagogy. For example, Wetzel et al. (2021) demonstrate a textual tool named DocuScope Classroom that allows a wide range of tasks:

Programs can make claims about particular curricular goals and align those goals with in-class instruction. We believe this approach facilitates a reconceptualization of assessment as both rhetorical and genre-based, but also as formative for instructional design, informing the vertical integration of writing skills across a curriculum as well as course-level instruction, for both academic and professional writing tasks. (p. 293)

DocuScope allows for easy comparison of documents and sections of documents. Students and teachers can use this tool to understand the rhetorical choices in student writing; students can use it to formatively analyze their own writing choices in comparison to others’ choices, while teachers can use the tool to assess and visualize aspects of students’ written work. Once lexical items and rhetorical choices have been identified, teachers and administrators can assess whether the students, the assignments, and the curricula work together to produce strong

writing outcomes. This type of tool can be used in technical communication classrooms and programs just as Wetzel et al. imagine it:

From a bird's-eye view, we can bridge the gap between university and workplace writing by mapping genre features according to their rhetorical purpose and function rather than their lexicogrammatical structure. Explicitly teaching rhetorical patterns across a variety of genres, through data-informed visualizations from DocuScope Classroom, may prime students to see relationships between writing tasks they encounter, enabling meaningful learning transfer (p. 319-320).

While specialized tools such as DocuScope Classroom are invaluable for certain types of analysis and outputs (such as visualizations), basic tools can also provide program-level insights. For example, Thomas Peele (2018) used a corpus analysis of 548 student essays as an “assessment tool, providing a microscopic view of a limited number of rhetorical moves. . . . As a result of our study, we hoped to be able to create assignments for research essays that responded directly to the patterns that we saw in our students’ essays” (p. 79). Comparing the rhetorical moves students actually made to the moves taught to them from *They Say/I Say* allowed the teachers to assess the students’ uptake of tasks at a programmatic level and create curriculum that responded to what they found. Thus, corpus analysis of classroom work that leads to programmatic assessment can work at a variety of levels of scale, complexity and experience: Peele noted that the researchers had “little prior experience with corpus analysis” and used the main functions of a standard corpus analysis tool (ANTCONC) (p. 79).

Other content types could be productively studied for administrative purposes. Web content from technical communication programs’ websites could be analyzed to identify ways that programs position themselves in relation to their universities, communities, theories, practices, or other concerns. Analysis of the types of news stories or updates that programs present on their websites may reflect pedagogical or administrative priorities. Analysis of terminology in frequently offered class names can shed light on areas of growth in the field and potential development for individual programs.

With ethical considerations in mind, corpus analysis can be a productive tool for student invention, classroom assessment, programmatic assessment, and curricular development. Ideally, students can take what they have learned about corpus analysis in their studies with them to the workplace. One of our reasons for writing this book is to encourage this sort of work in technical communication programs. While the field has matured, the area of program administration is one that has not taken advantage of corpus analysis work to the same extent as first year writing has (as evidenced by the comparatively smaller number of studies in technical communication on the topic). Program administration is a place where corpus analysis can help the field continue to grow and mature.

■ Practice

Corpus analysis can also aid technical communicators in the workplace, both in their daily work and research. Whether the aim is to understand a mountain of user input, customer queries, focus group answers, or usability test feedback, corpus analysis methods offer ways to draw meaningful conclusions and get work done. Technical communication practice is constantly evolving. Writers search for ways to gather better feedback and incorporate that feedback more thoroughly and consistently. Managers consistently seek improved efficiency and more effective oversight. These motives can be met through corpus analytic reflection on existing practice. This reflection can start in prominent places, such as trends in user feedback; considering less prominent places, such as language that facilitates efficient translation and more effective localization, may also be productive. Below, we demonstrate three areas where corpus analysis could offer benefits: user feedback, content management, and technical editing.

■ User Feedback

Consider user feedback, a common part of technical communication (Swarts, 2018). Technical communicators can handle many thousands of units of user feedback solicited from online forms and forums. Tom Johnson (2020) notes that “user champion,” a person who gathers user feedback from a variety of sources and presents the user’s opinions to the engineers, is an increasingly common role that technical communicators take on. Although the promise of starting a user community and crowd-sourcing some aspects of documentation and revision seem enticing, it is easy to get lost in the sheer amount of user feedback generated. Using corpus analysis to examine the patterns of user contributions may help reveal systematic ways that users make contributions to documentation projects. Upon learning what those modes of contribution are, for example, one could develop tools to better support those kinds of reader engagement. Corpus analysis is well-suited to surfacing trends from a variety of sources, so organizations interested in studying their user communities can use corpus analysis to make the job of finding trends from users easier. Corpus analysis also provides a way of quantifying the severity of concerns to engineers (e.g., “51 users from three sources of feedback are concerned about problem X”).

■ Content Management

Content managers have enormous amounts of digital text under their control, as some organizations maintain vast internal content management systems containing decades of carefully developed and curated content. Content strategists and web content management experts tell organizations how to get a grip on all of their content, inventory it, and know what is covered and how (e.g., Hackos, 2002; Halvorson & Rach, 2012). Content strategists developing content models may also make good use of corpus analysis. Content modeling is about finding

what content goes together and what should be separate, and, in particular, what elements connect the content that should be organized together (Andrews, 2020) and these kinds of content patterns may not be apparent without examining many examples of similar kinds of connections. This work will take a detailed approach from someone capable of reflecting analytically on the findings, because fully algorithmic matching is not enough to create effective categories. As Michael Andrews notes:

Humans decide taxonomies—even when machines provide assistance finding patterns of similarity. Users of taxonomies need to understand the basis of similarity. No matter how experienced the taxonomist or sophisticated the text analysis, the basis of a taxonomy should be explainable and repeatable ideally. Machine-driven clustering approaches lack these qualities. (2020, n.p.)

Corpus analysis offers a way to identify explainable and repeatable bases of taxonomies whose significance can then be validated through close qualitative analysis.

■ Technical Editing

Technical editors also can use corpus analysis in their work with large amounts of data. Johnson (2020) notes, “as an editor, you might also check to see how the content compares to the competitor’s content. For example, does the content cover the same topics as the competitor’s docs?” (n.p.). Comparing content across two large sets of text to identify points of comparison or similarity is a task to which corpus analysis is well-suited. Technical communicators dealing with API documentation may use corpus analysis tools to compare and contrast aspects of APIs that change over time. Johnson notes that

[a]nother non-writing role we play is as an editor who makes the content align with style guides and standards, who figures out whether the content uses the right terms, whether it aligns with industry best practices and style guides, and so on. (n.p.)

Identifying varied term use is an ideal use case for corpus analysis because corpus analysts can assess large amounts of content for questions like these.

Similarly, one might employ corpus analytic techniques to examine subtle differences in documentation that affect how the content is translated and localized in different global markets. A comparison of documentation that has been successfully localized versus unsuccessfully localized might reveal patterns of language use that could be associated with known constraints on localization processes. The result of such an analysis could more readily lead to the kinds of comprehensive guides which inform practical technical practice (e.g., Kohl, 2008). These examples demonstrate several potential organizational uses of corpora and corpus analysis; many more areas of technical communication practice that can benefit from corpus analysis.

To reiterate our point made prior to this review of potential areas for reflective scholarship, our aim has not been to identify areas of technical communication most in need of corpus approaches. Rather, our aim has been to argue that trends in ongoing areas of scholarship, teaching, and professional practice are already leaning into questions that reflect a maturation of the field, as well as a need for the kind of scope and vision that corpus tools can provide.

The specific ends that researchers, teachers, administrators, and students of technical communication will seek with corpus analysis tools will vary, but the tools each help point toward an overarching end: turning large amounts of text into insights that can positively affect the processes of writing for individuals. We will spend the rest of the book outlining in more detail the ideas, tools, and processes that allow people to conduct corpus analysis.

■ Chapters

In this section, we outline the chapters of the book. Each chapter illustrates how a step in corpus analysis research connects to practice and research in technical communication. We will cover initial ideas, tool use, data processing data, reporting findings and more.

Chapter 2 explains the basic terms, techniques, and concepts of corpus analysis. We cover the main necessary elements of corpus analysis, walk through some techniques of analysis (such as keyness and collocation), and explain the theoretical assumptions of corpus analysis. In each of these points, we tie the techniques back to their use in technical communication research. This chapter shows how the analytic functions of corpus analysis align with the questions of technical communication. It also lays the groundwork for future chapters.

Chapter 3 considers how to form research questions for corpus analysis research. We offer an overview of the steps needed to frame issues of technical communication research or practice as questions that can be addressed through corpus analytic techniques. The chapter first discusses the affordances and constraints of qualitative, hand-coded approaches to technical communication research and contrasts those with the affordances and constraints of corpus analytic techniques. We take the concept of a “theoretical framework” to discuss how to use literature and our experiences to frame research questions that are answerable through corpus analytic means. The balance of the chapter provides an overview of question types that one can ask of corpora. We review research in technical communication that attempts to answer similar kinds of research questions through corpus analytic means in order to highlight different methodological decisions that researchers might make. We conclude with a discussion of how to answer these research questions by relying on corpora to approach the issues inductively or deductively.

Chapter 4 takes up the issue of corpus construction. Just as a good research project requires careful selection of research participants and/or thoughtful and

purposeful selection of texts for close analysis, corpora must also be cultivated with questions of representativeness, validity, and reliability in mind. The chapter first grapples with the issue of how to create a representative corpus and what representativeness means. We then discuss ways of building corpora through automated and non-automated ways, including the associated ethical issues. The chapter concludes with additional preparatory steps one might make to a corpus to prepare it for analysis, including annotation.

Chapter 5 explains technical aspects of the research infrastructure needed to complete corpus analysis. The chapter gives a brief overview of the capabilities of several corpus analysis tools and information on how to select the appropriate tool for a research project. We turn then to the process and ethics of gathering and sampling data. We conclude with a discussion of how to answer these research questions by relying on corpora to approach the issues inductively or deductively.

Chapter 6 offers a reflective demonstration of corpus analysis techniques applied to a question in contemporary technical communication scholarship: writing style in topic-based documentation. We present the chapter as a stand-alone study of technical communication that benefits from tackling questions at the level of whole corpora. In contextualizing and setting up the study, we reflect in a meta-discursive way about the nature of the problem (i.e., what is the style of topic-based writing) and why it is best answered through a comparison of corpora. We then walk readers through the analytic design, including meta commentary about methodological choices. We carefully and explicitly draw findings from the two study corpora: topic-based and book-based writing. We demonstrate how to carry out the analysis and document the findings with evidence drawn from the corpora.

Chapter 7 concludes the book. We turn our attention to concrete steps that can help develop corpus analysis as a legitimate and mature tool for knowledge creation in the field. We then discuss issues regarding field-level resources to ensure that the relatively challenging startup cost of corpus analysis can be offset by strategic moves as a field that would provide communal resources for supporting this kind of research. We close by arguing that the next step in the maturation process for technical communication is to further enter large conversations about interdisciplinary and transdisciplinary research problems via the legitimating force of big data (via corpus analysis).

Ultimately, the goal of the book is to build on the field's existing work in corpus analysis and present the currently specialized study of corpus analysis to a larger audience of technical communication scholars. This book is intended as a guide that helps scholars imagine how their work could be enhanced or aided by corpus analysis. This book does that by offering readers a window into the different steps of the process in corpus analysis. Each of these topics in the upcoming chapters can be studied in much greater length elsewhere: omnibus sources such as the 754-page Routledge Handbook of Corpus Linguistics (O'Keeffe

& McCarthy, 2022) offer a wider array of concepts, while specialized resources like *What's In a Word-list?: Investigating Word Frequency and Keyword Extraction* (Archer, 2009b) or *Corpus Annotation: Linguistic Information from Computer Text Corpora* (Garside et al., 2013) offer much more depth on individual topics.