Chapter 2. Designing the Analysis

In this chapter, you will develop a design to guide your analysis. Based on what you know of your phenomenon, either from a literature review or from exploration at the site, you will build a framework to determine what is of interest and how to study it. You will articulate your questions, build a descriptive framework, decide how to focus your analysis, build in a contrast for comparison, and then make decisions about how to sample cases.

Writing Memos

Throughout the book we will be asking you to document your thoughts, analyses, and investigations using a memo-writing process. As Saldaña (2009) discusses, there are many different purposes and occasions for writing memos, but what they have in common is reflection. A memo is a conversational moment with yourself, an opportunity to take stock of how your analysis is taking shape. At various moments, you can use memos to reflect on emerging themes, patterns in the data, potential points of significance in the analysis, problems, and solutions. Eventually, memos can lead you to an analytic design, a coding scheme, or an analysis.

Throughout this chapter and all chapters in this book, we will offer you memo prompts to document your analytic process. The prompts will include questions that we find helpful to consider, whether to make decisions about

the study design, to remind yourself of methodological choices, to make educated guesses about analysis, or to begin drafting part of your final write up. Of course you should also memo yourself at any point and on any question or thought that seems important.

Aside from options supported by Excel and MAXQDA (see Excel Procedure 2.1 and MAXQDA Procedures 2.1 and 2.2), the simplest approach would be to start a word processing document and keep your memos in a single file. Divide your memos into sections, using subheadings that correspond to the "Write a Memo" sections that you find in this chapter and throughout the book (e.g., Memo 2.1 Descriptive Framework). Another option for writing memos could be to create a file folder with subfolders that correspond to different phases of the verbal data analysis process (e.g., design of analysis, sampling, coding, analysis). If you are inclined, a more creative option would be to memo with a tool like WordPress, which allows you to insert labels to use for filtering memos at a later point.

Excel Procedure 2.1: Creating a Memo

https://bit.ly/2kL7ATv

1. Place your cursor in the cell to which you want to add a memo.
2. Select **Insert > New Comment**.
3. Resize the comment field to give yourself a visible field on which to write.
4. Write your memo and press enter to finish.
5. Mouse over the cell to see the memo.

MAXQDA Procedure 2.1: Creating a Memo

https://bit.ly/2kL7ATv

If you intend to use MAXQDA for your later analysis, you may use its free memo function to document your analytic processes. Unlike other memos in MAXQDA, free memos are not associated with specific locations in your data, but are general to your project as a whole. Once you create a free memo on some aspect of your analytic process, you can easily locate it, update it, and export it to be included in a later description of your methods.

1. Select **New Free Memo** from the **Analysis** menu in the toolbar.
2. Select a memo type, if desired (see Figure 2.1).

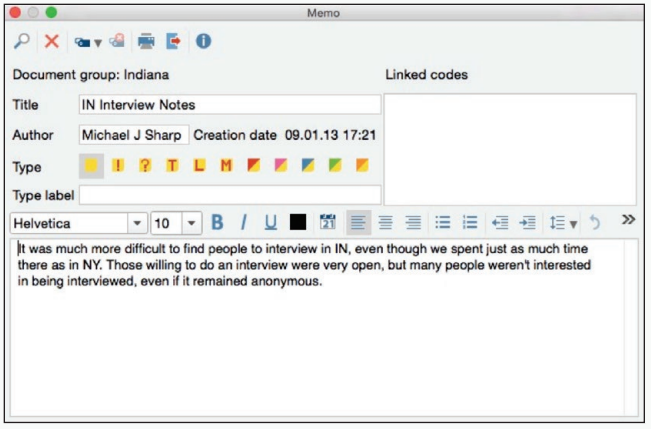


Figure 2.1: Writing a memo in MAXQDA.

3. Give your memo a title and write your thoughts.

MAXQDA Procedure 2.2: Reviewing a Memo

1. Open the **Overview of Memos** command by clicking on the **Overview of Memos** icon in the toolbar.
2. Click on the left-most table column to sort by type.
3. Click a specific memo to open and edit its contents.
4. If desired, click the **Export** icon to export your memo content to either an .rtf or an .HTML file.

Designing the Analysis

To get at the importance of an analytic design for doing verbal data analysis, it can be helpful to start with a polarized picture of different research approaches. In an over-simplified world, there are two kinds of research approaches: quantitative and qualitative. Each carries its own analytic assumptions that, in turn, shape how researchers envision their phenomena of study, ask questions, and decide how to acquire data.

Quantitative studies are largely based on empirical data about phenomena that researchers assume to be real and "out there" in the world. Quantitative researchers design their research by defining an object or problem to study, framing it in terms of a theory and associated variables, and testing it in some way, whether through an experimental or survey design. This approach sets constraints on the kinds of questions to be answered and the kind of data to be used in answering them.

Qualitative studies are also based on empirical data about phenomena in the world. However, those phenomena are taken to be constructed rather than objectively real. In this sense, constructed is commonly understood to mean constituted in and constitutive of a discourse. In other words, qualitative researchers study the discourses through which phenomena come to be (see Berger & Luckmann, 1991). For example, studying the phenomenon of online social activism might entail looking not only at the comments from community organizers who see their actions as social activism but also at other discourses that portray the same activities as civic disruption. A qualitative researcher designs her analysis to be exploratory, inductive, and emergent, and these qualities are reflected in more open-ended research questions whose answers require rich information gathered through case studies, ethnographies, and other descriptive studies. As a result, the scope and amount of data collected can be vast, often more than can be used productively.

The verbal data studies discussed in this book take a middle road and reflect more of a mixed-methods approach. Studies of verbal data are empirical and focused on the discourses that constitute or are associated with the phenomena we wish to study. However, we attempt to apply an analytic framework that

allows us to ask more precise questions and make judicious selections of data that are sensible within that analytic frame. The result is that a mixed-methods approach benefits from the analytic constraint of quantitative approaches while utilizing a data source that supports a more nuanced understanding of our phenomena. Yet to achieve this benefit of a mixed-methods approach, a clear analytic design is a necessity.

More is at stake in designing an analysis than just identifying a phenomenon of study. How you choose which streams of language to analyze and how you construct the comparative frameworks in which the analysis takes place will form the foundation of your study's credibility and applicability. You must be able to explain how the streams you have selected are related to the questions that drive your study. You must be able to articulate your process of sampling. And, in most cases, you must show how the comparisons you make are meaningful and relevant to the issues at hand. To meet these challenges, you must design your analysis.

Much of the advice in this chapter can be employed at one of two stages in a project. A first point comes when you have gathered your data and need to develop a strategy to direct your analysis. The data you have gathered may be comprehensive. Perhaps you have tried, to the best of your ability, to collect all texts and to tape record all interactions. Or the data may come from a preexisting archive, whether it be the paper archives of a historical collection or the electronic archives of a chat room. In all these examples, you have more data than you can possibly analyze in a reasonable period of time. You may focus your analysis by examining your data through a theoretical frame.

Another possible point for using the advice in this chapter follows an exploration of the data, a time when you have entered into a situation, become familiar with it, and seen something interesting. At that point you have developed a sense of what is interesting and worth exploring as well as an appreciation of what that data looks like and where it can be found. At this point, you may construct a design using the techniques in this chapter in order to guide your data collection. This kind of early design will not relieve you of the need to refocus your design when you come to the stage of data analysis, but it may considerably reduce the amount of data you face when analysis time rolls around.

Focusing on a Phenomenon

Before designing the analysis, you need to focus in on an object or phenomenon of study. Sometimes a phenomenon of study will present itself vividly and the exigence and audience for the study will be immediately apparent. Other times, a phenomenon of study might start from something that you have read or may arise out of an inkling that something interesting is going on at some site. Spending some time up front, focusing on what you are interested in studying, can help you develop a more focused analysis. There are two common and productive ways of focusing a study: reviewing the literature and exploring a site.

Reviewing the Literature

For many researchers, the phenomena they study arise from their review of the literature. You might start with some ideas of what you want to study, but the shape and significance of those ideas will come into focus as you situate a phenomenon in the literature of your field.

By reviewing the literature on or related to your phenomenon of interest, you gain awareness of it as a theoretical phenomenon: something to which others in your field have previously attached ideas and beliefs in an attempt to explain that phenomenon and articulate principles about it. This theoretical framework becomes clearer as you read the conversations that have helped define the phenomenon you want to study. A review of the literature can tell you the questions that others have asked about the phenomenon, the settings in which they have studied it, and questions that remain unanswered. Knowing what has been said already can help you determine what still needs to be said and how to talk about the results of your research in a way that fits into the conversations that people are already having about it.

Ideally, your review of the literature should allow you to do what John Swales (1990) called "create a research space" where the purpose is to:

- state why the phenomenon is important and worth studying,
- establish what is known or understood about the phenomenon, and
- articulate what is unknown or uncertain about the phenomenon.

Your contribution will be to occupy the space that you created by identifying what is unknown and uncertain. For example, if you are interested in studying legal mediation practices, a review of the literature might point to questions about the effectiveness of different strategies that you could test or observe in existing data. By reviewing the literature, you will have also acquired a sense of how to think about your phenomenon, including concepts, theories, heuristics or other analytic frames that you can develop, extend, or refute for your analysis.

Exploring the Data and Site

A second approach to focusing on a phenomenon of study is to explore it. While it is important to appreciate your phenomenon of study theoretically, you also need to build an empirical appreciation. Unlike the theoretical frame that you build through a review of the literature, an empirical frame is built up through experiences with the phenomenon and the sites where it is found. What does the data sound like? How is it received? How is it used? How formal or informal is it? There are no practical limits to the amount of exploration you can do.

It is possible to explore a site of research before identifying a phenomenon of interest or before situating that phenomenon in the literature. In this free exploration, your goal is to explore because your intuition tells you that something might be important or interesting. For example, in Jason's study of user forum traffic for open source science software, he noticed that the participants would often preface issues by talking about problems with technologies other than the one the forum was set up to support. These apparent digressions struck him as important to understanding the bigger picture of user support for open source science software. By approaching the study design in this way, you may not know what is going on at the site, what objects are available to study, what comparisons are meaningful, or what might be interesting cases to study. Even so, you could uncover a point of interest that had not previously been anticipated in the literature.

You can also explore a site in a way that is guided by your developing theoretical understanding of your phenomenon. Guided exploration allows you to enter a site with an idea of what you want to look for and what it might mean. Returning to the previous example, Jason recalled literature on networking and technological ecologies which suggested that the apparent digressive conversation about

other technologies might, in fact, be sketching a picture of the technology as a collection of networked technologies, where the user problem is not located in any one technology but is instead located across functionally-linked technologies. Your goal would be to observe where your phenomenon occurs, under what conditions and with what participants. This guided exploration will help you develop a sense of how to recognize and measure your phenomenon of interest.

Exploring with AntConc

One way to explore your data that is relatively simple and cost effective is to use any number of concordancing tools for examining the contents of your data. AntConc (http://www.laurenceanthony.net/software/antconc/) is one such tool. While AntConc is relatively simple to learn and use, it is too complicated to explore fully here. Instead, let us suggest two features that would be useful for guided and free exploration of your data: keywords and cluster analysis. We will discuss both tools again in Chapter 4 on coding.

Assuming that you have data in text form, save a portion of that data in a plain text (.txt) format. If your exploration of the data has already pointed to portions of the data that might be interesting to study, save those portions as separate .txt files. What you want are two collections of data: 1) a collection of the entire corpus of data, and 2) a collection of the subset of data that you wish to isolate for further study. For example, you may have data that includes the transcripts of a set of design meetings. All of the transcripts will be your first collection. Your second collection should be a selection of interesting transcripts taken from the entire collection. Procedure 2.1 shows how to carry this out.

Another exploratory analysis that may help you design an analysis is a cluster analysis, or an analysis of phrases and word groupings around a search term (see Procedure 2.2).

A related exploratory analysis is one where you simply want to see the most commonly occurring phrases in your data set (see Procedure 2.3). Here you are exploring units of language that offer slightly more context—just how much is up to you. The N-Grams tool, which is in the same tab as the Cluster tool, allows you to set the "N" or phrase length (e.g., 4-gram, 5-gram) and return a list of phrases within those parameters.

Procedure 2.1: Locating Keywords with AntConc

https://bit.ly/2kL7ATv

1. Select **Settings** > **Tool Preferences** > **Keyword List**.
2. At the bottom of the window you will see an option to upload a directory. Click **Add Directory** and navigate to the folder containing your collection containing all data.
3. When the files appear in the text field, click **Load** and then **Apply**.

This collection will become your reference corpus, the collection of what constitutes normal discourse in this setting. You will use the corpus to compare with the collection of data you set aside as potentially interesting.

4. Click **File** > **Open Dir** and navigate to the folder containing the collection that you want to analyze.
5. Click on the **Word List** tab and click **Start** to create a basic list of words appearing in your data set.
6. Click the **Keyword List** tab then **Start** to see a list of keywords.

The results show you what appear to be keywords in your second collection based on their "unusual frequency" in your data (Scott & Tribble, 2006).

7. Scroll through the keywords to see if some with higher keyness rankings are of interest.

Procedure 2.2: Exploring the Data with Clusters

https://bit.ly/2kL7ATv

1. Click on the **Clusters/N-Grams** tab.
2. Set the **Cluster Size** to the minimum and maximum you want to have returned.
3. In the search field, type a word that you feel is important to your analysis and click **Search**.

AntConc will return a list of phrases in which the searched word appears, sorted by frequency.

4. Click on each result to see the phrase in context.

Procedure 2.3: Exploring Data with N-Grams

https://bit.ly/2kL7ATv

1. From the **Clusters/N-Grams** tab, check the **N-Grams box** and set the **N-Gram size** to the minimum and maximum size N-grams you wish to receive.
2. Click **Start**.
3. Click any N-Gram to view it in a larger context.

Constructing a Descriptive Framework

Having focused your phenomenon, you probably have some sense of what is going on: who's involved, what they do, what resources they have available, and how things shift over time. To design an analysis, you begin by articulating that knowledge through what Miles and Huberman (1994) have called a descriptive framework: The descriptive framework is the first step toward explaining what you think is going on. It is your conceptual representation of the components of the verbal phenomenon you are interested in studying. At the same time, the framework is a decision about which components of that verbal phenomenon are significant for your study.

A few years ago, for example, Cheryl became interested in a senior capstone design course in mechanical engineering. She had explored the course in her role as the director of the writing intensive program of which this course was a part. The descriptive framework in Figure 2.6 is a graphical representation of its weekly events and their participants. Eventually, she had gathered a great deal of data about each of the components of the course:

- For each of fifteen weeks, teams of students met with the instructors in the course for a 1 ½ hour class meeting: she had recorded these class meetings.
- In addition, each team met twice, once with one of the instructors and a second time on their own: she had recorded the team meetings and gathered all of the texts from four of these teams.
- Each student also prepared work on their own. To track this work, she collected process logs from all of the team members. She also met for weekly interviews with a team contact as well as one other team member on a rotating basis.
- Finally, she knew the instructors attended a 1-hour staff meeting each week, in addition to whatever course preparation they did on their own. She attended and recorded each of these staff meetings and held a longer interview with the instructor at the beginning and end of the course.

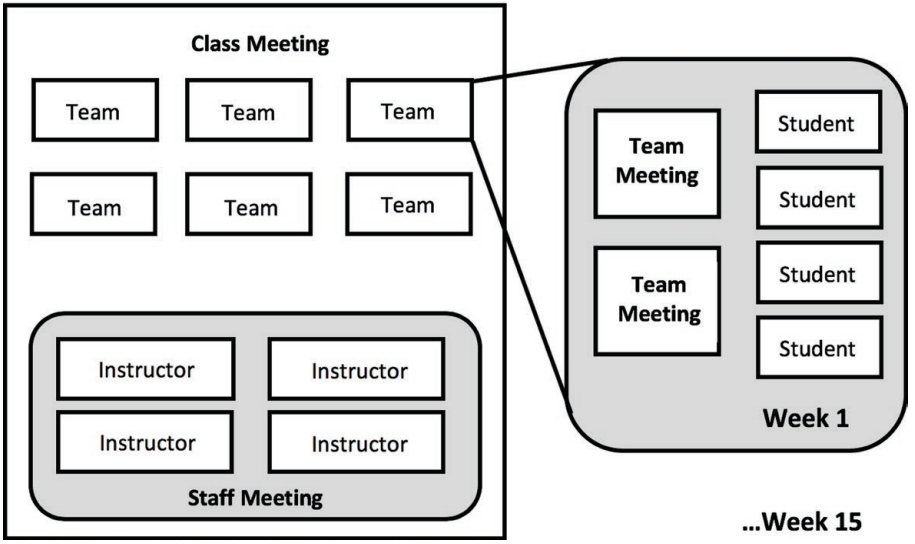


Figure 2.6: A descriptive framework for a capstone design course.

A good descriptive framework will allow you to focus on the major verbal events and their relevant attributes. In the descriptive framework in Figure 2.6, for example, three events have been identified: class meetings, team meetings, and staff meetings.

Second, a good descriptive framework will identify the relevant participants in the events making up the phenomena. In Figure 2.6, for example, we see that students attend the team meetings as well as the class meetings; that instructors join the students in class meetings but also meet on their own in staff meetings.

Third, a good descriptive framework will specify significant relationships among its entities. It will, for example, show the categories of which participants are members—such as the teams in Figure 2.6. It can also indicate other kinds of relationships such as hierarchy, opposition, and association—that an instructor "mentors" a team, that teams "compete" for the best design, that team members may "belong to" the same fraternity.

All of these relationships might be characterized as spatial because they describe the interconnections among various entities in the spatial dimension.

In addition, a good descriptive framework will have a temporal dimension. It should, for instance, indicate temporal routines—as Figure 2.6 shows how the class meeting, staff meeting, and team meetings combine to make up a weekly routine that repeats itself fifteen times during the course of a semester. Or, it might describe temporal change—that, for instance, a design team moves through three phases in the course of their work.

A good descriptive framework can also indicate which attributes or characteristics of the data will be of potential interest for your analysis. Attributes about the verbal phenomena might include length of sessions or syntactic complexity in team discussions. Attributes about the participants might include gender, experience, and frequency of contributions. Attributes about the contexts could include access (whether public or private) and location. Any of these attributes could influence your analysis, but not all of them will. Regardless, you should pay attention to the attributes suggested by your framework.

Exercise 2.1: Test Your Understanding

The descriptive framework in Figure 2.6 does a good job of indicating the components that shape the phenomenon. The framework is rather coarse in its depiction, focusing only on the simplest interactions and contexts. It does not focus on more specific spatial relationships between participants (e.g., hierarchy, opposition, affiliation, cooperation) that might be important in understanding the phenomenon. Modify this diagram, available on the book website (https://wac.colostate.edu/books/practice/codingstreams/), to include one or more of these additional relationships.

For Discussion: Is it possible for a descriptive framework to represent all of the relationships in the universe to be studied? If not, how can a researcher choose what to include and what to leave out in her descriptive framework?

Memo 2.1: Descriptive Framework

Construct your own descriptive framework and reflect on the attributes of the participants, settings, relationships, and data that might be significant in understanding your phenomenon.

Articulating Research Questions

At this point in the analytic design, you have learned a bit about your phenomenon by exploring it through the published literature and by examining the site where your phenomenon is found. You have sketched out a descriptive framework that gives a picture of verbal events, participants, relationships, and contexts that make up your phenomenon of interest. In doing so, you have been developing a tacit sense of what you want to study and how. Now is the time to be more explicit about those aims by articulating research questions that will drive your analysis.

Before articulating questions directly, take a moment to think about the aims of your research project. Most research projects have one or more of the following aims: to define or describe something that exists, to describe a relationship between variables, or to understand how one thing causes another. When studying verbal data, your aims will primarily be descriptive because of limitations on the amount and type of data that you can collect and analyze, as well as limitations on the contexts of study, make inferring causal relationships troublesome.

One source of research questions is your own curiosity and need to know. Look at your descriptive framework. What is it you want to know about this phenomenon? Is there something you suspect that is going on here? Is there something you feel a need to know more about? A second important source of questions is the literature you may have studied. What does the literature suggest is going on here? What gaps in the literature can be addressed by an investigation of your phenomenon?

These initial steps should help you arrive at a big research question like "How do students collaborate?" or "How are decisions made?" These large questions are what Creswell calls "grand tour questions" (1994, p. 70) and they are a necessary step in refining a set of research questions that are more directly and concretely answerable. A grand tour question is the overall question that you want to answer, but it is too broad to answer directly. Instead, you must come up with questions that have more concrete outcomes, that in answering you can speak to the grand tour question

In general, there are three kinds of questions that can be answered by the kind of descriptive analysis you will take forward in this book:

Questions of Kind: What kind of thing is this? What is it made up of?

Questions of Association: What is this thing associated with? When this occurs, what occurs with it? What is absent?

Questions of Time: How does this thing vary over time? What are its routines? How does it evolve?

As you articulate these questions, you might be tempted to include a fourth kind of question, a question of cause, to drive your research. Be careful about this. Strictly speaking, descriptive analysis cannot give a definitive answer to questions of causality: did this cause that? But we can make some headway on causality with questions of association—is this associated with that?—because there can be no causality without association. And we can also go some distance toward causality by using questions of time—does this precede that?—because there can be no causality without precedence. Thus, if you find yourself wanting to ask questions of causality, try to rephrase them as questions of association or questions of time.

While questions of kind, association, and time are straightforward in definition, it can be challenging to decide which questions to ask. One approach recommended by Booth, Colomb, and Williams (1995, pp. 39-41) is to begin with a list of every who, what, when, where, and how question you can think of, answers to which would get you closer to answering your grand tour question. Then categorize these questions as shown in Table 2.1.

Table 2.1: Mapping question objectives to question types

Questions defining parts/wholes	Questions of Kind
Questions identifying categories and characteristics	Questions of Kind
Questions about values and uses	Questions of Association
Questions about history and changes	Questions of Time

Memo 2.2: Research Questions

What are your question types and what kind of data would you need to collect in order to answer those questions? How definitively could you answer those questions? Where is that data to be found?

Defining the Focus for Analysis

From the universe of data mapped out by a descriptive framework, you need to select one or more foci for further analysis. The focus defines the object at the center of your research questions as well as the streams of language that you will use in your analysis. For instance, if you want to answer the question, how does a design evolve over time? appropriate answers would involve the object "the design"—thus, "the design" would be your focus.

The kinds of objects you may take as the focus of your analysis can vary considerably. Many analyses take the individual as its focus. Such analysis asks, "What has this individual been doing?" Using a focus on the individuals in the capstone design course, we might decide, for example, to select the following data to analyze:

- all of the texts written by an individual,
- all of the interviews with the individual, and
- all of the contributions the individual made in class and team meetings.

Other analyses focus on certain kinds of events. Such analysis asks, "What happened here?" In the data set for the capstone course, for example, we might focus on the team meeting as an event and pull the following data for analysis:

- transcripts of all of the team meetings,
- selections from interviews in which the team meetings are discussed, and
- all texts used during the team meetings.

And some analyses focus on specific activities that occur in the situation. Such analysis asks, "What gets done here?" It cuts across the individuals and events in a situation and may even involve other quite different situations. In the capstone data, for example, we might focus on the activity of engineering design and select the following data for further analysis:

- all of the texts and sketches that a team constructs for a design,
- all of the segments of team meetings in which the design was discussed,
- all discussions in the staff meetings about students design work in general, or a particular team's design, and

- interviews with the staff member responsible for mentoring the team through their design work.

By choosing a focus for analysis, you make a commitment to analyze a certain phenomenon and to discuss that phenomenon in reporting your results. Such commitments need not be final or exclusive. That is, taking one focus for analysis for your current project does not preclude you looking at the data with a different focus later on. One of the strengths of a approach to research is, in fact, that the data it produces is rich enough to sustain a variety of analyses. Making a choice about focus now just allows you to isolate the streams of language in which you are more likely to find the phenomenon that you want to study.

Building in Contrasts for Comparison

Before selecting a sampling scheme, there is one further consideration that is critical to the analytic process: choosing a built-in contrast for comparison. A built-in contrast allows you to examine your focused phenomenon in relationship to other phenomena that you take be *a priori* different and through that comparison to focus attention on qualities of the data that highlight those differences. Such contrasts become essential to shaping the coding and pattern detection we describe in later chapters. As it pertains to sampling, building in a contrast will help you think about what data you need to support the comparative analysis you are building.

It is through your choice of a sampling scheme that you build in the contrast. For example, if you have chosen to analyze a stream of language because you think it offers an idealized look at the phenomenon of interest, search for another stream that has a high probability of not being very good. If you have chosen what you think is a typical stream of language, look at the periphery of your data to search for streams that are less than typical. A failsafe strategy is always to include what appear to be negative streams or atypical streams for analysis.

Ideally, your contrasts come from the same data set as your core data. Streams of language which come from the same data set but exhibit contrast

help to define the boundaries of a phenomenon in a way that streams outside of that data set cannot. If, for example, we find the instructors in the capstone course consistently use sketches in ways different from the students, despite sharing a lot of the same context: the same course, the same design project, the same university, even the same field, then we may be on to something.

Contrastive streams may come, as in the comparison of instructors versus students, from variations in the spatial dimension, such as different sections, or the presence of supporting technologies. Do the student teams use different tools when sketching and what variations in the sketches might those tools explain? Other comparisons may be possible by looking for temporal contrasts. For example, do sketches vary significantly from the ill-defined early stages of design to the final stages of specification?

The source of appropriate comparisons often comes from the literature that you used to guide your study. Does the literature take a certain situation as paradigmatic, typical, desirable? Can you build in a contrast with your data set? Or, if such a contrast is not available in your own data, can you find data elsewhere that might make an appropriate comparison? Could we compare, for example, the ways students use sketches to design with the way they are used in the published literature in engineering?

Sampling Streams of Language

Once you have decided on a focus for your analysis and a contrast for comparison, you will need to decide how to sample among the language streams that your site presents. In almost all situations, you will have more than one choice. In the situation diagrammed in Figure 2.6, for example, if you decide to focus on students, which students? If you decide to focus on team meetings, which team? If you decide to focus on a team's design activity, which design? If you decide to focus on a whole course, why this course?

Convenience Sampling

One of the most commonly used strategies for sampling streams is also the least defensible: convenience. Using convenience sampling, we might choose a

student because he sits next to us. We might choose a set of meetings because they occurred at a time when we can easily attend. We might choose a certain team because we already know some of the members. We might choose this course because it was one we already knew about. Sampling by convenience, as these examples suggest, puts personal considerations ahead of other consideration that might be relevant to your study.

If convenience is your only answer to a question about sampling, you will lose a great deal of credibility and possibly miss out on finding your phenomenon of interest. By the same token, however, convenience is almost never totally irrelevant in the design of a study. If your desired focus is difficult to access for whatever reason, you may need to consider what is possible for you. If access depends upon a history of interaction in a site that is difficult and costly to build, you may want to rely upon rather than abandon what you already have access to.

Snowball Sampling

Related to the convenience sample is a snowball sample and next to convenience sampling, it is one of the more popular and widely used techniques. In a snowball sample, you gather the data that you can and then work from those data sources to find other sources. Often this sampling entails working through participants who can introduce you to other participants who are similar. Snowball sampling can be useful if you are studying phenomena that are difficult to locate or are found within populations or settings that are difficult to access. In our example, we could gather a snowball sample by first finding a student team willing to share data with us and then asking them to introduce us to friends and fellow students who might also be willing to participate. As with the convenience sample, the snowball sample has similar limitations, but sometimes the networks of affiliation that participants use to create the snowball sample may be relevant to the analysis.

Typical Case Sampling

Can you pick a stream or streams that are typical in your site: a typical student,

a typical team, a typical meeting, a typical course? To use the strategy of typicality you will need to have some kind of data available about the range of relevant variation across streams of language at your site. If most of the teams in the capstone design course, for example, are made up of both men and women, we may want to make sure to pick mixed-gender teams for my cases. Not all variations are relevant, of course: if we find out that most of the students in the course own cats, we may still feel we do not have to worry about whether the students on the team we choose are cat-owners.

Extreme/Deviant Case Sampling

A counter point to the typical-case sample is to choose outliers or extreme cases that show the range of conditions where you would find your phenomenon. Sometimes it can be beneficial to your analysis if you can show how the phenomenon differs from what is typical by examining the unusual circumstances in which it occurs. In the example of the capstone course, one could sample by looking at only the teams with the highest scores or the teams with the lowest scores.

Best Case Sampling

If the phenomenon in which you are interested only occurs in some streams, you may want to employ a sampling strategy that maximizes your chances of finding it. If, for example, you are interested in describing successful design activities, and you know that about a third of the teams in any capstone design course will not be judged successful by their instructors, you may want to try to find a way to sample teams with a high probability of being successful.

Criterion-based Sampling

Best case sampling is a special variety of the more general strategy of criterion-based sampling. With criterion-based sampling, you specify a certain relevant criteria and choose all streams that meet that criteria. If, for example, I

want to study mixed-gender communication patterns in student design teams, I might decide to study all of the mixed-gender teams formed in a particular semester at a particular university.

Stratified Sampling

With a stratified sampling strategy, you take advantage of knowing something about the existing variations in a site. If you know, for example, that in a certain site, most design teams are either all male or all female, but that a few are mixed gender, you may want to make sure that you study a certain number of each of these three kinds of teams: male-only, female-only, and mixed. If you know that some design teams succeed and others fail, you may want to make sure to interview students from both types of teams.

Random Sampling

With random sampling, you chose streams based on the patterns established by a randomly generated sequence of numbers. Using a random sampling strategy involves three basic steps (see Excel Procedure 2.2).

Using the example of student teams, we can easily use Excel to build a random sample to include in our analysis. An alternative to working with Excel to generate a random sample is to use an online random value generator, such as Research Randomizer (https://www.randomizer.org/).

Comprehensive Sampling

The final strategy may not seem like a strategy at all yet, in some situations, it is appropriate. In comprehensive sampling, you choose all of the streams available to you in a site. If there are 50 teams, you analyze 50.

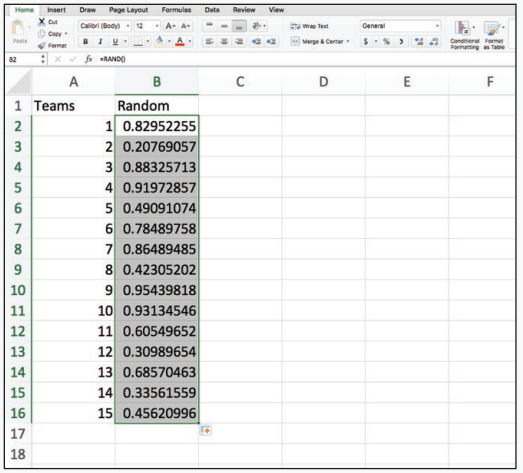
Excel Procedure 2.2: Generating a Random Sample in Excel

https://bit.ly/2kL7ATv

1. Open a blank Excel worksheet.
2. In the first column enumerate the universe of samples.

For example, with 50 teams we would type 1–50 in the first column.

3. In the second column type =RAND()
4. Drag this formula down to generate a random number for each sample (See Figure 2.7).



	A	B	C	D	E	F
1	Teams	Random				
2	1	0.82952255				
3	2	0.20769057				
4	3	0.88325713				
5	4	0.91972857				
6	5	0.49091074				
7	6	0.78489758				
8	7	0.86489485				
9	8	0.42305202				
10	9	0.95439818				
11	10	0.93134546				
12	11	0.60549652				
13	12	0.30989654				
14	13	0.68570463				
15	14	0.33561559				
16	15	0.45620996				
17						
18						

Figure 2.7: Generating a random number for each sample.

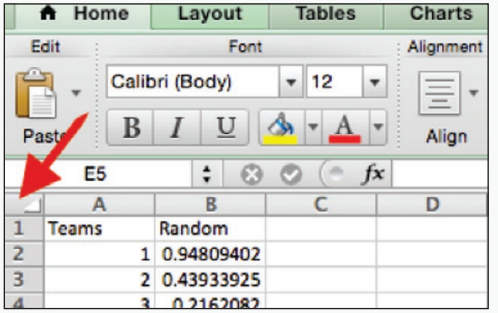


Figure 2.8: Clicking the box where the row and column numbers originate.

5. Select all of the random values and click **Edit > Copy > Paste Special** and check the option for "values." Click **Okay** to paste the values over the random values.

Cutting and pasting the random numbers as values keeps the numbers the same instead of recalculating new random values.

6. Select the whole worksheet by clicking as shown in Figure 2.8.
7. From the menu bar, choose **Data > Sort** and sort by the column containing your random values

To get a sample size of 12 teams, take the first 12 in your sorted list.

Exercise 2.2. Try It Out

#	Gender Composition	Score in Class	Coop
1	Women-Only	2	yes
2	Women-Only	6	no
3	Mixed	18	no
4	Mixed	10	no
5	Women-Only	46	yes
6	Men-only	1	yes
7	Men-only	23	yes
8	Women-Only	1	no
9	Men-only	44	yes
10	Men-only	12	no
11	Women-Only	11	yes
12	Men-only	32	no
13	Men-only	20	yes
14	Men-only	23	no
15	Mixed	36	no
16	Women-Only	11	no
17	Men-only	18	no
18	Mixed	25	yes
19	Women-Only	11	yes
20	Men-only	17	no
21	Men-only	47	yes
22	Men-only	38	no
23	Men-only	3	yes
24	Men-only	30	no
25	Women-Only	12	yes

#	Gender Composition	Score in Class	Coop
26	Men-only	17	yes
27	Women-Only	19	no
28	Men-only	35	no
29	Men-only	1	yes
30	Mixed	18	yes
31	Women-Only	14	no
32	Men-only	15	yes
33	Men-only	11	yes
34	Men-only	39	no
35	Mixed	49	yes
36	Men-only	49	no
37	Men-only	14	yes
38	Women-Only	4	no
39	Men-only	34	yes
40	Men-only	43	no
41	Men-only	37	no
42	Women-Only	14	yes
43	Men-only	36	no
44	Men-only	18	no
45	Men-only	8	yes
46	Women-Only	19	yes
47	Women-Only	11	no
48	Mixed	46	no
49	Women-Only	37	no
50	Men-only	44	yes

Figure 2.9: A sample of 50 cases.

Suppose I have gathered data from 50 teams in the order shown in Figure 2.9. Complete the table on the following page, available on the book website (https://wac.colostate.edu/books/practice/codingstreams/) showing the gender characteristics of the following samples:

- A convenience sample of 12.
- A typical case sample of 12, with typicality defined by gender.
- A best-case sample of 12 with "best" defined as top-scoring.
- A criterion-based sample of ten using the criterion of experience.

- A stratified sample of 12 stratified by gender composition.
- A random sampling of 12 (use the following random number sequence: 16, 50, 36, 9, 25, 1, 5, 27, 19, 29 or one you generate on your own using the directions in Excel Procedure 2.2).
- An extreme case of high and low performing.

Kind of Sample	Teams Sampled	# of men-only teams	# of women-only teams	# of mixed-gender teams
Convenience				
Typical				
Best-Case				
Criterion-Based				
Stratified				
Random				
Comprehension				

For Discussion: Given what you find, discuss the benefits and costs of each strategy.

Choosing a Sampling Strategy

Deciding what sampling strategy or combination of strategies to use is more an art than a science. The guiding principle is to choose the strategy that will best support the credibility of your eventual analysis and support your checking of that analysis with contrastive cases. Obviously, for example, if you are concerned with a phenomenon that is not widely encountered, you may want to use some variation of criterion-based sampling. If you wish to describe a best-case phenomenon, it makes sense to use a best-case sampling technique. But if you want to show how a phenomenon is distributed over the range of your data, you would not want to restrict yourself to best cases: a stratified, random, or even comprehensive sample may be best.

Your choice of a sampling strategy should be driven by three concerns. The first is whether the sampling strategy you choose will ensure that you select enough data to observe the phenomenon you are interested in studying. The second is that the sampling strategy corresponds to and represents the larger phenomenon that you are trying to study and draw conclusions about. The third is that the sampling strategy includes cases that productively contrast with those you are interested in analyzing. Sometimes, these goals can be in conflict. For example, if we choose a sampling strategy to maximize chances of seeing a particular design activity, we may end up with a sample that over-represents one kind of student team. However, you may find that it is necessary to weight the sample in order to reveal more of your phenomenon. Just be aware of the potential bias of representation and temper your conclusions accordingly.

To make this discussion of sampling choice less abstract, consider the capstone example again. Suppose we want to describe the ways that sketches are used in the engineering design process. And suppose we choose an example of the use of sketches from the opening lecture by the primary instructor, a lecture which stuck in our mind for the skillful and interesting way that the instructor used sketches.

When we complete the analysis and write up the results for the readers, what could we conclude? Minimally, we could suggest that the patterns in this stream are examples of the kind of patterns that may occur when engineers use sketches to design. If we were to go further, however, and claim that these patterns were somehow typical, somehow best, or somehow characteristic of professional practice, the credibility of our claims would be undermined by our casual approach to the sampling strategy that lead to the stream to begin with, for we had only picked an example that stood out in our minds without considering what made it exemplary—or typical, and so on.

A little more consideration of sampling would have gone a long way toward putting the analysis on more solid ground. We could have, for example, used a comprehensive strategy to select transcripts of all interactions that involved the use of sketches and then tried to describe the stream in relation to that comprehensive sample. We could have done a stratified sample, choosing to look at a some streams produced by students in team meetings compared

to streams produced by instructors in class meetings. We could have used a random sample to pick 10 streams of sketch use. Any of these strategies might have put us on more solid ground with respect to the credibility of the results.

The drawback to this approach is that the analysis becomes that much more complex and the analysis is not so sharply focused on the phenomenon that you have an interest in studying. Each approach has its advantages and disadvantages.

The next consideration in choosing a sampling strategy is to be aware of what your selection assumes to be true about your phenomenon and its distribution throughout the setting where it is found. Strategies like a stratified sample or a random sample are systematic approaches, designed to ensure that all data points have equal probability to be selected. These approaches assume that the phenomenon of interest is either equally distributed throughout the data or equally distributed throughout the strata. Non-systematic or purposive sampling assumes that some researcher judgment is required to locate the phenomenon and isolate it for analysis. Since the phenomena tracked in verbal data analysis tend to be exceptional and not random, purposive sampling like best-case, typical-case, and criterion-based sampling may make fair assumptions about the phenomenon you are studying. In any case, the most important lesson is to have an articulated sampling strategy and to be able to explain your strategy to readers in a way that enhances rather than undermines your credibility.

Finally, be sure that your sample allows you the opportunity to select streams that contrast or give negative examples of the kind of phenomenon that you are interested in analyzing. If you want to study high scoring teams and the materials they use to support collaboration be sure to select low scoring teams for comparison.

Memo 2.3: Sampling Scheme

Consider two or more sampling schemes and reflect on how those approaches would change the shape of the data you collect. What data will be included and excluded with each sampling scheme? How might your analysis be affected?

Acquiring the Data

Once you have chosen the appropriate unit for your segmentation, you will need to apply it to your data in preparation for coding. Essentially, this will involve five steps that we cover in the rest of this chapter: acquiring the data, cleaning the data, segmenting the data, moving the segmented data into your analytic application, and then labeling the segments.

Your goal in the first step, acquiring the data, is to put the data into the word processing program that you will use for segmentation. In our examples, we will use Microsoft Word for Mac (Version 15.28), which uses a look and feel similar to Microsoft Word for Windows. You may choose to use one of these or any other word processing application.

The Ethics of Acquisition

Before you begin to acquire data, an important consideration in any research project, but especially those that involve streams of verbal data produced by people, are the ethical guidelines that govern how research participants are to be recruited and treated in the course of that research. If you are working in a university context, your institution may already have a research office dedicated to helping you consider the ethical considerations entailed in your research.

Most research offices in the United States that make determinations about human ethics base their guidelines on those recommended in the so-called Belmont Report, which was drafted in 1978, partly in response to the Tuskegee Syphilis Studies. The Belmont Report centers around three basic principles that ought to guide all research:

> **Respect for Persons:** Researchers should respect the autonomy of all research participants. People who are autonomous act in accordance with their own goals and have the right to hold opinions and make decisions in pursuit of those goals. In cases dealing with populations that have their autonomy restricted (e.g., prisoners) it is important for the researcher to allow for special protections from harm.

Beneficence: Researchers must take steps to act in the best interests of research participants. This principle asks researchers to consider a balance between, on the one hand, doing no harm to the research participant and, on the other hand, the prospect of conducting research that provides some larger benefit, even if the research participants themselves are not the direct beneficiaries.

Justice: Researchers must ensure that the risk entailed in participating in a research program does not unfairly burden one particular population, especially if that population is not the direct beneficiary of that positive outcome.

While these principles were largely developed to govern biomedical research, the precepts are written with enough room for broader application to other realms of research involving human participants, including verbal data analysis. As the guidelines pertain to your research project, consider the participants who will be offering their verbal data for your analysis. How aware are they of your analytic intentions and do they know enough about what you are collecting and analyzing to determine whether participation in your study is in accord with their goals? In other words, do your participants have enough information and do they comprehend it well enough to make a decision about whether participation in your study fits with their goals? In the case of accessing publicly available information, how aware are the contributors that their contributions may be used in your analysis? Have you articulated for yourself and for your participants what the purpose and ultimate benefit of your research will be? Will your participants receive some benefit from your research, and if not, will they at least be protected from harm or injury as a result of your research (e.g., can any direct quotes be traced back to participants in some identifiable and damaging way?)

As you can see, applying these ethical principles to research outside the realm of biomedical research requires us to stretch our understanding of concepts like participants, privacy, and benefits. Even so, taking verbal data from face-to-face contexts like classrooms and other kinds of meetings does not present insurmountable obstacles for applying the same basic guidelines for conduct. Looking for verbal data to study on the internet, however, does pose additional problems.

Because it is entirely likely that you may be interested in studying streams of language from online settings, it is useful to broaden our understanding of how general principles of research ethics (such as those found in the Belmont Report) might need to be expanded to address the challenges of doing research online. One assumption of human-subjects research is that the focus of study is human subjects. And in the case of verbal data analysis, this is often the case. However, a human subject on the internet is not so unproblematically defined as in a lab setting. On the internet, human subjects are present as information, registering as potentially personally identifiable information that someone may not be aware they are sharing. What we choose to study can, if we are not careful, capture a wider range of this data than a person might knowingly give. For this reason, we need to exercise a bit more ethical discretion when utilizing what might otherwise seem to be publicly available information on the internet (e.g., online video, forum exchanges, blog posting, tweets, customer reviews, email list contributions, etc.).

The Association of Internet Researchers (AoIR) have developed a heuristic and set of considerations for researchers to use when gathering verbal data found online. Among the considerations one should keep in mind are:

- How is the research context defined? Given that the internet is a collection of specific locales, you ought to be aware of expectations in this context.
- How do you gain access to the research context? Is the context publicly accessible or for members only? How aware are people of your presence?
- What is your ethical stance with regard to the aims of the locale you are accessing and the expectations of participants regarding how their verbal data flows?
- Who is involved in your study and what options do they have for opting in or out?
- What is your object of study and how may that object be traceable to individuals or not (e.g., direct quoting from forum postings).

Ultimately, the AoIR panel suggests that each instance of internet-based research presents its own ethical context and the factors that matter when making an ethical evaluation should be built up from the specifics pertain-

ing to that case, the context of the research, the people directly and indirectly reached, and their expectations about how their verbal data is supposed to be used (see Nissenbaum, 2012).

The internet is an altogether different kind of public space in which the very notion of public and private is complicated. While a forum or email archive might provide you public access to a stream of verbal data, it is an open question whether the people who participated in those exchanges considered their contributions private or public. Early concern over online privacy still rings true today: "the end user can never be sure [. . .] who has access to his or her information, under what conditions and limitations such access is granted, and so on" (Jones, 1994) and even in cases where the law does not specifically prevent information from being made public, it may be the participants' expectations that their contributions were private, and that expectation ought to be respected and balanced against consideration of the benefits that derive from your research. Further complications arise when we consider the various ways that you might recruit participants into a study, especially when you utilize cloud computing resources like file sharing services, online drives, and other resources, in that unintended, private information may very easily be uploaded along with the contributions that participants are aware they are providing.

By all means, do use streams of language data that are accessible on the internet. There is significant good that can come of research on information seeking practices, user assistance, crowdsourcing projects, and other creative projects. However, it is your responsibility as a researcher to be aware of the ethical issues that touch upon the work that you doing.

Please familiarize your self with the Belmont Report guidelines for human subjects research as well as the adaptations recommended for internet-based research as supported by the Association of Internet Researchers (AoIR) and always be sure to submit your research plan to your university's research ethics board to seek guidance and approval before starting.

Transcribing Audio

If you are beginning the acquisition process with audio files, you will need to transcribe them. If you have the budget, you may want to employ someone to

transcribe them for you, and there are many services that specialize in transcription that can save you a lot of time. But you should be aware that you will still need to listen to and review the transcripts to correct for errors that result from arcane vocabulary, overlapping speakers, or poor quality recordings. We have done such corrections numerous times in previous projects and can testify to the fact that uncorrected transcriptions can contain so many errors as to make their analysis hardly worthwhile.

Another route you might consider is voice recognition software. Ideally, this kind of software produces a complete transcript from an audio file. While some voice recognition software works well when trained to recognize the voice of a single speaker, it does not yet do well enough with unknown speakers or multiple speakers. For most of the audio data that we encounter, then, the quality will not be adequate.

If you are resigned to making your own transcriptions, you will want to use transcription software that allows you to adjust the speed of playback and automatically rewinds a few second each time you stop to transcribe. If you are transcribing from video, the ability to insert timestamps into your transcript will also be useful to help you sync the transcript with the video. As well, a footpedal can really speed up the transcription process, so if you are considering using one, make sure the transcription software you choose is compatible.

Some qualitative analysis software has transcription functionality built in. The Standard package of MAXQDA, for instance, supports transcription of audio and video and provides automatic timestamping that links to the media from the transcript. If you are using or have access to any CAQDAS package, check to see if it offers transcription functionality.

You may also decide to use stand-alone transcription software. The three applications listed below vary by price and by the kind of features that they offer. Note that the free application, Express Scribe, does not do automatic rewinding.

- Express Scribe (http://www.nch.com.au/scribe/) version free for non-commercial purposes
- Transcribe (http://transcribe.wreally.com) for $20 US a year
- F4transkrip for Windows or F5transkrip for Mac (https://www.audio-transkription.de/english/f4.htm) for about $45 US

Scanning Texts

If you are working with texts on paper—without digital files—you will need to scan them and then convert them into readable form with optical character recognition (OCR). Many higher-end copiers combine scanning and OCR functionality and can process many pages quickly. You may also use a standalone scanner; many of these also come with OCR software. Or if you have scanned a small number of texts, you may find that a free online conversion service like Online OCR (http://www.onlineocr.net) will be sufficient to produce readable texts.

Scanning texts can take a lot of time, especially if you have to do optical character recognition as a separate step. For this reason, if you have a lot to scan, you may want to try to organize access to a high end photocopier that can do both at once. And keep in mind that even the best scan will introduce errors that will need to be dealt with using the techniques described in the section on Cleaning the Data.

Exporting PDFs

If you begin your work with files in Adobe PDF form, you will need to convert them to editable documents. To do so, open each file with Adobe Acrobat or Adobe Acrobat Reader and save them as either an .rtf file (preferred) or a .txt file. Again, keep in mind that the conversion will introduce errors that will need to be dealt with using the techniques described in the section on Cleaning the Data.

Memo 2.4: Data Acquisition Process

Describe the format in which your data is now recorded (hand-written or typed or audio-recording? on paper or in electronic form? graphic or text?). What steps will it take to move this data into a word processing application for segmentation? Plan for the resources (people, machines, time) you expect to need in order to carry out this process. Try out the process to see if it works. Document your final data acquisition process.

Cleaning the Data

Transcribing, scanning, and or exporting your data from another form often introduces errors that you will want to correct before segmentation. Sometimes these errors arise from transcription error. More often, they are the result of format conversion errors. In either case, you should clean your data before moving forward with segmentation and further analysis.

Manual Cleaning

Manually cleaning data requires you to read through the data word by word, comparing it to the source data, to insure that it has not been altered in the acquisition process. If you have employed someone to transcribe the data for you, for example, you will want to review the transcript while listening to the audio file. Sometimes transcription errors arise because the discourse is too technical for the transcriptionist. Other times, the voice is too low or distorted for the transcriptionist to make out. Many times this kind of listening can be done in double speed, slowing down only for difficult passages. In any case, it is useful to have someone familiar with the domain and the data to review the transcript and correct for errors and omissions.

In one of our earlier projects, for example, a transcriptionist had recorded a contribution by a speaker in a focus group meeting as shown in the left-hand side of Figure 2.10. The areas highlighted contained transcription errors. The first "talked to" is a mis-transcription; the remaining errors are areas where the transcriptionist could not understand what the speaker was saying. You can see the cleaned up transcript in the right-hand column. Five new phrases have been added that significantly clarify the meaning of the conversational turn. The turn also grew by 24% after cleaning.

If you have acquired your data using optical character recognition (OCR) from a graphics file, you will often find errors from failures in the OCR conversion process. On the left-hand side, Figure 2.11 shows an original graphic file of a text with some highlighting. The conversion produced by OCR is shown on the right-hand side. Many errors have been produced and will need to be cleaned manually.

Actually I had known my advisor before I came here. I had the chance to work with him and also knew my project so when I came here, I just talked to…… so I also obtained my MS degree from another institution. So nearly …… so I guess I had …. . so the answer for your question is sometimes my advisor says that… he told me his…… So I had the chance to work on a project but there are some disadvantages, too. Your advisor sees you like a colleague, so sometimes I wanted him to be my only advisor but not be my colleague. …

Actually I had known my advisor before I came here. I had the chance to work with him and also knew my project so when I came here, I just started working on this topic so I also obtained my MS degree from another institution. So nearly I took like 20 grad level courses so I guess I had ….. enough background so the answer for your question is sometimes my advisor says that… he told me he saw me like his colleagues. So I had the chance to work on a project but there are some disadvantages, too. Your advisor sees you like a colleague, so sometimes I wanted him

Figure 2.10: A conversational turn. The right-hand column shows the transcript before cleaning. The left-hand column shows the turn after cleaning. Five phases have been added and the turn has grown by 24%.

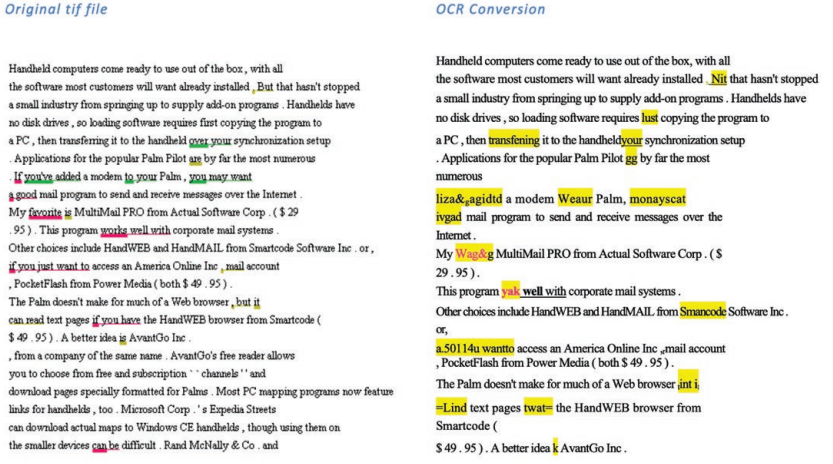


Figure 2.11: A graphic image of a text file with highlighting, shown in the left-hand column, has been read optically to produce the editable text shown in the right hand column. Many errors, shown highlighted, have been introduced by the conversion and will need to be cleaned up before analysis.

Using Find and Replace

One step up from manually cleaning your data is the use of a simple find and replace command in your word processor (see Procedure 2.4). If, for example, you find that a transcriptionist has routinely used an incorrect term, you might use a find and replace command to correct all of the errors at once. Or you could use a global find and replace to change a speaker's actual name to a chosen pseudonym.

Using Regular Expressions

A further step up in ingenuity for cleaning your data involves the use of regular expressions in your Find and Replace commands. Regular expressions are special text strings that create flexibility in search. They are sometimes called wildcards (see Procedure 2.5). For instance, if you want to search for all words starting with "th," you could search for "th*" which uses the wildcard * to match on any string. Many regular expressions exist, but only those shown in Table 2.2 are recognized by Microsoft Word.

To see how you might use regular expressions, consider a search for demonstrative pronouns, a major category of indexicals. You could use four separate searches, one for this, one for that, one for these and one for those, or you could use the single regular expression:

<([**Tt**]**h**[**iaeo**][**st**][**e**])>

Here the letters in square brackets indicate alternative acceptable matches. The angle brackets limit the search only to words that contain the full expression. [Tt] allows matches for both upper and lower case words starting with t; [iaeo] allows matches on thi . . ., tha . . ., the . . ., and tho . . .; [st] allows matches on this . . ., that . . ., thes . . ., and thos . . .; [e] allows for matches that end after the s (this) or the t (that) or those which take another e (these, those).

Another time you could use regular expressions is to clean up files scraped from the web. You may use regular expressions to find and delete HTML tags with the wildcard search string

\<*\>

Procedure 2.4: Cleaning a Transcript Using Find and Replace in Word

1. Open your transcript in Microsoft Word.
2. From the Edit menu select **Find > Replace**.
3. In the top field, type the character string you want removed and in the bottom field type the character string you want inserted as the replacement.
4. Click **Replace All** to replace all matches at once or **Replace** to replace them one at a time.

Procedure 2.5: Using Wildcards in Microsoft Word

https://bit.ly/2kL7ATv

Open your transcript in Microsoft Word.

1. From the Edit menu, select **Find > Advanced Find and Replace**
2. Expand the menu in the lower left of the dialogue box and check the box for **Use Wildcards** (Figure 2.12).
3. Type in the string with the RegEx wildcard in the **Find what** field.
4. Click **Find Next** to see the next match.

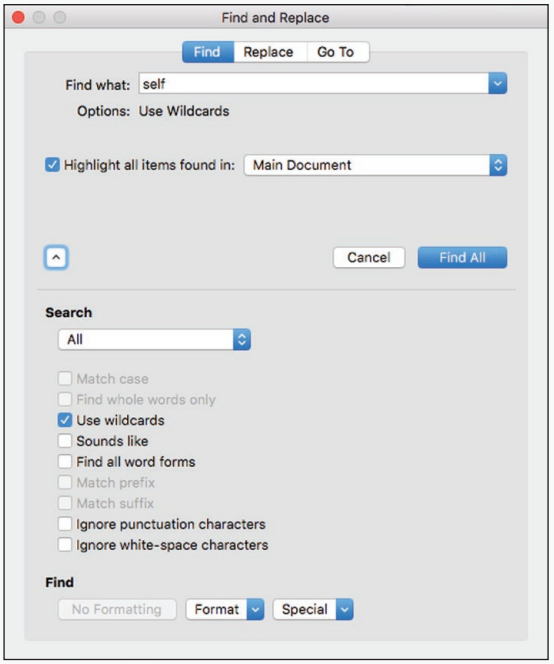


Figure 2.12: Enabling wildcards with find and replace in Microsoft Word.

which will match any HTML tag. To find and delete pairs of HTML tags along with all the characters between them, you can use the regular expression,

\<**tag*tag**\>

replacing "tag" with the name of the specific tag. To find all style tags, for example, search for

\<**style*style**\>

With this regular expression, you could find and replace all of the material between an opening style tag and its close with something like a simple paragraph return.

To preserve the content enclosed by the tags, modify the expression to

\<*\>

You may want to consult an excellent guide to using wildcards created by Graham Mayer, *Finding and Replacing Characters Using Wildcards* which can be found at http://word.mvps.org/faqs/general/usingwildcards.htm.

Table 2.2: Regular expressions recognized in Microsoft Word [1]

To find	Use this	For example
Any single character	?	s?t finds "sat" and "set."
Any string of characters	*	s*d finds "sad" and "started."
One of the specified characters	[ ]	w[io]n finds "win" and "won."
Any single character in this range	[-]	[r-t]ight finds "right" and "sight" and "tight." Ranges must be in ascending order.
Any single character except the characters inside the brackets	[!]	m[!a]st finds "mist" and "most" but not "mast."
Any single character except characters in the range inside the brackets	[!x-z]	t[!a-m]ck finds "tock" and "tuck" but not "tack" or "tick." Ranges must be in ascending order.
Exactly n occurrences of a character or expression	{ n}	fe{2}d finds "feed" but not "fed."

To find	Use this	For example
At least n occurrences of a character or expression	{ n,}	fe{1,}d finds "fed" and "feed."
A range of occurrences of a character or expression	{ n, n}	10{1,3} finds "10," "100," and "1000."
One or more occurrences of a character or expression	@	lo@t finds "lot" and "loot."
The beginning of a word	<	<(inter) finds "interesting" and "intercept" but not "splintered."
The end of a word	>	(in)> finds "in" and "within," but not "interesting."

1. Available at https://support.office.com/en-us/article/Find-and-replace-text-or-formatting-in-Word-for-Mac-ac12f262-e3cd-439a-88a0-f5a59875dcea

Using Macros

Before you segment the data, you will often need to clean up the data by removing extraneous carriage returns which could be mistaken for segmentation breaks. Particularly if you have copied electronic interactions to a file, you may find unwanted carriage returns at the end of each line. In texts, you may find unwanted blanks lines (i.e., 2 carriage returns) between each paragraph.

Removing unwanted carriage returns by hand can be tedious in the extreme. In Word, however, you can create a macro to do the task, assign it to a keyboard shortcut and then apply it repeatedly with far less tedium. Basically, you are linking three separate commands together into a sequence which can then be repeated as a single command. The sequence that will remove a carriage return immediately preceding a line break is as follows:

Backspace delete + insert space + move to beginning of next line

You can assign a sequence of segmentation moves to a single macro in Microsoft Word (see Procedure 2.6).

Setting Up The Data

During the analytic process described in this book, you will store and manipulate verbal data as a set of documents. In Excel, each document will be placed in a sheet in a data workbook (see Excel Procedures 2.3 and 2.4). In MAXQDA, each document will be imported into a Document System (see MAXQDA Procedure 2.3).

Note: You should not import any data into EXCEL or MAXQDA until it has been segmented as discussed in Chapter 3.

Procedure 2.6: Using Macros to Assist Multi-step Segmentation in Word

https://bit.ly/2kL7ATv

1. Open your transcript in Microsoft Word.
2. From the Tools menu, select **Macro > Record New Macro**.
3. Give the macro a name (spaces not allowed).
4. Click on the keyboard icon to assign the macro to a keystroke.
5. Type the keystroke combination you wish to use.

   For example: Cntrl+Option+J.
6. Click **OK** to start recording.

The only indication you will have that a macro is recording is the change in the shape of your cursor.

7. Perform the action you wish to assign to the macro.
8. Return to the **Tool** menu, select **Macro > Stop Recording**.
9. Return to the transcript and press the keystroke assigned to your macro to use throughout.

MAXQDA Procedure 2.3: Setting Up a Document System

https://bit.ly/2kL7ATv

The Document System window is usually found in the upper left-hand side of the MAXQDA interface.

1. To import data into the document system, choose **Documents > Import Document(s)**
2. Navigate to the data file you want to import and click **Open**.

The file will be imported as a document and appear in your document system window.

Excel Procedure 2.3: Setting up a Data Workbook

https://bit.ly/2kL7ATv

When you open Excel, it will provide you with a workbook with a set of empty worksheets as shown in Figure 2.13.

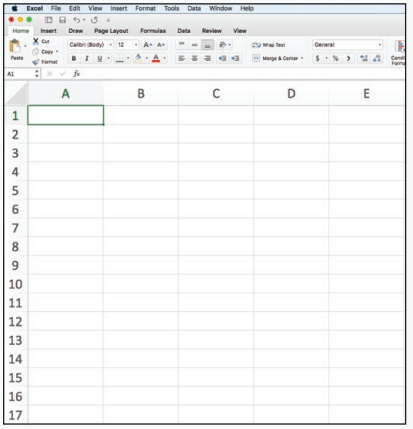


Figure 2.13: An empty Excel workbook.

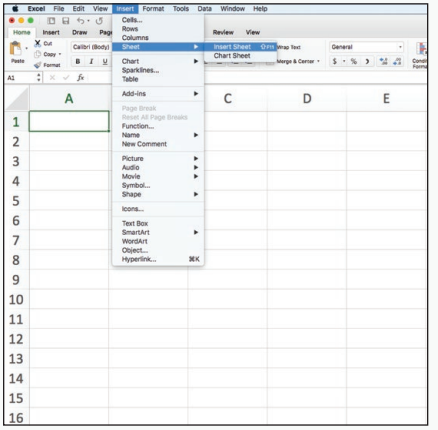


Figure 2.14: Inserting a new worksheet.

1. To add additional worksheets, choose **Worksheet > Insert** as shown in Figure 2.14.
2. For each worksheet choose a one-word name that will help you remember its contents (e.g., Design1).

Choosing a one-word name for each sheet will make it possible to use named data ranges that will simplify later analytic processes. Get in the habit of naming a worksheet as soon as you create it in order to save yourself much confusion later on.

3. To name a worksheet, select **Name > Define Name** from the **Insert** menu, as shown in Figure 2.15.
4. Type in the name you have chosen.

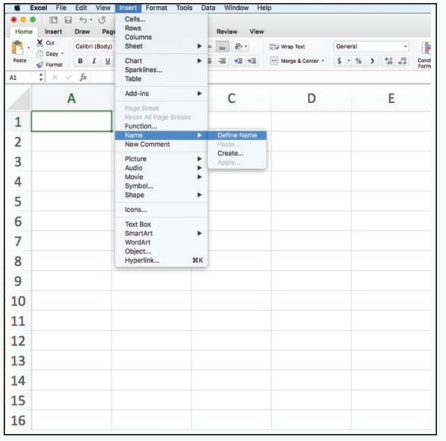


Figure 2.15: Naming a worksheet.

Setting Up a Data Table

A Data Table catalogs all of the data available to you in your data set. This established the universe of data available for possible analysis. As the sample table in Figure 2.16 suggests, it includes the following kinds of information:

- The type of data, listed in columns across the top
- The dates of collection/recording/publishing, listed in rows down the side
- The weeks/months of the projects, numbered.
- The data label: e.g., Technical Meeting #1
- The accession number: This might be the number of the tape on which it was recorded, the document number under which it was filed, the name of the computer file in which it is stored.

Excel Procedure 2.4: Creating a Table of Contents for the Workbook

https://bit.ly/2kL7ATv

One of the best ways to manage your workbook is to create a special worksheet that serves as a table of contents for the rest of the data workbook.

1. Open a blank Excel worksheet.
2. Rename the worksheet TOC.
3. Write a header for the first three columns:
   - Worksheet
   - Date Created
   - Description
4. Place your cursor in the cell with the first worksheet name.
5. From the Insert menu, select **Hyperlink > Webpage** or **File > Select**.
6. Navigate to the corresponding worksheet and click **OK** to link.
7. Repeat to link all of your data sheets.

As long as you do not change the folders where the workbook and coding scheme are stored, this file path will continue to work.

A	B	C	D	E	F	G
Week	Date	Technical Meeting	General Meeting	Don	Joe	Sue
0	8/28/18					#1
1	9/3/18			#1		
	9/3/18		#1			
	9/3/18					#2
2	9/10/18				#1	
	9/10/18		#2			
	9/10/18					#3

Figure 2.16: Sample data table.

Selected Studies Using Sampling

Elliot, N., Kilduff, M., & Lynch, R. (1994). The assessment of technical writing: A case study. *Journal of Technical Writing and Communication, 24*(1), 19-36. https://doi.org/10.2190/53LM-VWV5-JFTV-B7H7 (best case sample)

Mackiewicz, J. (2010). Assertions of expertise in online product reviews. *Journal of Business and Technical Communication, 24*(1), 3-28. https://doi.org/10.1177/1050651909346929 (typical case sample)

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*(1), 57-86. https://doi.org/10.1177/0741088309351547 (comprehensive sample)

Rude, C. D. (2009). Mapping the research questions in technical communication. *Journal of Business and Technical Communication, 23*(2), 174-215. https://doi.org/10.1177/1050651908329562 (criterion/stratified sample)

Schryer, C. F., Afros, E., Mian, M., Spafford, M., & Lingard, L. (2009). The trial of the expert witness: Negotiating credibility in child abuse correspondence. *Written Communication, 26*(3), 215-246. https://doi.org/10.1177/0741088308330767 (criterion sample)

For Further Reading

Alun Jones, R. (1994). The ethics of research in cyberspace. *Internet Research, 4*(3), 30-35. https://doi.org/10.1108/10662249410798894

Bauer, M., & Aarts, B. (2000). Corpus construction: A principle for qualitative data collection. In M. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound* (19-37). London: Sage.

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge*. London: Penguin.

Booth, W., Colomb, G., & Williams, J. (1995). *The craft of research (Chicago guides to writing, editing, and publishing)*. Chicago: Chicago University Press.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications

Department of Health and Human Services. (2010). The Belmont Report. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.HTML

Goetz, J. P., & LeCompte, M. D. (1984) *Ethnography and qualitative design in educational design* (pp. 63-92). Orlando, FL: Academic Press.

Hamel, J., with Dufour, S., & Fortin, D. (1993). *Case study methods* (pp. 40-44). Newbury Park, CA: Sage.

Markham, A., & Buchanan, E. (2012). Ethical decision-making and Internet research: Version 2.0—Recommendations from the AoIR Ethics Working Committee. Retrieved from https://aoir.org/reports/ethics2.PDF

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (pp. 16-49). London: Sage.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*(1), 119.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analyzing talk, text, and interaction* (pp. 54-56, 83-218). London: Sage.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage Publications Ltd.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.