Visualizing the Future of Interaction Studies: Data Visualization Applications as a Research, Pedagogical, and Presentational Tool for Interaction Scholars

Corinne Weisgerber and Shannan H. Butler
St. Edward's University

Abstract: The advent of the social web, or web 2.0, has brought with it an array of data visualization tools developed to help amateur and professional researchers alike uncover new patterns in large data sets. This paper examines one such tool and illustrates how visualization types such as tag clouds and word trees can be incorporated into language and social interaction scholarship and pedagogy. The paper contends that data visualizations can enhance all phases of qualitative coding by facilitating hypothesis formation and moving analysis from a mere description of the data into the realm of conceptualization and theory development. Visualization technologies are also argued to play a role in the validation of research findings by supporting qualitative validity measures such as member checks and reflexive journaling. Finally, issues of data privacy are addressed and pedagogical and presentational applications are explored.

The last few years have seen not only an explosion in the amount of data available on the Internet, but also a transformation of the web itself into a more participatory space. The advent of the social web, or web 2.0, has brought with it an array of content creation tools designed to encourage the creation and sharing of user-generated content. Among those tools are data visualization applications developed to help amateur and professional researchers alike uncover new patterns in large data sets. While these tools are often used to display the dynamics of social networks, their adoption in communication research and pedagogy has been limited. Although communication professionals now have access to an unprecedented amount of interactional data as well as free web applications capable of analyzing that data and displaying it visually (i.e. Many Eyes, TagCrowd, Neoformix, Swivel, etc. [1]), so far, few research studies in our field have taken advantage of these technological advances. This is unfortunate considering that visualization applications promise to provide powerful new ways for discovering relationships and patterns in data sets and formulating visual arguments.

Data visualization as a tool for language and social interaction research

In this paper, we seek to introduce visualization tools and illustrate how they could be applied to language and social interaction scholarship. Specifically, we will demonstrate how such tools could inform the coding procedures used in grounded theory research. We have chosen IBM's Many Eyes website as the focus of this paper because of its ease of use and ability to generate a large variety of data visualizations. The concepts and examples discussed here can however be applied to most other data visualization tools available to date. Many Eyes, which launched in January 2007, "is a web site that provides collaborative visualization services, allowing users to upload data sets, visualize them, and comment on each other's visualizations" (Danis, Viégas, Wattenberg, & Kriss, 2008, p. 1). The site currently allows lay and expert users to choose between 17 different types of visualizations, ranging anywhere from tracking changes over time, to examining relationships among data points. Probably of greatest interest to language and social interaction scholars are the visualization types that handle textual data such as tag clouds, and word trees.

Tag clouds were made popular by bloggers wishing to give readers a quick sense of their blog's content. In the blogosphere, a tag cloud therefore refers to a list of common tags or keywords used to describe the content of a blog. To make it easy to grasp that content, the list displays the most frequently used tags in a bigger font size. Their application is not limited to the blogosphere. Any textual data can be pulled into a tag cloud generator such as Many Eyes or TagCrowd and transformed into a display of word frequencies in which the "size of a word corresponds to the quantity associated with that word" (Tag Cloud Guide, n.d.). Figure 1 is an example of a sample tag cloud generated from a transcript of the third presidential debate between Barack Obama and John McCain.

Figure 1.
Tag cloud of the 200 most frequently used words during the third presidential debate.

Note. To see a full-size image with improved legibility, click on this image.

Many Eyes tag clouds contain a tooltip that provides contextual information such as the context a word was used in and the number of occurrences of a particular word in the text. This tooltip is activated when a user rolls the mouse over a certain word. In the example above, the tooltip shows some of the contexts in which the word "campaign" occurred and indicates that this word was used 32 times during the debate.

For a more detailed analysis of the contextual information, users may want to create another type of visualization, called a word tree. A word tree "is a visual search tool for unstructured texts" (such as speech), which lets users pick a word or phrase from the text and then locates all the different contexts in which that word or phrase appears (Word Tree Guide, n.d.). These contexts are then "arranged in a tree-like branching structure to reveal recurrent themes and phrases" (Word Tree Guide, n.d.). Once the visual has been generated, users can switch between viewing all the phrases that appear either before or after the search term. Many Eyes calls the word tree tool an experimental visualization technique that uses the book concordance as a model. Figure 2 was generated from the same transcript of the third presidential debate and shows a sample word tree for the term "plumber."

Figure 2.
Sample word tree displaying the context in which the term "plumber" occurs.
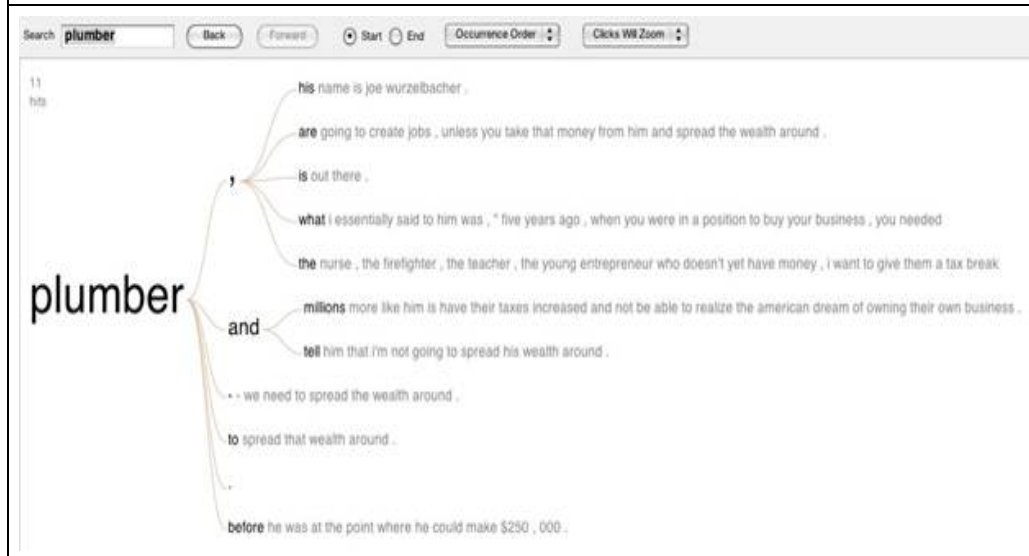
Note. To see a full-size image with improved legibility, click on this image.

To illustrate how these tools could be incorporated into language and social interaction scholarship, a data set from a previous study on the social construction of a sleep disorder (Weisgerber, 2004) was uploaded into Many Eyes. The data consisted of 646 publicly available messages to an online forum devoted to the discussion of a mysterious disorder named sleep paralysis. In order to load the data into Many Eyes, all individual posts had to be merged into one large text document. Since the original data set contained formatting such as subject and reply lines, those lines were removed to avoid generating a tag cloud with the words "subject:" and "re:" as the most frequently occurring words. Many Eyes automatically ignores common English words such as articles when generating the visual, so no further action was needed to address that issue.

One of the questions we sought to answer in the original study was how people who suffer from sleep paralysis make sense of their experience. This question was all the more intriguing considering that sleep paralysis is a poorly understood disorder (Buzzi & Cirignotta, 2000) and that few people in the Western world know what to attribute their symptoms to when they first experience them (Hufford, 1982). The matter is complicated by the fact that the symptoms are rather bizarre and include waking up to a fully paralyzed body and experiencing things such as visual and auditory hallucinations, chest pressure, and/or a sensed evil presence (Cheyne, Rueffer, & Newby-Clark, 1999). In order to answer the question of how the disorder is constructed online, the data was originally coded using the grounded theory approach of open, axial, and selective coding (Carpenter, 1999; Strauss & Corbin, 1990). Results from that analysis suggested that sleep paralysis sufferers invoke three different explanatory frameworks in their sense-making process. While most attributed religious or spiritual meaning to the experience, some perceived it as a paranormal phenomenon, and only few embraced a medical explanation (Weisgerber, 2004).

Tag clouds as research tools

Although visualization tools such as those discussed in this paper couldn't actually replace the line-by-line coding process that led to these results, they could however be used to support and even enhance the

analysis of social interactions. In the following sections, we will discuss how visualization applications could inform qualitative data analysis in general and the grounded theory method in particular. According to Strauss and Corbin (1990), there are three types of coding: open, axial, and selective. Open coding seeks to label the various phenomena discovered in the data and results in the identification of codes and concepts. Through the constant comparative method, these concepts are then compared to one another and similar concepts are merged into higher level categories. During open coding, the investigator reads a transcript under study line-by-line and attempts to label participants' statements and behaviors. This process is supposed to "open up the inquiry" (Strauss, 1987, p. 29) and to break the data down into smaller units (Rice & Ezzy, 1999).

In order to demonstrate how tag clouds could support both of these tasks, we uploaded the sleep paralysis data set into Many Eyes and displayed the results in the form of a simple tag cloud (Figure 3). During the initial coding phase, qualitative researchers are supposed to "look for what they can define and discover in the data" (Charmaz, 1988 p. 113). A look at that visualization allows for a first glimpse into the question of how the disorder is constructed through online discourse. The words that immediately stand out with regards to this research question are: evil, demons, figure, presence, sense, shadow, shadowy, spirit, spirits, spiritual, god, Jesus, obe (out-of-body experience), and disorder.

Figure 3.
Single word tag cloud of the 200 most frequently occurring words in the data set.

Note. To see a full-size image with improved legibility, click on this image.

By visualizing the words contained in the data set in this manner, a picture starts to emerge about the explanatory frameworks evoked by sleep paralysis sufferers. This initial picture, albeit incomplete, points to the prevalence of a spiritual/religious explanation of the phenomenon where hallucinations that accompany the disorder are interpreted as evil demons, spirits, and shadows, and God and/or Jesus are called upon for help. Figure 3 also identifies other ways of making sense of the experience, but the relative size of the words "obe" and "disorder" combined with the lack of related words, suggests that these explanations may not have enjoyed as much popularity on the online forum. The prevalence of other tags in the data set such as "fear," "scared," "scary," "terrified," "terrifying," and "terror," illustrates the subjective experience of this dis order and may suggest that these phenomena deserve further investigation. At this point, we suggest that the researcher return to the transcript for more in-depth coding of the phenomena brought to light by the tag cloud. The visualization thus doesn't replace traditional data analysis methods; it simply helps identify where to focus the coding.

While this single-word tag cloud may be useful in helping the researcher develop ideas about what is happening in the data, it doesn't provide much contextual information. In order to make the tag cloud more contextually relevant, users can generate a two-word tag cloud, which examines the frequency of two consecutive words in a data set. Users can create such a visualization either for an entire data set (Figure 4), or for specific words and their correlates (Figure 5). Figure 4 displays the 200 most common word combinations from our sleep paralysis data set.

Figure 4.
Two-word tag cloud of the data set.
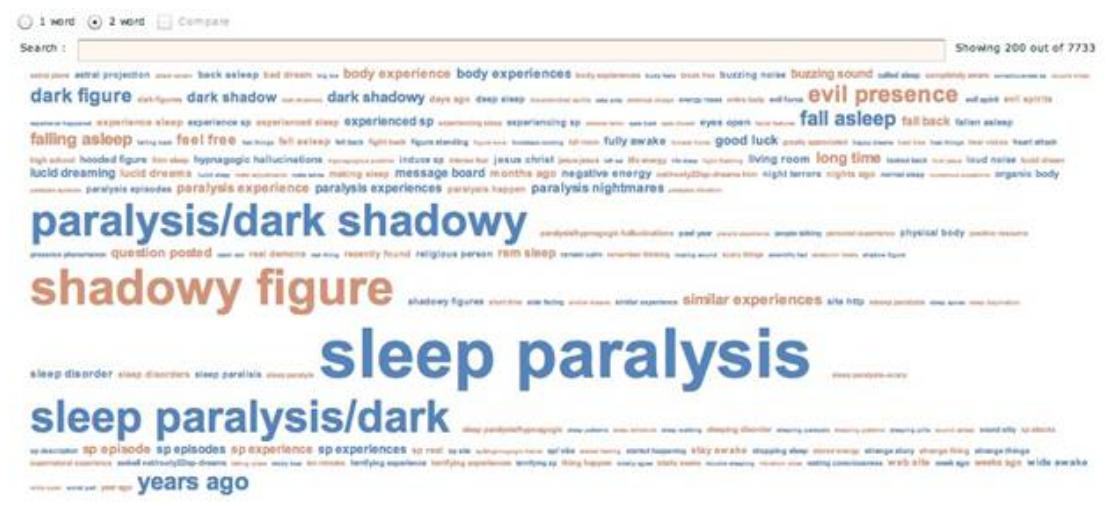
Note. To see a full-size image with improved legibility, click on this image.

These word combinations prominently sum up the symptoms of the disorder and suggest that the subjective disorder narratives cluster around a dark, shadowy figure, which is perceived as an evil presence. To illustrate the purpose of tag clouds in open coding, let's take a closer look at this display of the most common phrases used in the 646 posts that made up the data set. The adjectives "dark," "evil," and "shadowy" immediately stand out in this visual because of their relative size (indicating frequency) and their repeated use (i.e. dark figure(s), dark shadow, dark shadowy). As a result, the researcher may wish to use these adjectives as in vivo codes and label all similar experience descriptions with the code "dark shadowy figure." Similarly, the visualization seems to suggest that sleep paralysis sufferers often refer to their disorder as an "episode" or an "experience." Again, this insigh t could lead to the discovery of another in vivo code, which could then be applied to the data during line-by-line coding. Besides giving researchers an idea for what is going on in the data set and helping them identify in vivo codes, tag clouds also allow researchers to approach the data in a new manner. As Charmaz (1988) contends, "using the grounded theory method necessitates that the researcher look at the data from as many vantage points as possible" (p. 114). Visualization tools have been argued to do just that by helping us discover new ways of understanding the world (Anderson, 2008).

While the conclusions drawn from Figure 4 should be further analyzed with traditional line-by-line coding, they should also be examined through a more complex visualization technique. The search feature built into the two-word tag cloud generator allows users to create more complex tag clouds by identifying the words that cluster around a certain term. Figure 5 provides an example of a two-word tag cloud generated for a particular keyword, in this case the adjective "evil."

Figure 5.
Tag cloud of the words clustering around the tag "evil".
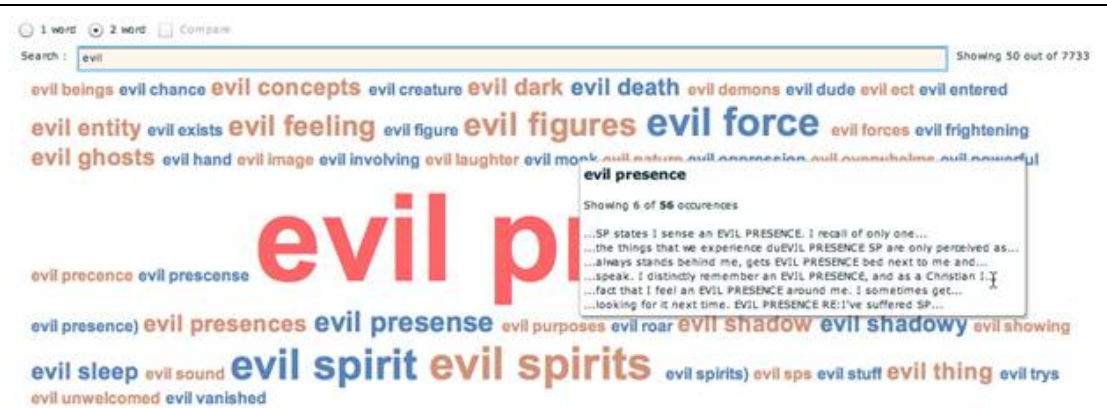
Note. To see a full-size image with improved legibility, click on this image.

The visual displays the 50 words most commonly associated with this adjective. It also allows users to view "information about the occurrences of that word and the context it was used in" (Tag Cloud Guide, n.d.) by rolling their mouse over a particular word pair, in this case the phrase "evil presence." This type of visualization is helpful during the axial coding phase, in which a researcher seeks to detect the dimensions, properties, conditions, consequences, interactions, and strategies that are "associated with the appearance of the phenomenon referenced by the code or category" (Strauss, 1987, p. 36) and tries to make connections between categories and sub-categories. If a "sense of evilness" was identified as a category during the open coding process, axial coding would require the researcher to try to understand that category in its relation to other codes and ca tegories. This is where the visualization type displayed in Figure 5 would be helpful. Since the goal is to understand the concept of sensed evilness in all its different incarnations as expressed in the data set, a two-word tag cloud generated through a keyword search could help guide this inquiry.

In this case, the visualization could help the researcher understand subtle nuances in the data surrounding the idea of evilness. The tag cloud for instance suggests that the word evil tends to be associated with the words presence (in all its different spellings), figure(s), and spirit(s). This raises important questions for axial coding. Why are some people describing the experience as an evil presence while others are speaking of evil figures? What are the conditions of invoking an evil spirit explanation? To start to answer these questions, it is necessary to look at the actual text in which these phrases occur. Figure 5 displays the context for 6 out of 56 occurrences of the phrase "evil presence." Two of them are listed below for illustration purposes:

- "In most (90%) of my SP states I sense an evil presence. I recall of only one time that I have seen a dark shadowy figure."
- "I have never seen it because it always stands behind me, gets on the bed next to me and applies pressure on me."

Although it would be premature to base any definite conclusions on the analysis of the few text occurrences that Many Eyes provides in its visualization, it is possible to start hypothesizing that the evil presence relates to a sensed experience while the evil figure refers to an actual visual perception of a figure or shadow. This hypothesis would of course have to be tested through more thorough line-by-line

coding of all pertinent posts. To identify the conditions of invoking an evil spirit explanation, a similar two-tag word cloud could be generated. The idea is that visualizations can facilitate hypothesis formation (Apitz & Lin, 2007; Ware, 2004) and move the analysis from a mere description of the data into the realm of conceptualization and theory development - a move that has been described as the hardest part of doing qualitative research such as grounded theory analysis (Kendall , 1999). The hypotheses generated through this visualization process should therefore be seen as providing "the basis for further exploration" of the data (Apitz & Lin, 2007, p. 60). This ability to promote theorizing was also observed in a study by Danis and colleagues (2008), which revealed that visualizations caused users "to ask questions [of the data] rather than simply look for answers" (p. 9).

Word trees as research tools

As qualitative researchers embark on this hypothesis creation process and start exploring potential patterns and relationships in the data, the word tree visualization tool becomes particularly interesting. This tool allows users to search a data set for a particular word or phrase and to display the matches in a tree-like branching structure. For instance, knowing that our data suggests that some sleep paralysis sufferers construct their experience as an actual medical disorder, we may want to explore the narratives surrounding the term "medical".

To create a word tree, a user first enters a search term. The computer then locates all the occurrences of that term in the data set, and displays them on the screen along with the phrases that appear either before or after it (depending on which display mode the user selected). Branches can be ordered from top to bottom by order of occurrence in the text (the default setting), by overall branch size, or by alphabetical order (Word Tree Guide, n.d.). Figure 6 displays the word tree visualization for the term "medical."
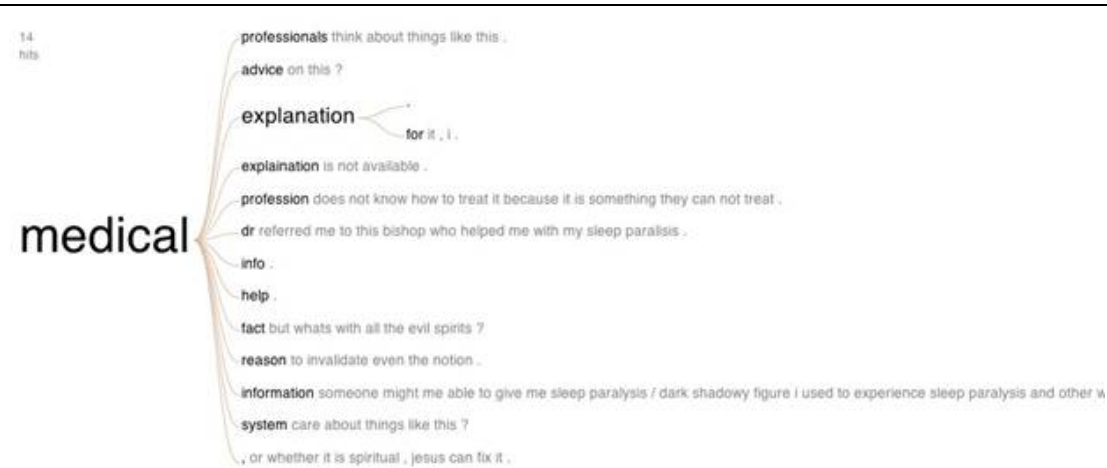
A quick analysis of that word tree reveals that most of the narratives are characterized by a rather negative valence and exhibit guarded optimism for the prospect of receiving help from the medical field. While it is possible to garner the meaning of some of the messages displayed in this word tree at first glance, others don't contain enough context to make sense of them. In that case, the user has the option to alter the tree to view the context occurring before, rather than after, the keyword or keyphrase. Doing so for the first tree branch displayed in Figure 6 would reveal the following message of mistrust: "I learned many years ago not to care what medical professionals think about things like this."

To test the idea that sleep paralysis sufferers view the medical profession with suspicion, further word trees may be created using related words or in vivo codes. Figure 7 presents such a word tree. Most of the branches of that tree seem to further reinforce that idea, or at least provide an explanation for the guarded review doctors have received in this online forum. In it, sleep paralysis sufferers report having been brushed off by doctors, diagnosed with mental health problems, or treated with anxiety drugs without, in most cases, receiving an accurate explanation of their symptoms.

Figure 7.
Word tree displaying the context in which the term "doctor" appears.

Note. To see a full-size image with improved legibility, click on this image.

This again poses some important questions capable of driving the analysis forward. Is it possible that these past negative experiences with health professionals have tainted sleep paralysis sufferers' faith in the medical system and influenced their inclination to accept non-medical explanatory frameworks? In other words, is a past negative experience a condition of an alternative explanatory model? Posing and subsequently answering those types of questions is essential to successful axial coding. Although further line-by-line coding is needed to provide a definite answer, the word trees displayed in Figures 6 and 7 proved invaluable in calling immediate attention to these questions and moving theory development along. Since these word trees can aid researchers in understanding how different coding categories are integrated to form a theoretical framework, they play an equally important role during the selective (Strauss & Corbin, 1990) or theoretical coding (Glaser, 1992) process. Both selective and theoretical coding are concerned with the integration of categories discovered through open and axial coding. During the selective coding process, a researcher focuses on those codes that relate directly to the core variable in an effort to unify all codes in respect to the core variable and, thereby, to densify the emerging theory. According to Strauss (1987), a core variable needs to possess the following characteristics: (a) it needs to be central, which means that it needs to be related to as many other categories as possible; (b) it needs to appear frequently in the data; (c) it needs to relate easily to other categories; (d) it needs to have clear implications for a more general theory; (e) the theory needs to evolve as it becomes more detailed; and (f) it needs to account for a maximum amount of variation in the analysis. As the previous discussion has illustrated, data visualizations can help us identify at least the first three of these conditions more easily. Since tag clouds are based on frequency counts and word trees are designed to reveal recurrent themes and phrases in a data set, both types of visualization should prove useful in the quest to detect core variables.

Using visualizations to establish the validity of the findings

So far, we have discussed how information visualization technologies could be used to complement data analysis and drive theory development. What we haven't addressed is the role these technologies could

play in the validation of research findings. To ensure the trustworthiness of the claims derived from qualitative data analysis, Lincoln and Guba (1985) suggest the use of reflexive journals and member checks among other procedures. Reflexive journals "display the investigator's mind processes, philosophical position, and bases of decision about the inquiry" (Lincoln & Guba, 1985, p. 109). These journals or memos contain an internal dialogue concerning the data collected and are designed to help raise the data to the level of conceptualization. Again, it is possible to see how data visualization technologies such as the ones described above could be used to explore observatio ns and assumptions expressed in reflexive journals and ultimately lead to higher-level abstractions. Research further supports the idea that visualization applications can provide powerful tools for reflection (Viégas, Boyd, Nguyen, Potter, & Donath, 2004). A study of email visualization tools conducted by Viégas and colleagues (2004) indeed concluded that "users readily utilized the visualizations to revisit past experiences and reflect on their relationships with others" (p. 7). Since reflexive journals are introspective by nature, the incorporation of visualization tools capable of promoting reflection should only enhance the journaling process.

Besides journaling, Lincoln and Guba (1985) also recommend using member checks as a way to guard against bias in a study. Member checks entail sharing analyses, interpretations, and conclusions with participants of the study in an effort to solicit their feedback on these matters. Doing so allows research participants to play an active role in the study and to integrate their insights into the final research report. By including data visualizations in their presentation of interpretations and conclusions, researchers would not only make academic language more accessible to study participants, but also increase the amount of lay input. Research indeed suggests that visualizations can be extremely successful in encouraging conversation and collaboration around a data set. "The compelling presentation of data through visualization's advanced techniques generates a surprising volume of impassioned conversations. Viewers ask qu estions, make comments, and suggest theories for why there's a downward trend here or a data cluster there" (Wattenberg & Viégas, 2008, p. 30). Researchers at IBM's Visual Communication Lab, which has developed Many Eyes, therefore "believe that visualizations become even more powerful when multiple people access them for collaborative sensemaking" (Visual Communication Lab, n.d.).

By inviting study participants to join in this visual sensemaking process, language and social interaction researchers could not only perform vital member checks, but also give people outside academe a chance for "talking and thinking about data in a new way" (Viégas and Wattenberg, 2008, p. 49). As Viégas and colleagues' (2004) study on email visualizations suggests, study participants actually welcome the idea of examining visualizations that capture their own activities and behaviors. Their study revealed that "users felt compelled to tell stories around the data they saw in the visualizations" (p. 2) and that they spent considerable amounts of time exploring them. Such an involvement on the part of study participants could lead to the collection of additional data, which in turn could help saturate th e data. Since sites such as Many Eyes were designed as collaborative spaces "where users can share their own dynamic, interactive representations of big data" (Horowitz, 2008, p. 121), researchers could further empower study participants by allowing them to interact with the data themselves and produce their own "vernacular visualizations" (Viégas and Wattenberg, 2008, p. 49).

While data visualization tools can enhance the trustworthiness of a study through member checks and reflexive journaling, their ability to pinpoint inaccuracies or problems in the data set provides another important quality control mechanism (Jana, 2007; Ware, 2004). "With an appropriate visualization, errors and artifacts in the data often jump out" at the researcher and can be fixed before they affect study findings. To illustrate this ability to spot problems, let's take a look at Figure 4, which represents the phrases "sleep paralysis/dark" and "paralysis/dark shadowy" as particularly prominent. Considering that the data set is comprised of naturally occurring talk, it seems odd that so many forum users would have used a slash in this particular phrase. A closer analysis reveals that the phrase "sleep paralysis/dark shadowy figure" was fi rst used as a subject line and then simply repeated in the response line of a

number of subsequent posts. To mitigate potential data analysis problems caused by this line, we simply deleted all response lines from the data set and focused our examination on the body of the message. That way we avoided attributing undue importance to a phrase contained in the subject line instead of in the actual post, which may have contained a message unrelated to the response line. Without the tag cloud however, it would have been difficult to detect this problem this easily.

Privacy issues

Despite the many promises visualization applications such as Many Eyes hold for researchers, there are some important limitations that need to be discussed. Of primary concern to language and social interaction researchers dealing with human subjects should be the issue of data confidentiality and privacy. Because Many Eyes and other similar web applications were developed as social sites, "all artifacts on the site--data, visualizations, user comments—are visible to the public" (Danis et al., 2008, p. 2). The concept driving the site is the idea that many eyes are better at spotting new connections in a data set than a single pair of eyes (Wattenberg & Viégas, 2008). Although this concept of collaborative data analysis may be the ultimate strength of the site, it also poses a series of problems to researchers required to protect the privacy of the data they collect. In order to receive IRB approval for a study, many language and social interaction researchers have to provide safeguards to protect the confidentiality or anonymity of their sources. This means that uploading transcripts of interactions to a site that will make the data set publicly available is either problematic or simply out of the question.

While we realize that the current state of affairs may not allow all researchers to take advantage of these visualization tools quite yet, we do believe that these sites will eventually offer "walled-in communities for proprietary data" (Danis et al., 2008, p. 8) and options to switch to a private mode that would safeguard data confidentiality. The New York Times Visualization Lab (Visualization Lab, 2008), which relies on the Many Eyes technology, suggests that in the future private visualizations only visible to the user may be added depending on user interest. Swivel, another social data visualization website, already offers a service that allows its users to keep their data private in exchange for a monthly fee. However, as of spring 2009 Swivel supports only spreadsheet based data and does not offer a visualization tool for unstructured text data.

For now, experimentation with these tools and incorporation of visualizations into research procedures may be limited to researchers who study publicly available interactional data, such as online discourse. The data set we used as an example in this study for instance was openly available on the Internet and did not require the audio or video taping of private conversations, which would require greater privacy safeguards. According to Viégas and Smith (2004), discourse that occurs in "computer-mediated conversational spaces, which are intrinsically recordable" are fair game for researchers to analyze (p. 9). The data set for the original sleep paralysis study also contained transcripts of in-depth interviews with forum users. Because informed consent was required for those interviews and interview participants were assured that their data would be kept confidential, the transcripts f rom those interviews were not loaded into Many Eyes for purposes of this paper.

In order to address the concerns of users interested in protecting the privacy of their data, the developers of Many Eyes suggest anonymizing the data and concealing the metadata. The former can be achieved by replacing the names of participants with pseudonyms, while the latter would require investigators to "give obscure titles to their data sets and visualizations" (Danis et al., 2008, p. 8) in an effort to remove some contextual information. This latter procedure has been found to effectively render the data contained in visualizations useless to people outside the study (Viégas et al., 2004). Indeed, participants in Viégas and colleagues' study felt that the visualizations of their email usage patterns would be useless in the hands of strangers who would be at a loss to "understand the stories behind the images" without their

explanations ().

Data visualization as a pedagogical tool

Visualization tools also offer many possibilities for pedagogical application. In the past, students needed access to expensive qualitative or quantitative data analysis software packages in order to be able to complete even small research projects. Departments without a significant budget for research software, lab facilities, and training staff could offer students only a theoretical knowledge of these tools or at best give them a chance to apply them to limited assignments. Besides being expensive, tools such as NUD*IST, NVivo, HyperResearch, and SPSS also have a steep learning curve for first-time users that may make it difficult to incorporate them into class projects where software instruction time is limited. However, with freely available data visualization tools such as Many-Eyes, students can now tackle research projects outside the classroom with little or no software training. Although these projects may not be as detailed or intensive as full-fledged resear ch projects, they still allow students to explore a data set and share their insights. As our previous discussion has shown, visualizations created with such tools may be more successful in encouraging academic discussion than their traditional counterparts. Since the data would be socially available for other students and academics to comment on, there is a stronger chance of encouraging conversations about the data. Indeed, research by Viégas and Wattenberg (2006) suggests that "visualization apps can enable interactions between people in powerful and unexpected ways" (p. 801).

This type of pedagogical experience would likely be conducted over the course of several days or even weeks in order to allow for analysis, presentation, feedback, and revision. Since the data set would be housed online, students could manipulate the same data in a variety of ways and share different interpretations. Considering that "visualization is often used collaboratively" (Danis et al., 2008, p.3), the social nature of the data set and the resulting visualizations could facilitate collaboration between students, classes, and even academic institutions. From a pedagogical perspective, the possibility of increased levels of interaction between students and scholars seems particularly promising and may herald a new era of collaborative research and pedagogical opportunities.

Besides their application in classroom research projects, visualization technologies also provide a useful teaching tool. As educators, we are well aware of the fact that today's students expect to receive quality instruction coupled with the latest in visual presentation and media. Theirs is a generation that has been raised in a world that preferences the visual. To accommodate, most professors, it seems, now supplement their lectures with slideware presentations and multimedia. This transition within the academy has not met with resounding approval from some academics who fear the visual hegemony that it may foster. In an article on hegemonic visualism, Tietje and Cresap (2005) reference Marshall McLuhan, in asserting that although slideware looks like a cool medium, it is actually a hot one that fills in all of the blanks and impedes students from thinking on their own. Their fears rest mostly on the assumption that visual pr esentations move education closer to entertainment and further away from collaboration. We would however suggest that the problem is rooted not so much in the limitations of the particular media, but rather in the way that media are utilized by the course instructor. One way to get students involved is, of course, to make the media more interactive. Data visualization tools are particularly useful in this respect since they allow analysis of data in real-time with input from both the instructor and students. This real-time nature of data visualization also helps override any sense within the classroom that the data presented has been cherry-picked to support the instructor's particular viewpoint, thereby increasing its potential to foster authentic interaction and discovery.

While research has shown that college students learn better when images or graphs are added to textual slides, an overabundance of audio and visual information may actually hinder learning (Mayer, 2001).

From a pedagogical standpoint, it is important then to find the right balance between enough but not too much information. This optimal "signal to noise ratio" (Reynolds, 2008, p.122) is reached by presenting the minimum effective visual information without additional artifice or ornamentation. Reducing visual noise means presenting only relevant data and avoiding undo verbiage and distracting graphic elements. The visualization tools we have discussed in this article all produce high signal to noise ratio graphics since they create clean, easily legible visualizations (Jana, 2007) that allow the instructor to focus on one particular aspect of the data without combing through a huge textual display or providing copious notes. By incorporating such visualizations into presentations of course material, educators could make the data far more relevant and digestible for students.

The possibilities for classroom application are virtually limitless. A professor teaching interpersonal communication might upload a data set from a traditionally recorded and transcribed conversation for in-class analysis of gender biased language. An instructor of computer-mediated communication could upload a set of email exchanges to analyze social interaction and networks. A teacher of rhetoric might choose to upload a speech from the night before for in-class study as a cluster criticism or fantasy theme analysis. By applying different visualizations, students can generate different responses to the data and become actively engaged in the search for patterns of meaning. We believe that data visualization tools offer a host of new pedagogical opportunities for any course that studies language and interaction.

Data visualization as a research presentation tool

Many of the research and pedagogical advantages offered by visualization tools also apply to the presentation of research findings to peers. Conference presentations could benefit equally from the inclusion of data visualizations presented either in real time or as screenshots created ahead of time. Such visualizations could help illustrate and clarify research findings while simultaneously increasing audience involvement by inviting them to enter into a conversation with the presenter. Not only could the use of visualizations help engage the audience, but it could also aid in the long-term recall of key information (Reynolds, 2008) and help condense complex arguments into a format fit for a standard 15-20 minute conference presentation.

It may be time for us to take a cue from the natural sciences, which often consider oral presentations of research at conferences as "one of the most serious and most highly valued kinds of speech" (Rowley-Jolivet, 2004, p. 146) especially considering the role they play in the creation of scientific knowledge. In the sciences, where the addition of data visualizations is not seen as artifice, but rather as a necessity for the effective presentation of data and the successful formulation of scientific arguments, the visual semiotic has long enjoyed greater rhetorical force than the spoken word (Rowley-Jolivet, 2004). This is due to the fact that in comparison to the visual semiotic, language is relatively inept at expressing certain relationships between scientific data (Lemke, 1998). As a result, for many data sets visualization is the only effective means to convey th e information to an audience. Additionally, visualizations not only allow for the more effective and elegant exchange of information in a conference setting, but they also act as indispensable time saving devices, enabling researchers to sum up relationships on one well-crafted graphic instead of spending considerable time describing them verbally. This preference for visual arguments seems so deeply rooted in scientific conference presentations that audiences tend to trust the visual data presented more than they trust the verbal information in cases where the verbal message contradicts the visual material presented (Rowley-Jolivet, 2004).

Some scholars within the humanities and social sciences may be hesitant to accept visual presentations of textual data considering that there has long been a hierarchy within the academy privileging the spoken word over writing, and writing over images. Even in the face of postmodern theory, which embraces the visual, the academy tends to remain logocentric. However, just like numbers in scientific data, words are

visual symbols of something other than themselves. If graphing mathematical calculations can provide insight into our material cosmos, then the graphing of words should offer insight into our inner and social worlds. As data visualization expert Martin Wattenberg observed, "language is one of the best data compression mechanisms we have. The information contained in literature, or even email, encodes our identity as human beings" (Horowitz, 2008, p.121). With the aid of data visualization tools we can make our research more invigorating for our colleagues and more understandable and relevant to those outside of our field of study.

Discussion

As our discussion of the research, pedagogical, and presentational applications of data visualization techniques indicates, these tools hold many promises for enhancing not only the way we conduct and present research, but also how we teach our students. Visualization tools have been found to help us analyze large data sets (Anderson, 2008; Ware, 2004), discover new ways of understanding the world (Anderson, 2008), reveal patterns and draw new connections between data (Bulik & Klaassen, 2006; Jana, 2007; Viégas & Wattenberg, 2006), facilitate hypothesis formation (Danis et al., 2008; Ware 2004), detect inaccuracies in the data (Jana, 2007; Ware, 2004), invite collaboration and academ ic debate (Danis et al., 2008; Viégas, Wattenberg & Dave, 2004; Wattenberg & Viégas, 2008), and encourage interaction around a data set (Danis et al., 2008; Viégas & Wattenberg, 2006; Wattenberg & Viégas, 2008). Despite these obvious advantages, it is necessary to note that the goal of any data visualization application "is not to supplant but to augment the scholar in the creation of knowledge" (Apitz & Lin, 2007, p. 60). These tools merely provide an additional, albeit important, resource to a researcher concerned with making sense of a set of data. They are not meant to replace our traditional research methodologies, but to simply help us perform them more effectively. A s these technologies mature, they will provide researchers with ever more powerful means to explore data and allow for the analysis of ever increasing data sets.

At least one researcher (Anderson 2008) has argued that the massive amounts of data available online and the increased computational power now available to analyze and visualize them spell the death of the scientific method built around hypothesis testing and the end of "every theory of human behavior, from linguistics to sociology" (p. 108). However, the tools and the examples described in this paper show how visualization techniques can be used to formulate hypotheses and develop theory. Considering that the Internet has made available huge amounts of interactional data and that the humanities and social sciences tend to work with large data sets (Apitz & Lin, 2007), it only seems appropriate that language and social interaction scholarship start taking advantage of tools designed to support the analysis of large collections of textual data.

References

Anderson, C. (2008, July). The end of theory. **Wired, 16**(7), 108-109.

Apitz, G., & Lin, J. (2007). Interfaces to support the scholarly exploration of text collections. **Proceedings of the CHI 2007 Workshop for Exploratory Search and HCI**, 60-63.

Bulik, B., & Klaassen, A. (2006, July 24). Tag Clouds. **Advertising Age, 77**(30), 19.

Buzzi, G., & Cirignotta, F. (2000). Isolated sleep paralysis: A web survey. **Sleep Research Online 3**(2), 61-66.

Carpenter, D. R. (1999). Grounded theory as method. In H. J. Streubert & D. R. Carpenter (Eds.),

Qualitative research in nursing: Advancing the humanistic perspective (pp. 99-116). Philadelphia: Lippincott.

Charmaz, K. (1988). The grounded theory method: An explication and interpretation. In R. M. Emerson (Ed.), **Contemporary field research: A collection of readings** (pp. 109-126). Long Grove, IL: Waveland Press.

Cheyne, J. A., Rueffer, S. D., & Newby-Clark, I. R. (1999). Hypnagogic and hypnopomic hallucinations during sleep paralysis: Neurological and cultural construction of the night-mare. **Consciousness and Cognition, 8**, 319-337.

Danis, C. M., Viégas, F. B., Wattenberg, M., & Kriss, J. (2008). Your place or mine? Visualization as a community component. **Proceedings of the 26th Computer Human Interaction Conference,** 1-10. Retrieved August 7, 2008, from http://www.research.ibm.com/visual/papers/chi08.pdf

Glaser, B. (1992). **Emergence vs. forcing: Basics of grounded theory analysis**. Mill Valley, CA: Sociology Press.

Horowitz, M. (2008, July). Visualizing big data. **Wired, 16**(7), 121.

Hufford, D. J. (1982). **The terror that comes in the night: An experienced-centered study of supernatural assault traditions**. Philadelphia: University of Pennsylvania Press.

Jana, R. (2007, June 12). IBM's data visualization champion. **Business Week Online.** Retrieved August 8, 2008, from Academic Search Complete database.

Kendall, J. (1999). Axial coding and the grounded theory controversy. **Western Journal of Nursing Research, 21**(6), 743-757.

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), **Reading science: Critical and functional perspectives on discourses of science** (pp. 87-113). London: Routledge.

Lincoln, Y. S., & Guba, E. G. (1985). **Naturalistic inquiry**. Beverly Hills, CA: Sage.

Mayer, R. E. (2001). **Multimedia learning**. New York: Cambridge University Press.

Reynolds, G. (2008). **Presentation Zen: Simple ideas on presentation design and delivery.** Berkley, CA: New Riders.

Rice, P. L., & Ezzy, D. (1999). **Qualitative research methods: A health focus.** Melbourne, Australia: Oxford University Press.

Rowley-Jolivet, E. (2004). Different visions, different visuals: A social semiotic analysis of field-specific visual composition in scientific conference presentations. **Visual Communication, 3**(2), 145-175.

Strauss, A. (1987). **Qualitative analysis for social scientists.** Cambridge, England: Cambridge University Press.

Strauss, A., & Corbin, J. (1990). **Basics of qualitative research: Grounded theory procedures and techniques.** Newbury Park, CA: Sage.

Tag Cloud Guide. (n.d.). Retrieved August 7, 2008, from http://services.alphaworks.ibm.com/manyeyes /page/Tag_Cloud.html

Tietje, L. & Cresap, S. (2005). Hegemonic visualism. **Radical Pedagogy, 7**(1), 1-9.

Viégas, F. B., Wattenberg, M., & Dave, K. (April, 2004). Studying cooperation and conflict between authors with history flow visualizations. **Proceedings of the 2004 Computer Human Interaction Conference**, 1-10**.**
Retrieved August 7, 2008, from http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf

Viégas, F. B., & Smith, M. (2004). Newsgroup Crowds and Authorlines: Visualizing the Activity of Individuals in Conversational Cyberspaces. **Proceedings of the 37th Hawaii International Conference on System Sciences**, 1-10. Retrieved August 6, 2008, from http://alumni.media.mit.edu/~fviegas/papers /authorlines.pdf

Viégas, F. B., Boyd, D., Nguyen, D. H., Potter, J., & Donath, J. (2004). Digital artifacts for remembering and storytelling: PostHistory and Social Network Fragments. **Proceedings of the 37th Hawaii International Conference on System Sciences**, 1-10. Retrieved August 6, 2008, from http://alumni.media.mit.edu/~fviegas/papers/posthistory_snf.pdf

Viégas, F. B., & Wattenberg, M. (2006). Communication-Minded Visualization: A Call to Action. **IBM Systems Journal, 45**(4), 801-812.

Viégas, F. B., & Wattenberg, M. (2008). Tag clouds and the case for vernacular visualization. **Interactions, 15**(4), 49-52.

Visual Communication Lab. (n.d.). Retrieved August 7, 2008, from http://www.research.ibm.com/visual /index.html

Visualization Lab. (2008). Retrieved October 30, 2008, from http://vizlab.nytimes.com/page/FAQ.html

Ware, C. (2004). **Information visualization: Perception for design (2nd ed.)**. San Francisco: Morgan Kaufman.

Wattenberg, M. & Viégas, F. B. (2008). Emerging graphic tool gets people talking. **Harvard Business Review, 86,** 30-32.

Weisgerber, C. (2004). Turning to the Internet for help on sensitive medical problems: A qualitative study of the construction of a sleep disorder through online interaction. **Information, Communication & Society, 7**(4), 111-132.

Word Tree Guide. (n.d.). Retrieved August 7, 2008, from http://services.alphaworks.ibm.com/manyeyes /page/Word_Tree.html

Notes

[1] These data visualization tools offer varying degrees of functionality. For the researcher interested in trying them out, the URLs are as follows: Many Eyes (http://manyeyes.alphaworks.ibm.com); TagCrowd (http://tagcrowd.com); Neoformix (http://www.neoformix.com); Swivel (http://www.swivel.com)