Situational Awareness in Mass Emergency:

A Behavioral and Linguistic Analysis of Microblogged Communications

by

Sarah Elizabeth Vieweg

B.A., University of Illinois, 1999

M.A., University of Colorado, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Technology, Media and Society

Alliance for Technology, Learning and Society (ATLAS) Institute

2012

UMI Number: 3508050

UMI

Dissertation Publishing

UMI 3508050

ProQuest®

This thesis entitled:
Situational Awareness in Mass Emergency:
A Behavioral and Linguistic Analysis of Microblogged Communications

written by Sarah Elizabeth Vieweg
has been approved for the Alliance for Technology, Learning and Society Institute

Leysia Palen (Chair)

Kenneth M. Anderson

Clayton Lewis

James H. Martin

Martha Palmer

Date

The final copy of this thesis has been examined by the signatories, and we find that
both the content and the form meet acceptable presentation standards of scholarly
work in the above mentioned discipline.

ABSTRACT OF THE DISSERTATION

By Sarah Elizabeth Vieweg
(Ph.D. Technology, Media and Society; ATLAS Institute)

Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of
Microblogged Communications

Dissertation Directed by Professor Leysia Palen

In times of mass emergency, users of Twitter (a popular microblogging service) often communicate information about the event, some of which contributes to situational awareness. Situational awareness refers to a state of understanding the "big picture" in time- and safety-critical situations. The more situational awareness people have, the better equipped they are to make informed decisions. Given that hundreds of millions of Twitter communications (known as "tweets") are sent every day and emergency events regularly occur, automated methods are needed to identify those tweets that contain actionable, tactical information. The purpose of this dissertation is to explore how Twitter is used in mass emergencies, and to inform mechanisms for automatically identifying information that contributes to situational awareness.

This dissertation provides a three-part analysis of Twitter communications from four different mass emergency situations. I first perform discourse analysis on tweet content to uncover and explain the information Twitter users communicate during mass emergencies. This analysis serves as the basis for a qualitative coding scheme of specific information types that lead to situational awareness. Second, using this coding scheme, tweets from each of the four emergency events are coded by multiple annotators and inter-annotator agreement rates are calculated. The results of this process provide an overview of the information Twitter users contribute in these emergency situations. My final analysis identifies linguistic characteristics of those tweets that convey situational awareness information and serves as a foundation for natural language processing classifiers that can be trained to identify such tweets.

In sum, this dissertation contributes (1) an analysis of tweet content that is relevant to situational awareness, (2) an overview of the information Twitter users communicate that contributes to situational awareness, and (3) a computational linguistic resource for the development of natural language processing tools that automatically extract tweets containing information relevant to situational awareness.

DEDICATION

To my parents, who taught me the importance of friendship,
and the value of laughter.

ACKNOWLEDGMENTS

school in the fall of 2005—have been wonderful writing partners, editors, listeners and companions.

This PhD is also the result of much love and encouragement from family. My mom has supported me through every adventure I've taken, and every journey I've traveled. She is a smart and funny woman, and a kind and loving mother. I treasure her witty observations about life, and her ability to provide perspective. My dad passed away before my graduate career began, but his presence is with me always. He was a truly kind man who is greatly missed by many, many people. He led by example and taught me to never take myself too seriously. Finally, I want to thank my partner in life, Adam Hodges. His constant support and love got me through this process. The notes taped to my desk, the frequent talks and emails and the many hugs always came at the right time, and were a constant source of encouragement.

CONTENTS

TABLES

Table

FIGURES

Chapter 1

Introduction

The use of microblogging services during mass emergencies is on the rise. People use these services to communicate tactical, actionable information that can help them understand an emergency situation. In this dissertation, I analyze communications broadcast via the popular microblogging service Twitter. I identify what information people communicate and how they linguistically construct that information. These findings provide the foundation for natural language processing techniques that can classify particular types of information, and help affected populations gain *situational awareness* during times of mass emergency.

1.1 Situational Awareness

During mass emergencies, affected populations construct an understanding of the situation based on incomplete information (Harrald and Jefferson, 2007). Often, potential victims, members of formal response agencies, and/or concerned outsiders gather available information before they decide what action to take regarding an emergency. This process of gathering information, or situational assessment, leads to a state of *situational awareness* (SA).

Situational awareness is a state of knowing what is happening in your immediate environment and understanding what that information means for a particular situation—including perception of the elements in the environment and how those elements relate to each other (Endsley et al., 2003; Sarter & Woods, 1995).

Attaining situational awareness requires understanding "the objects in the region of interest," as well as "knowing the relations among the objects that are relevant to the current operation" (Matheus et al., 2003).

In her research on situational awareness in studies of air traffic control, aviation, large-systems operation and similar settings, Mica Endsley (1995) states:

> [O]perators must do more than simply perceive the state of their environment. They must understand the integrated meaning of what they are perceiving in light of their goals. Situation awareness, as such, incorporates an operator's understanding of the situation as a whole, forming a basis for decision making.

Endsley's point indicates that situational awareness is a complex process that requires perception and comprehension of the elements in one's environment and leads to predictions of what will happen in the near future (Endsley, 1995; Endsley et al., 2003).

Gaining situational awareness involves familiarity with the elements in the environment, and understanding the significance of those elements. In other words, in gaining situational awareness people grasp the meaning of the information they are receiving (Harrald and Jefferson, 2007). During mass emergencies, the knowledge people process may include the status of a hazard agent, damage done to buildings and infrastructure, the location of evacuation centers, and the number and location of injured people and/or animals. Such knowledge provides decision-makers with information that contributes to an understanding of emergency situations, and can help them decide what actions to take.

In their research on human behavior in disaster, Fritz and Marks (1954) state:

> [T]he *immediate* problem in a disaster situation is neither uncontrolled behavior nor intense emotional reaction, but deficiencies of coordination and

organization, complicated by people acting upon individual…definitions of the situation.

Fritz and Marks' assertion that people define disasters individually, which can lead to problematic outcomes, speaks to the need for common situational awareness among affected populations. Complete information is not attained during mass emergency, else it would not be a mass emergency (Sood et al., 1987; Harrald and Jefferson, 2007; Palen et al., 2011). However, the more information people have and the better their situational awareness, and the better equipped they are to make tactical, strategic decisions.

Sonnenwald and Pierce (2000) look at a command-and-control exercise for army battalions and explain the processes of attaining 1) individual situational awareness, 2) *intragroup* situational awareness, whereby SA is shared *within* groups, and 3) *intergroup* SA, which is shared *across* groups. These processes combine to create *interwoven situational awareness*, which advances "response to dynamic, constraint-bound situations" (ibid.) Sonnenwald and Pierce highlight the need for people faced with uncertain situations to collect and analyze information from a variety of sources, and subsequently construct a coherent picture of events while under time constraints. The authors point to the need for individuals with different specialties and knowledge to collaborate to execute optimal plans.

Though aviation and formal military operations differ from mass emergencies, they have several things in common: the need to combine information from various sources, the requirement to quickly understand a situation, and the ability to act with as much information as possible within a constrained time period. However, when

facing mass emergencies, members of the public lack the same communication infrastructure and organizational protocols as military personnel; locating and assessing information are not preordained tasks that involve tacit, methodical procedures. Yet research on Twitter use during mass emergencies shows that despite the lack of a formal system to communicate information during mass emergency, people turn to Twitter to broadcast information that can potentially contribute to situational awareness (Vieweg et al., 2010).

1.2 Using Twitter to Attain Situational Awareness

Twitter is a popular microblogging service that allows users to send 140-character messages, known as "tweets," for anyone to view.[1] Recent research shows that users converge on Twitter in times of mass emergency (Palen et al., 2009; Mendoza et al., 2010; Starbird et al., 2010; Vieweg et al., 2010; Sarcevic et al., 2012).

Twitter is one means by which people share information that contributes to situational awareness during times of mass emergency. On Twitter, people may find information such as updates on property damage, reports of personal injury, areas under evacuation, and sheltering sites. Additionally, people may find information about where to direct monetary and material donations, and read messages of support and prayer. The former type of information—that which is tactical in nature—is the focus of this dissertation.

People experience disasters differently, and may require different information types to best handle their own state of affairs. The goal is for those faced with a mass

---

[1] Some Twitter users maintain private accounts, which allows only approved followers to view their tweets.

emergency situation to make good, timely decisions with the most information possible. Information communicated via Twitter can help attain this goal.

Additionally, Mileti (1999) considers collective response in mass emergency and presents a five-step strategy that people experience when coping with hazards:

1. Assess hazard vulnerability
2. Examine possible adjustments
3. Determine the human perception and estimation of the hazard
4. Analyze the decision-making process
5. Identify the best adjustments, given social constraints, and evaluate their effectiveness

These steps offer a high-level understanding of how people cope with the immediate aftermath of mass emergency events, and we see these same coping behaviors exhibited in Twitter communications during recent mass emergencies (Vieweg et al., 2010). Mileti's conceptualization explains how Twitter activity fits into the processes people attend to when dealing with worrisome—often dangerous—situations.

We know people use Twitter to offer information, ask questions and request help during mass emergency situations. New questions are arising as to how microblogging services are changing or augmenting human behavior in disaster situations, and how situational assessment and the eventual state of situational awareness are changing with the use of Twitter.

1.3 Contributions

This research offers the following contributions: (1) a description of the information contributing to situational awareness that is broadcast via Twitter; (2) an outline of information types relevant to situational awareness, based upon the examination of Twitter communications, sociology of disaster literature, and

government documentation; (3) a description of information types communicated via Twitter during four mass emergency events; (4) an outline of linguistic features used in tweets that communicate situational awareness information that provide a foundation for the development of natural language processing techniques to classify tweets based on their content.

1.4 Overview of Dissertation Chapters

After this introductory chapter, the dissertation is comprised of five research chapters, and a final concluding chapter. Chapter 2 provides a review of relevant literature on crisis informatics, microblogging, discourse analysis, and computational linguistics. Chapter 3 outlines the research design and methods. Chapter 4 provides an analysis of the types of information that contribute to situational awareness communicated via Twitter, as well as an analysis of sociology of disaster literature and government publications on mass emergency. Chapter 5 examines how the information types identified in Chapter 4 are communicated in four different mass emergency situations. Chapter 6 describes the analysis and identification of specific verbs and VerbNet classes that have high representation in tweets that contain situational awareness information, and how these findings can be used for the development of natural language processing techniques. Chapter 7 concludes this dissertation by summarizing findings and offering future directions for this research.

Chapter 2

Online Communication in Mass Emergency

In recent years, the proliferation of social media and microblogging services has provided researchers new opportunities to study human behavior in times of mass emergency. Prior to, during and immediately following the impact of a natural hazard (and other types of disaster situations), many turn to social media websites and/or microblogging services to communicate (Palen et al., 2009; Qu et al., 2009; Starbird et al., 2010; Vieweg et al., 2010; Starbird and Palen, 2011). Users of services like the popular microblogging service Twitter may provide personal status updates; inquire about the wellbeing of others; and ask about road closures and evacuation routes, property damage, sheltering options for people and animals among other inquiries and statements. They may offer support and/or prayer and request monetary and/or material donations (Vieweg et al., 2010). Timely, often crucial, information is broadcast via Twitter during mass emergency. This information has the potential to provide affected populations with more or additional knowledge than they would otherwise have access to in these situations—information that may lead to situational awareness.

Studying and understanding human behavior as it takes place via Twitter involves a multi-disciplinary approach that draws from several fields. Sociologists, anthropologists and information scientists, among others, study the activities, thoughts and reactions that take place in response to mass emergency with the goal of understanding and theorizing about human behavior in these time- and safety-critical

situations. Methods researchers use to gain knowledge about mass emergency and its effects include (but are not limited to): conducting interviews and/or surveys, analyzing artifacts, historical analysis, and observing activity during disaster situations. In the following sections, I discuss the various areas of research that underpin the study of Twitter communications in mass emergency from behavioral, linguistic and computational points of view. This chapter is divided into two parts. In Part 1, I discuss disaster and information communication technology (ICT) research. In Part 2, I discuss the approach and methods I use to understand microblogging behavior that takes place during mass emergency.

2.1 Part 1: Disaster and ICT Research

The first part of this chapter provides an overview of foundational research in the study of ICT use in disaster situations. Part 1 covers crisis informatics, microblogging and the sociology of disaster.

2.1.1 Crisis Informatics

Crisis informatics is a field of study first identified by Hagar (2006), and was subsequently articulated and developed by Palen et al. (2007; 2009), who describe crisis informatics as a field that:

> [E]xtends consideration of emergency response to not only include official responders (who tend to be the focus in policy and technology matters) but also members of the public. Therefore, crisis informatics views emergency response as an expanded social system, where information is disseminated within and between official and public channels and entities. Crisis informatics wrestles with methodological concerns as it strives to develop new theory and support sociologically informed development of both ICT and policy.

By attending to the integral role that members of the public play in crisis situations,

the field of crisis informatics "views emergency response as a social system where information is disseminated within and between official and public channels and entities" (Palen et al., 2010). Crisis informatics also includes research on uses of technology other than those that use computer-mediated communication. For example, Hagar and Haythornthwaite (2005) wrote about farmers' reliance on local radio stations to gather information during the 2001 foot and mouth disease outbreak in the United Kingdom. Overall, the body of literature on crisis informatics encompasses research on the use of information generation and consumption via technology during times of crisis.

Much of the study of technology use in crisis focuses on Web activity. The September 11, 2001 terrorist attacks in the United States spurred research in an attempt to understand the role of the Internet in crisis communications. In the aftermath of the attacks, many turned to the Internet for a variety of purposes: to gather information, to check on the wellbeing of others, to offer support and sympathy and to show solidarity with victims and their families (Schneider and Foot, 2004). At the time—over ten years ago—Internet users began to use and appropriate online tools that allowed them to share their experiences and communicate with others about the crisis situation; results from a survey conducted by Schneider and Foot about Internet use related to the September 11 attacks show that almost one fifth of respondents claimed that information obtained via the Web helped them "a lot" (ibid.)

The Internet has changed how people gather information—in times of mass emergency and otherwise. Twenty-five years ago, Sood et al. (1987) wrote: "news

producers are usually regarded as the only credible information source for the public during a disaster event." The credibility of mass media outlets as opposed to (or in addition to) members of the public is outside the scope of the research presented here. However, the relatively recently-acquired ability for everyday citizens to act as information sources for those outside their personal friend and family networks is changing how information is distributed in time- and safety-critical situations. These changes lead to new ways for affected populations to gain situational awareness when faced with or in the midst of a mass emergency. One of the conclusions reached by Schneider and Foot (2004) in their research on Internet use following the September 11 attacks is to show "the importance of the internet, and particularly the Web, *as a significant component of the public sphere*, enabling coordination, information sharing, assistance, expression and advocacy in a crisis situation" (ibid., emphasis added.)

In recent years, researchers of computer-mediated communication have taken an interest in studying social media and microblogging activity during mass emergencies. Research in this area looks at how citizens used popular discussion forums to communicate during the 2008 Sichuan Earthquake (Qu et al., 2009), the use of information and communication technology by affected communities during the 2007 Southern California Wildfires (Shklovski et al., 2008), and the use of Twitter during the 2010 Chile earthquake (Mendoza et al., 2010). Additionally, recent research I performed with my colleagues considers micro-blogging activity during the 2009 Red River flood (Starbird et al., 2010). This work identifies "mechanisms of information production, distribution and organization" (ibid.) that took place among

Twitter users immediately before, during and after the flood occurred. In this research, we point to the collective behavior that takes place in mass emergencies, and how such behavior is manifest when we examine the Twitter activity of those users concerned with this specific flooding event. My colleagues and I also consider the use of Twitter during the same 2009 Red River flood, as well as during the 2009 Oklahoma grass fires (Vieweg et al., 2010). In this research, we focus on information that contains "situational update" information broadcast by Twitter users during each event, and identify features of information found in tweets that provide the foundation for information extraction techniques. When faced with a mass emergency, people are prone to increased information seeking—the affordances of services like Twitter complement this behavior by providing the potential for up-to-the-second information from a vast extent of sources.

When a wildfire spreads or a hurricane is predicted to make landfall, people in the affected area require information such as whether and how to evacuate, weather predictions and sheltering options. Affected populations also have questions, and services like Twitter provide an outlet for them to quickly ask, and potentially receive, answers. People who use CMC applications have "access to a much broader pool of information sources" than those who rely upon conventional means to ask questions, such as directly speaking to neighbors or friends (Sproull and Kiesler, 1993, p. 133). In times of mass emergency, the option to rapidly present to and survey a large audience is of great worth to those offering and seeking specific information while under severe time constraints.

2.1.2 Collective Behavior

Throughout this dissertation, I use the term *mass emergency* to describe the societal impact that floods, fires and earthquakes and other natural hazards have when they impinge upon social, physical and built environments. A mass emergency is "an unexpected or undesirable event which requires the resources from most or all municipal departments and limited assistance from outside agencies may be needed" (Drabek, 1996). To expand upon this definition, Quarantelli and Hultaker (1986) state that mass emergencies "are often created by natural disaster agents, technological accidents, violent intergroup conflicts, shortages of vital resources, acute and chronic environmental threats, and other kinds of major hazards to life, health, property, well being, and everyday routines." It is important to note that mass emergencies do not include those that do not exceed capacities of local agencies and do not significantly affect the built environment or social order.

The above definitions are broad and encompass events ranging from natural hazards to human-induced emergencies (such as those involving terrorism). However, what they do not include is an indication of a temporal element. For this, I turn to Fischer's (1998) explanation of the "five stages in the life cycle of a disaster," which can be extended to the present discussion on mass emergency. These stages outline the activities people engage in at different times during mass emergencies, and are helpful for understanding human reactions.

> *Pre-impact period*: some preparatory activity may take place if area residents are aware of the need, are given enough warning, and if they are properly equipped.

Impact period: this is usually the shortest, but most dangerous period in the life cycle of a disaster. This is the period during which a hurricane hits land, a tornado touches down, a fire affects the built environment. Fischer also notes that some disasters have prolonged periods of impact, such as floods.

Immediate post-impact period: search and rescue activities commence, necessary repairs and reconstruction efforts are identified and tended to (e.g. telephone service, electricity, etc.), official emergency response takes place, local, state and federal agencies become involved where necessary, and the media begin to report about the event.

Recovery period: cleanup efforts are completed, infrastructure is restored, insurance claims are filed by those affected, reconstruction begins, and societal norms begin to be re-established.

Reconstruction period: reconstruction of the affected area is completed, residents return to a sense of normalcy.

Additional attention to the socio-temporal aspects of disaster are discussed by Palen and Liu (2007), who refer to Dynes' (1970) socio-temporal stages of disaster in their discussion of civilian actions that took place during several mass emergencies, particularly with regards to communications via information communication technology (ICT). They explain that "citizen-to-citizen communications…during and following a disaster arise out of needs to help and be helped. Its forms depend on how the physical characteristics of the disaster agent affect the built and social environment, which in turn results in different spatial and temporal arrangements for communication." We now know that part of these stages of disaster involve people turning to ICT to communicate and gather and disperse information. Understanding the periods of mass emergency, along with the actions people take during these periods, is important for understanding the context for Twitter communications that take place during mass emergency situations.

Research in the sociology of disaster provides the background for understanding what happens "on the ground" during times of mass emergency as I study communications in "the Twitterverse" during and immediately following these events. Instances of collective behavior during crises include physical convergence to a specific geographic location, heightened information search from multiple sources, and the spread of information (Dynes and Tierney, 1994; Kreps and Bosworth, 1994); the activity that takes place among users via Twitter during times of mass emergency is also a form of collective behavior that may involve instances of self-organizing, and information gathering and dissemination (Starbird et al, 2010; Starbird and Palen, 2011; Vieweg et al., 2010). The information communicated via Twitter may provide people with information that can lead them to a state of situational awareness, and this information is often the product of collective behavior.

2.1.2.1 Rumor

One type of collective behavior that frames the discussion regarding the information contained in Twitter communications is that of *rumor*. The popular conception of rumors is that they include malicious, sometimes cruel and often untrue accounts of people, activities or events. However, scholars who study rumor depart from the common understanding, and treat rumor as a person-to-person communicatory process that takes place through informal channels, involving information that is unverified by official sources (Turner, 1994, p. 244). Mass emergencies are characterized by a lack of information and collective ambiguity about the event, yet affected populations need to know what is happening so they can make decisions. In these situations of information dearth, when institutionalized

sources (e.g. mass media channels, formal response agencies) are unable to provide necessary information, people generate rumors (Aguirre and Tierney, 2001). In times of informational ambiguity combined with high levels of anxiety, the information-seeking behavior of people increases, and rumors begin to spread.

In his landmark sociological study of rumor, Tamotsu Shibutani (1966) defines rumor as:

> A recurrent form of *communication through which men (sic) caught together in an ambiguous situation attempt to construct a meaningful interpretation of it by pooling their intellectual resources*. It might be regarded as a form of collective problem solving (p. 17, emphasis in original).

Much of the information communicated via Twitter during mass emergency can be viewed as rumor—information that is not verified by an official source which is sent via an informal channel and which provides a collective picture of an event. During crisis events, affected populations find themselves in uncertain and often dangerous situations. They recognize common goals—one of which is to attain information—and work together to reach them. The collective transaction of information that takes place through Twitter may help people understand the situations they find themselves in. When formal response organizations are unable to provide people with information or advice in times of crisis, affected populations may act on information they receive from rumor (Shibutani, 1966, p. 34), which in the present age, may be communicated via Twitter.

If we think about Twitter activity during mass emergency, the study of rumor is helpful in theorizing about the information people communicate, as well as the goals that may be reached with regards to attaining situational awareness. Twitter activity involves many people who utilize an informal communication channel to

gather and disperse information about an ambiguous situation; this information can help attain situational awareness, but complete knowledge of an event is never possible. Therefore, affected populations make *satisificing* decisions (Simon, 1996, p. 27). In times of mass emergency, it is difficult to know how to react or what actions to take, so people work to make satisficing, or "good enough" decisions and they may incorporate information communicated via Twitter in that process.

In times of crisis, people are thrust into situations that require immediate action; they must quickly assess their surroundings, understand what is happening around them, and make decisions. This process of contributing to the "big picture" that describes an event, which is a collective effort, is one way to describe the activity that takes place via Twitter as people attempt to attain situational awareness.

2.1.2.2 Tradition of Command-and-Control

The discussion of the use of ICT to attain situational awareness in mass emergencies would be incomplete without mentioning formal emergency response organizations. Formal emergency response in the United States is governed by a command-and-control model (Dynes, 1994; Neal and Phillips, 1995). Research on the command-and-control structure found in emergency response shows that it erroneously portrays mass emergency situations. Below are excerpts which detail some of the inaccurate assumptions made by members of the public, formal response organizations and others, which have been highlighted by scholars studying disaster response:

> [I]nformation outside of official channels is lacking or inaccurate (Britton, 1989a, p. 13 in Dynes and McEntire, 2003)

> [C]itizens are inept, passive or non-participants in disaster operations (Dynes, 1994, pp. 142, 147; Britton, 1989a, p. 13; Mileti, 1989, p. 67 in Dynes and McEntire, 2003)

> [A]d hoc emergence is counterproductive (Neal and Phillips, 1995, pp. 327-8; Mileti, 1989, p. 67; Dynes, 1994, p. 147 in Dynes and McEntire, 2003)

However, sociologists of disaster have long drawn attention to the activities of civilians in times of mass emergency, and point out the integral role they play in response efforts (Tierney et al., 2001). First responders in times of mass emergency are not the fire brigade nor the police force; community members from local and surrounding areas are the earliest to arrive at the scene of mass emergency situations—they administer first aid, transport victims to hospitals and clinics, and perform search and rescue operations (Palen and Liu, 2007).

The advent of Twitter and other forms of Web 2.0 technologies have introduced a new means through which emergent behavior can take place during mass emergencies. The publically available nature of the information shared via Twitter is causing emergency managers to take note, and some are realizing the positive impact microblogging services can have. In September 2010, US Federal Emergency Management Agency (FEMA) director Craig Fugate, said "social media provides FEMA and other emergency management agencies the chance to get 2-way input from disaster survivors to better shape response to their needs in a crisis." With regards to the 2010 San Bruno, CA Fire, Fugate stated that he "got better situational awareness [from Twitter] before we got official word…Four or five years ago I wouldn't have gotten that quality of information" (Spellman, 2010).

Fugate's statements speak to the prevalence—and potential value—of Twitter use in mass emergency. His remark about emergency management's ability to communicate via Twitter with disaster survivors to better attend to their needs gets at the potential for Twitter to empower members of the public, to give them voice in how emergency response transpires. A potential benefit of Twitter will be to connect members of the public and formal response organizations through its ability to provide an information-sharing outlet available to all.

The proliferation of social media and microblogging services makes the vital role played by members of the public in mass emergencies—which has long been recognized and commended by sociologists (Tierney et al., 2001)—increasingly difficult to disregard. The research on collective behavior is integral to understanding where Twitter activity is positioned within the space of mass emergency. This dissertation contributes to furthering the study of collective behavior in disaster by describing Twitter activity in mass emergency situations.

2.1.3 The Role of Microblogging

We know that every day, millions of people use Twitter to communicate a variety of information. The access to millions of other users, instantaneous publication of one's message and character limitation all make for an as-yet-unseen communicatory phenomenon. Beginning in 2007, researchers began to study Twitter and provide high-level explanations of how people were using it. Early work focuses on the growth in use, basic user demographics, user networks and usage trends (Barnes & Böhringer, 2009; Huberman et al., 2009; Java et al., 2007). In an attempt to provide an understanding of who of Twitter are (among other findings), Java et al.

(2007) identify primary types and claim that Twitterers tend to take on one of three roles: *information source, information seeker*, or *friend*. Java et al. (ibid.) also explain the intent behind Twitter use, and provide a taxonomy of user intentions, which include *daily chatter, conversations, sharing information/URLs*, and *reporting news*. The later of this research by Java et al. takes a *user-centered*, as opposed to *content-centered* approach, which is the focus of this dissertation.

In 2010, Naaman et al. (2010) provided a study of Twitter messages. In this paper, the authors take an inductive approach to qualitatively analyzing tweet text and find that there are nine major information categories that describe tweet content; four categories comprise the majority of tweets. While this paper provides results which extend earlier findings on the importance of network in Twitter use (Java et al., 2007), Naaman et al.'s study was performed using 3379 tweets which were qualitatively coded by analysts. Though it is possible that performing the same study on a larger sample would produce similar results, the small number of both users and tweets is a limiting factor in this research.

Additional research on tweet content was done by Honeycutt and Herring (2009), who analyze Twitter messages to understand whether and how conversations start and are maintained, and the importance of the "@" symbol in carrying on conversations via the "noisy" Twitter environment. The authors provide a breakdown of tweet content and explain the potential of tools like Twitter for online collaboration and information sharing. Their results include findings about the language of Twitter, how users employ "@" in Twitter discourse, tweet content, and how users carry on conversations. Honeycutt and Herring provide an in-depth look at Twitter behavior

which leads to claims regarding the design of microblogging systems, and successful adoption of new technology. Though the authors performed analysis on a sample of 200 tweets, their research serves as a starting point for future work that examines how microblogging services are used and may be improved or enhanced for a variety of purposes.

The work of Naaman et al. (2010) and Honeycutt and Herring (2009) delves into the subject matter of Twitter communications. These researchers take an interest in understanding not just who is using Twitter, how frequently and what their online social network looks like, but in the *information* that is being conveyed via Twitter. These studies recognize that though Twitter communications are constructed through the written word, there is a sense of orality to these short bursts of text; they are often conversational and informal in nature. Honeycutt and Herring's (2009) work is particularly situated to uncover the position of Twitter at the interstices of written and oral communication in their application of traditional conversation analysis methods to Twitter discourse.

The vein of inquiry taken by Naaman et al. and Honeycutt and Herring is in line with Project EPIC's research on Twitter use during mass emergency situations, in which we consider information production and distribution, (Starbird et al., 2010) as well as the specific types of information communicated via Twitter that have the potential to contribute to situational awareness (Vieweg et al., 2010). This work takes a critical approach toward not only examining and explaining Twitterer behavior, but toward contributing theoretical understandings of how people use Twitter in mass

emergency situations, and about the role of Twitter in mass emergency going forward.

Following on the work of the Project EPIC team, in 2010, Mendoza et al. (2010) published research on the use of Twitter during the 2010 Chile earthquake. The authors were particularly concerned with information propagation, and pay close attention to re-tweets. Mendoza and colleagues conclude that the relevance of a tweet to a specific community is related to how often it is retweeted (ibid.) Research on the use of Twitter during mass emergency will continue to grow as more disasters and crises occur, and more people turn to Twitter to gather and disperse timely information.

The research described above provides background on the questions scholars are asking about Twitter use and its implications. Much of this early work is descriptive in nature; researchers ask questions about who uses Twitter, for what purposes, how their communications are structured and how users are connected to each other. Subsequent research has worked to penetrate more deeply into the content of Twitter communications, how it is appropriated for various purposes by users, and the role Twitter will play in collective, collaborative events going forward.

This dissertation expands this body of literature by providing an inductive, in-depth analysis of Twitter communications that take place during mass emergency; an analysis that considers both the meaning of the tweet content as well as tweet construction at the linguistic level. Research on the use of Twitter during mass emergency will lead to greater understanding of its role during such situations; it will also begin to shed light on the potential changes to mass emergencies that may be

taking place due to the incorporation of Twitter into understanding and responding to such situations by a variety of populations.

2.1.3.1 The Potential of Twitter in Disaster

When Twitter was created, no one could have predicted how it would be used or the role it would have in mass emergency situations or otherwise. We now know that it is a service used by many for various reasons, one of which is to communicate potentially vital, timely information during disasters. One objective of this dissertation is to consider such aspects of Twitter, and how to harness the power of this communication phenomenon can contribute to mitigation of emergency situations.

The introduction of new technology can influence and alter how people interact with one another. However, technologies are not deterministic; they are negotiated and directed by technology users (Orlikowski, 1992). Microblogging services like Twitter have not changed essential characteristics of human behavior. People have always communicated with each other; Twitter is a new communication channel that extends existing behavior. However, when someone uses Twitter to communicate, its affordances affect the details of how that communication is arranged and how it influences or impacts others. A Twitter user can reach millions of people, and the possibility for feedback, dialogue and communication proliferation is enormous given the instantaneous one-to-many communication that Twitter offers. Twitter and other microblogging services have an impact on how people communicate and who they communicate with by the very nature of their technological affordances; the speed, accessibility, ability to search for specific

information and opportunity for one-to-many communication are all aspects of microblogging services.

Services like Twitter offer users a way to communicate through text that involves conversational qualities and ephemerality. These attributes, combined with rapid broadcasting, provide researchers with new ways to understand human communication. Twitter also allows for new perceptions of information movement and dissemination. Users of microblogging services are realizing value in the ability to quickly absorb and easily distribute large quantities of small fragments of information in real-time, and the benefits of being able to search the Twitter space for previous communications. We find ourselves in an opportune time for understanding how services such as Twitter will affect how people communicate, gather and disperse information during times of mass emergency.

2.2 Part 2: The Nuance of Online Communication

The second part of this chapter provides an overview of the linguistic and computational aspects of this research. This includes background on computer-mediated discourse, discourse analysis, pragmatics and natural language understanding.

2.2.1 Communication in Context

Communicating about disaster (or any other topic) is not always explicit or straightforward. Human communication is a nuanced process that involves more than knowledge of a sound system, lexicon and a grammar. When we communicate with others, we participate in an activity that takes into account the deliverer of an

utterance[2], the meaning of words, the mode of delivery (e.g. verbal or textual) and the surrounding environment. The focus of this dissertation is on a specific type of communication which occurs in a particular setting: computer-mediated discourse in mass emergency.

2.2.1.1 Computer-Mediated Discourse

Linguist Susan Herring (2001) defines computer-mediated discourse (CMD) as "the communication produced when beings interact with one another by transmitting messages via networked computers." Herring goes on to define the *study* of computer-mediated discourse as "a specialization within the broader interdisciplinary study of computer-mediated communication (CMC), distinguished by its focus on *language and language use* in computer-networked environments, and by its use of methods of *discourse analysis* to address that focus" (ibid., emphasis in original). The research I present here includes the use of discourse analytic methods on computer-mediated text.

The earliest research on CMD dates back to the mid-1980s, when researchers began to examine this new type of communication. Scholars were initially interested in the empirical study and description of computer-mediated discourse, which led to the challenge of popular conceptions which claimed that CMD was "anonymous," "impersonal," and "egalitarian" (Herring, 2001). This early research serves as the precursor for later work on the study of CMD in different contexts, and how users of

---

[2] I use 'utterance' in the linguistic sense, defined as an uninterrupted chain of spoken or written language.

technologies appropriate CMD to attain various goals (del-Teso-Craviotto, 2006; Panyametheekul and Herring, 2003).

Computer-mediated discourse is sometimes viewed as "speech like" in spite of it being a written form of communication—there is no clear delineation between spoken, written and computer-mediated discourse. Those who communicate via CMD may experience it as a blend of the spoken and the written, but CMD also comes with its own set of constraints and affordances. Computer-mediated discourse is affected by the features of the technology used to communicate it, and while those technologies are not necessarily deterministic (meaning that users have little to no ability to manipulate the built-in constraints of the system), it is important to note the role that the medium has on the message and to consider how technological affordances may affect message content (Herring, 2001).

2.2.2 Pragmatics

To successfully communicate, humans combine semantic knowledge, which is knowledge of the meaning of words, with *pragmatic knowledge*, which is—broadly speaking—how humans combine semantic knowledge with other types of knowledge to make sense of what they hear or read. A clear, agreed-upon definition has yet to be established for *pragmatics*, and to illustrate some of the various approaches toward the study of pragmatic phenomena, I offer three definitions provided by Levinson (1983):

> 1. Pragmatics is the study of the relations between language and context that are basic to an account of language understanding (p. 21)

2. Pragmatics is the study of the ability of language users to pair sentences with the contexts in which they would be appropriate (p. 24)

3. Pragmatics is the study of deixis (at least in part), implicature, presupposition, speech acts, and aspects of discourse structure (p. 27)

For purposes of the research presented here, I use the term *pragmatics* in a general sense, and define it as the study of meaning in *context*. I draw attention to the notion of context because it too is not an easily defined notion. In her discussion of context, Ochs (1979) writes:

> The scope of context is not easy to define….one must consider the social and psychological world in which the language user operates at any given time. It includes minimally, language users' beliefs and assumptions about temporal, spatial, and social settings; prior, ongoing and future actions (verbal, non-verbal), and the state of knowledge and attentiveness of those participating in the social interaction at hand (p. 5).

I use Ochs' description as the basis for how I view context in the forthcoming analysis of Twitter communications described in the following chapters.

To introduce how humans rely on pragmatics and context to communicate, in the sections that follow I explain various types of knowledge and pragmatic phenomena as they appear in Twitter communications.

2.2.2.1 Background Knowledge

To make sense of human communication, we rely on *common-sense* or *background knowledge* in understanding utterances. Saeed (2009, p. 201) treats such knowledge as information the speaker "might calculate others would have before, or independently of, a particular conversation, by virtue of membership in a community." The *community* to which Saeed refers may be any number or type of communities: speakers of the same language; residents of the same city, region or

state; work colleagues; members of the same political group, and so on. Community membership is indicative of shared knowledge, and language users rely on that shared knowledge when they wish to communicate in their interactions with others. To demonstrate just how ensconced shared cultural knowledge is in our communicatory practices, I turn to a mundane, everyday exchange between two people:

> Person A: I'm running late.
> Person B: I'll lend you my car.

This example makes sense to English speakers because we understand that a car is a mode of transportation that conveys people from point to point. Using a car is faster than walking or biking, and does not require the traveller to wait for public transportation. We understand this exchange due to the background knowledge we possess regarding timeliness and transportation.

In addition to background knowledge and context, added pragmatic phenomena speakers and hearers often rely on are those of *markedness* and various types of *inference*. Below, I illustrate how this pragmatic knowledge is integral to successful communication.

2.2.2.2 Markedness

I borrow the notion of *markedness* from linguistics and philosophy. "Markedness is a very broad notion applying to all levels of analysis. Generally speaking, a marked form is any linguistic form which is less usual or less neutral than some other form—the unmarked form—from any number of points of view" (Trask, 1999). In this dissertation, I use markedness "to explain how certain places, landmarks or items become taken-for-granted and expected when referred to in more

general terms" (Vieweg et al., 2010). For example, in the 2009 Red River flood

dataset, Twitter users refer to *the river*, by which they mean the Red River; it is tacitly

understood that referring to *the river* refers to the Red River by those who are

communicating about it. For example, the following tweet from the 2009 Red River

flood dataset includes reference to *the river*:

> ```
> @ksassi (March 27, 02:52): The river is now over 40 feet and
> there has been a breach in the dike; neighborhood a few blocks
> south of us is evacuating.
> ```

Conversely, in that same dataset, we see an example of a Twitterer writing about the

Sheyenne River. It is necessary to "mark" the Sheyenne, rather than simply referring

to it as *the river* (the unmarked form).

> ```
> @Fictoris (March 23, 10:08am): The Sheyenne River went up 10
> feet in 12 hours…
> ```

This is one example of marked versus unmarked phenomena present in Twitter

communications during the 2009 Red River flood; additional examples are presented

in chapter 4.

2.2.2.3 Inference

Additional pragmatic phenomena present in tweets represent conversational

inference, which Gumperz (1982, p. 153) defines as "the situated or context-bound

process of interpretation." To understand an utterance, a hearer must do more than

grasp the meaning of words and how they are grammatically strung together.

Understanding utterances frequently requires hearers to make *inferences* which

"connect what is said to what is mutually assumed or what has been said before"

(Levinson, 1983, p. 21).

Regarding Twitter communications that take place during mass emergency, previous research shows examples of a particular type of inference known as *implicature*, which is the implication(s) that can be drawn from an utterance based on expected conversational norms (Crystal, 1997, p. 191). For example, in the aftermath of the 2010 Haiti earthquake, Twitter users refer to "bodies," as in this example:

```
http://twitpic.com/y39h9 - Bodies line roads as Haiti waits
for help++Shit is crazy! THIS IS REALITY!!
```

The content of tweets in this example explicitly tells the reader that "bodies" line roads. However, those who read the content of this tweet within its correct context understand that it implies people have died, and adequate management of the corpses of those who have passed is not available. The reader of this tweet may infer that this information serves as an indication of the grave situation facing Haiti, though no mention of mass deaths is explicitly mentioned in the tweet text. The notion of implicature is particularly germane in instances such as these, because when dealing with implicature "what is conveyed is generally far richer than what is directly expressed" (Horn, 1999).

Implicature can impact the analysis of tweets in which content is frequently shortened due to the constraints of tweet length. Often, what is meant reaches beyond the number of characters typed. When we consider the character limit of tweets, implicature provides a useful lens with which to frame information being communicated in the context of time and safety-critical situations.

2.2.2.3.2 Anaphora

Another type of inference language users commonly rely on is that of *anaphora,* which is the "process of a linguistic unit deriving its interpretation from some previously expressed unit or meaning" (Crystal, 1997, p. 19). Hearers continuously make sense of utterances that include anaphoric references—they are able to easily link pronouns with preceding or understood entities (known as antecedents) and therefore create a coherent understanding of a situation (Saeed, 2009 p. 212). An example of anaphoric usage follows:

> The train was on time, the lunch was delicious and the show was spectacular. I think it really made her happy.

In this example, the speaker uses the pronoun *it,* but which of the preceding entities does *it* refer to? *It* may refer to the train, the lunch or the show. However, what is most likely is that *it* refers to the entire situation, which we can think of as an outing or excursion. Anaphora is another phenomenon humans rely on in communication practices, and which is present in tweets which have a limited character length, as anaphora allows speakers and writers to use short pronouns to refer to previously stated or understood entities.

2.2.2.4 Interpreting Tweets

To illustrate how humans depend on pragmatic and background knowledge in our interpretation and understanding of tweets, I turn to the following tweet examples which were broadcast during the Red River flood of March-April 2009. The Twitter users who sent these tweets are local (meaning they live in or near areas under threat

30

of flood) and individual (meaning they are members of the public, and not part of any

organization, agency or other group):

1. `@adkinsjm Not yet. Deciding today whether family should leave.`
   `I plan to stay here.`

2. `@JanetStewartCBC can they blow it up?  explosives?  #redriver`
   `#flood09`

3. `Finished evacuating Doug and Jan... moved all their belongings`
   `to the top floor with help from my brothers.`

As an analyst familiar with this particular flood, I use background knowledge

of flooding situations in general, background knowledge of the 2009 Red River flood,

as well as pragmatic knowledge of how human communication works, to understand

what is happening in these tweets and how the information contained therein is

relevant to situational awareness. In the first tweet, the Twitterer is replying to

someone, as evidenced by the use of the conventional form of Twitter address "@" at

the start of the tweet; the first phrase appears to be an answer to a question (*Not yet*).

Next, the Twitterer is deciding if her/his family should leave, while s/he plans to stay

here. Yet how do I know what question s/he is answering? Why would her/his family

leave? Where is *here*?

When this tweet is analyzed in context, I understand that the phrase *Not yet* is

in response to a question that was likely asking if the Twitterer has evacuated or has

made a definite decision to evacuate; I know this due to the succeeding phrase about

deciding whether her/his family should leave the area. I also know that *here* is the

Twitterer's home, which is in a flood-prone area. I know the Twitterer lives in a

flood-prone area because evacuations are only considered in areas under threat.

Overall, the first tweet provides information that contributes to situational awareness

by telling readers that people are considering evacuation, which indicates that the flood threat is serious and precautionary measures are being taken.

The second tweet includes references to *they* and *it*. Who are *they*? What is *it*? These two pronouns are anaphoric references. With knowledge of pragmatics, and with an understanding of the circumstances of this particular flood as well as how flooding situations typically transpire in this region of the United States, I can infer that *they* are members of a formal response agency and *it* is an ice jam. Context, background information and inference provide the tools to understand this tweet—I know that only formal response agencies have the authority to use dynamite and other explosives during a mass emergency situation, and a common method for dispersing ice jams (which can wreak havoc in low-lying areas during a quick thaw) is to explode them and allow the smaller chunks of ice to flow downstream unimpeded. The second tweet contributes to situational awareness by informing readers that ice jams are causing problems, and that formal response agencies are involved in flood mitigation and prevention efforts.

In the third tweet, readers learn that the Twitterer evacuated two people, and moved their belongings to *the top floor*. Why are people being evacuated? Why are their belongings being moved? What is *the top floor*? Again, to understand the information in this utterance, I draw on my knowledge of the types of physical structures people who live in this region of the United States tend to live in, what happens to inhabited areas when rivers flood, and what precautions are typically taken. I therefore understand this tweet to explain that evacuations are taking place due to an impending flood, that basements and lower floors will experience water

seepage and damage before upper floors, hence the moving of belongings to *the top floor*, and that the evacuees live in a house with at least two stories, and the top story is most safe for items they do not want damaged. The third tweet is similar to the first one in communicating that people are evacuating, and that precautionary measures are being taken—with this type of information, others may make the decision to evacuate and/or move valuables to safe areas of their homes.

It is important to explicate the complexity involved when people decipher utterances and successfully communicate. People perform the complex operations involved in communication on a constant basis, but we do not often recognize all of the work that goes into understanding utterances and making sense of situations.

The tweets above serve an additional purpose in providing examples of the *types* of information that may lead to situational awareness in mass emergency situations. The first offers information about evacuation—we know the flood is serious enough for area residents to consider evacuating. In the second tweet, we learn that formal response agencies are involved in flood mitigation efforts, and that the situation is serious enough to warrant the potential use of explosives to destroy ice jams. In the third tweet, we again learn about a decision to evacuate, and we are also provided insight and/or advice on how to protect belongings. The types of information contained in tweets is the subject of later chapters, and serve as a foundation for how situational awareness is communicated in disaster.

2.2.2.5 Tweets and Situational Awareness

The explanation of how information is communicated via Twitter in mass emergency, and the potential for that information to lead to situational awareness leads to the thrust of this dissertation. We know people use Twitter to communicate timely, useful information during crises. We also know that during crises, people seek information from multiple sources in an attempt to make locally optimal decisions within given time constraints (Palen et al., 2009). The first objective, then, is to identify what tweets that contribute to situational awareness "look like"—i.e. what specific information do they contain? This leads to the next objective, which is to identify how information is communicated at a linguistic level. This process provides the foundation for tools that can automatically extract pertinent, valuable information—training machines to correctly "understand" human language involves the identification of the words people use to communicate via Twitter when faced with a disaster situation.

2.2.3 Natural Language Understanding

During a mass emergency, when time is limited and safety may be in question, people must act quickly to gather information and assess their situation. If an individual faced with disaster turns to Twitter to help him/her attain situational awareness, he or she must decide how to locate useful, meaningful information. At any given time, millions of tweets are broadcast on Twitter, making it difficult for a single user to locate the type of information in which he or she is interested. Therefore, it is necessary to implement computational methods to help humans find that information which is most useful to them. In this section, I provide a brief

overview of natural language process (NLP) methods used to automatically identify

and organize text. Greater detail about the NLP aspect of this research is provided in

chapter 6.

2.2.3.1 Data Annotation

Data annotation is an integral component to natural language processing and

natural language understanding tasks. Many of these tasks, including named entity

recognition (Bikel et al., 1999), word-sense disambiguation and semantic role

labeling (Hovy et al., 2006; Hwang et al., 2010) and syntactic parsing (Gabbard et al.,

2006), make use of the supervised machine learning approach, which relies heavily

on annotated corpora. In the supervised approach, the manually labeled data is used to

train classifiers; it is also used as the "gold" standard in evaluating the performance of

these classifiers. Additionally, manually annotated data can be used to provide rich

feature sets that classifiers use to automatically categorize and label text. The

supervised approach is a good example of how natural language processing relies on

annotated corpora because the classifier is given the gold labels as examples for

learning purposes, then in the testing stage, the classifiers' predictions are evaluated

against the gold standard annotations.

For a machine to successfully locate tweets that contain useful, timely

information during mass emergencies, it needs to be trained to identify tweet text that

is most likely to convey such information. This process requires the identification of

the types of information tweets communicate; e.g. information about the social, built

and physical environments, or more specifically, information about, for example,

evacuations, structural damage and weather. Tweets are annotated with these various

categories, and further analysis reveals how information that leads to situational awareness is constructed linguistically. In the research presented here, this further analysis involves an examination of the verbs used to convey situational awareness information.

2.2.3.2 Leveraging Verbs

In English, the verb is generally the central element in a sentence which organizes all other elements (Manning & Schütze, 1999), and which conveys the meaning of the event taking place (Palmer et al., 2010). VerbNet (Kipper-Schuler, 2005) is a lexical resource based on verb classes originally devised by Levin (1993). VerbNet provides groups of verbs which are placed into classes based not only on their meaning but also according to their syntactic and semantic behavior. VerbNet is organized such that each class is listed along with an accompanying class number, e.g. *pour-9.5, destroy-44, entity_specific_modes_being-47.2, modes_of_being_with_motion-47.3* and so on. The closer the class numbers, the more similar the verbs, i.e. verbs in the *9* classes are more similar to verbs in the *10* classes than they are to verbs in the *51* classes. Each class in VerbNet includes examples of the verbs in context, and descriptions of various semantic (meaning) and syntactic (use) attributes (Palmer et al., 2009). The class descriptions include:

- A list of each verb in the class
- Semantic roles, which describe the conceptual relationship sentence elements have with the verb, and their typical semantic class restrictions, e.g. *agent [+animate]* which indicates an animate entity is the causer of the event described by the verb.
- Syntactic frames that explain the syntactic behavior of sentence elements by providing a formal description of their ordering in a sentence, e.g. *NP V NP PP.instrument (*noun phrase, verb, noun

phrase, prepositional phrase.instrument—The man damaged the car with a pipe.)
- Semantic predicates that attempt to capture a portion of the shared semantics of the verbs in the class, e.g. *cause(Agent, E) destroyed(result(E), Patient),* where the agent is the causer of an event that results in the patient being destroyed.

For example, the verb class *destroy-44* includes twenty-five verbs such as *damage, demolish, destroy, maim* and *ruin.* Each of these verbs can take the following semantic roles:

- agent [+int_control]: the doer of the event may be an agent that exhibits internal control, e.g. *the man*
- patient [+concrete]: the undergoer of the event is a physically tangible item, e.g. *a building*
- instrument [+concrete]: if an instrument is used in the event, it is physically tangible, e.g. *a baseball bat*

The syntactic frames in the *destroy-44* class include:

- NP V NP: e.g. The Romans damaged the city.
- NP V NP PP.instrument: e.g. The national guard destroyed the ice jam with dynamite.
- NP.instrument V NP: e.g. The explosives demolished the ice jam.

The semantic predicates in the destroy-44 class are:

- cause(Agent, E) destroyed(result(E), Patient): e.g. The Romans damaged the city.
- cause(Agent, E) use(during(E), Agent, Instrument) destroyed(result(E), Patient): e.g. The national guard destroyed the ice jam with dynamite.
- cause(?Agent, E) use(during(E), ?Agent, Instrument) destroyed(result(E), Patient): e.g. The explosives demolished the ice jam.

The essential thrust of VerbNet and Levin's original verb classes is that a set of verbs in the same class behave in much the same way; they have similar underlying semantic components and appear in similar contexts. For computers to successfully process and understand natural language text, they have to detect the events the text

describes, as well as the event participants (Palmer et al., 2010). The purpose of using VerbNet for this research is to identify verbs used in tweets that convey particular events which describe situational awareness information, as well as their VerbNet classes. We can then examine whether or not the VerbNet classes provide useful generalizations.

This phase of my research serves as the foundation for future work which involves the development of natural language processing classifiers that can predict if tweets contain situational awareness information based on features providing the correct identification of the verb and its corresponding VerbNet class.

2.2.3.3 Consistencies in Disaster

Disasters caused by natural hazards occur frequently; and though each disaster situation is unique, regularities exist as to how these events transpire. When a wildfire ignites or the threat of flood comes to pass, previous experience with these types of situations provides people with the background knowledge necessary to predict what will likely happen as the event unfolds. Correspondingly, by uncovering, understanding and describing regularities in Twitter communications, automatic methods can be used to locate vital information that we can expect affected populations to communicate in mass emergency situations. For people to gain knowledge about a disaster event that is transmitted via Twitter, computers must be able to identify pertinent information. A foundational step in this process is training classifiers to locate particular events, often described by particular verbs.

2.2.4 Summary

The use of Web 2.0 technologies in times of mass emergency includes people harnessing the power of social networking sites and microblogging services to gather and distribute timely, critical information. As I describe Twitter communications that take place during mass emergency situations and how they lead to situational awareness, I draw on each of the areas of research discussed in this chapter. Subdisciplines within linguistics and sociology, the study of situational awareness, and crisis informatics provide insight into the meaning and structure of Twitter communications in times of mass emergency and how these communications can contribute to a "big picture" understanding of these situations. The regularities in human behavior that take place during crises include how affected populations communicate; human experience is bound up in how people use language to describe events. Recognizing and closely analyzing the work language does to explain disaster situations as experienced or observed by those affected is fundamental to understanding how these events are shaped by human communication.

Chapter 3

Research Design and Methods

Understanding the role of Twitter in mass emergency as well as its potential in future emergencies is an emergent area of study. This chapter describes the research questions developed based on preliminary studies of what information Twitter users communicate in time- and safety-critical situations. I present my research questions in section 3.1, provide an overview of the research design I implemented to answer those questions in section 3.2 and summarize the research design process in section 3.3.

3.1 Research Questions

People use the microblogging service Twitter every day to communicate a variety of information, to ask questions, and broadcast news. They tell jokes, express opinions, link to interesting stories and exchange pleasantries (Naaman et al, 2010). As of June 2011, Twitter traffic was an average of 200 million tweets per day,[3] and previous research shows that when a disaster occurs, Twitter users contribute to tweet traffic by communicating about the event. Twitter users provide general reports, offer support and prayers and ask after the wellbeing of others (among other communications) (Starbird et al., 2010; Starbird and Palen, 2011; Vieweg et al, 2010). In this dissertation, I investigate the behavior of Twitter users in times of mass emergency with a focus on communications that include information which

---

[3] http://techcrunch.com/2011/06/30/twitter-3200-million-tweets/

contributes to situational awareness. I am interested in what information is communicated, how that information is constructed at the linguistic level and how researchers can use this knowledge to contribute to the development of computational tools that can help interested parties quickly identify timely, accurate information broadcast via Twitter in mass emergency situations.

These inquiries lead to the following research questions:

1. What types of information do people communicate that contributes to situational awareness via Twitter during mass emergency? Are there differences in the types of information Twitter users communicate across different disaster situations?

2. What generalizations in verb use exist in tweets that communicate situational awareness information?

3.2 Research Design

In this section, I outline the methods and processes I developed and implemented to answer my research questions. The analytic procedure detailed in this section is influenced by sociologists of disaster, who tend toward inductive, contextual analyses of data (Phillips, 2002; Quarantelli, 2002). I combine these methods with those of sociolinguists and computational linguists by using discourse analysis and corpus analysis in interpreting and analyzing data.

3.2.1 Mass Emergency Event Selection

I focus on four different mass emergency events in this dissertation. The first step in this research was to select the events to study, which involved a combination of previous familiarity with some mass emergency situations, as well as an interest in broadening the scope of the disasters examined by including a categorically different event with the initial datasets. Three of the events examined in this dissertation are

local, or *focalized*, in nature: the 2009 and 2010 Red River floods and the 2009

Oklahoma grass fires. The fourth disaster, the 2010 Haiti earthquake, is a *diffused*

disaster.

> According to Dynes (1970, p. 54), a disaster is *focalized* when it:

> may affect certain population groups within the community and/or result in the demolition of certain buildings within a delimited area…community coordination remains relatively unaffected, and the impact area provides a locus for activity.

Focalized disasters are contrasted with *diffused* disasters, which Dynes (ibid.)

characterizes as:

> single or multiple impacts which result in greater disruption of the community…existing patterns of community coordination are disrupted, the delay in developing new patterns of adaption and coordination creates great delays in rescue and rehabilitation.

> I treat both Red River floods and the Oklahoma grass fire as focalized. The

Red River floods are delimited by nature of the river as the center, or focal point, of

the mass emergency events. Though the river extends for over 500 miles and spans

two countries, area residents and others concerned about flooding in the region

acknowledge that the flooding situation revolves around the level and status of the

Red River and its tributaries—activity regarding the flood centers on the river. In

addition, when the Red River floods, it is an expected event that does not disrupt

communication infrastructure. People in the affected area have access to mainstream

media reports, cellular devices and the Internet. Formal response agencies and

members of the public are able to maintain contact with one another, and organization

of flood preparation and relief efforts persists throughout the flooding event.

Similarly, the 2009 Oklahoma grass fires are focalized in nature. Though the fires were unexpected, burned large swaths of land and area residents were concerned about their spread, formal response agencies were able to delimit the areas under threat by placing a specific number of counties under a state of emergency and focusing on those cities and counties that suffered the most damage and injury. In addition, like the Red River floods, communication infrastructure remained in tact during the fires, and area residents were able to stay in touch with other residents, as well as gather information from area fire departments and other response agencies.

In contrast to the Red River floods and the Oklahoma grassfires, the 2010 Haiti earthquake was a diffused disaster. To say a community was disrupted by the earthquake does not fully describe the catastrophic impact the earthquake had on the entire country. Existing patterns of community coordination were disrupted and rescue efforts required involvement from myriad sources, both domestic and international. Though the earthquake had an epicenter, its repercussions spread throughout an entire nation, and rehabilitation efforts are ongoing.

3.2.2 Emergency Events

The following sections provide background and a description of each mass emergency event examined in this dissertation. An "on the ground" perspective of the 2009 Oklahoma fires, Red River floods from 2009 and 2010 and the 2010 Haiti earthquake are provided.

3.2.2.1 2009 Oklahoma Fires

On the morning of April 9, 2009, an unknown person or group of people intentionally set a fire in Midwest City, Oklahoma[4]. A combination of low humidity and high winds fueled the fire, which eventually spread throughout central and southern Oklahoma (National Oceanic, 2009a). By late afternoon, many fires were burning in the region due to the prevalence of dry brush, grass and trees that further fed the flames (newsOK.com, 2009). At 3:22 pm, the National Weather Service (NWS) in Norman, Oklahoma began to issue fire warnings. According to the National Oceanic and Atmospheric Administration (NOAA), the 2009 Oklahoma fires were "Unlike any wildfire event seen before….they required Fire Warnings in rapid succession" (National Oceanic, 2009a). The fire warnings included details of areas under threat referenced by street names and provided information on escape routes. The NWS office in Norman consistently monitored weather conditions which provided much-needed information on wind speed and direction, location of the dryline[5], and humidity levels.

The fires put over 2,000 homes at risk and ultimately destroyed almost 270 buildings, 228 of which were homes; over 100,000 acres were burned. Fire officials requested voluntary evacuations and at least sixty people were reported injured.

---

[4] The exact cause of the 2009 Oklahoma fire is not known. Fire marshal Jerry Lojka stated the fire that started the series of blazes was intentionally set in Midwest City, Oklahoma in a populated area (at SE 29th and Post Rd.), but was not deemed to be arson due to lack of proof that the person or people who started the fire did so with malicious intent. http://newsok.com/how-and-why-the-midwest-citychoctaw-fire-started-is-unknown/multimedia/video/19562250001#ixzz1QJy7VjXu

[5] The dryline is what separates moist air from the Gulf of Mexico (to the east) and dry desert air from the south-western states in the United States (to the west). http://www.crh.noaa.gov/glossary.php?word=dry%20line

(McNutt, 2009). Officials declared thirty-one counties a state of emergency and eight counties experienced damage. Total damage to the area was estimated at over $30 million based on claims filed with insurers (Oklahoma Wildfire Damage, 2009).

3.2.2.2 Red River Valley

The Red River is situated on the border of North Dakota and Minnesota in the United States, and runs north into Manitoba, Canada where it empties into Lake Winnipeg. Along its course are several urban areas with populations ranging from 7,000-100,000, including Fargo, ND, Moorhead, MN, Grand Forks, ND, East Grand Forks, MN, Whapeton, ND, Brekenridge, MN and Winnipeg, Manitoba as well as many small towns. The Red River is 550 miles long; 395 of those miles lie within the US border (Federal Emergency, March 17, 2009).

Several characteristics of the Red River valley make it an ideal setting for regular floods. As the spring thaw takes place every year, snowmelt begins in the south and moves north. The river is flowing north, and when the thaw happens, the ground becomes saturated with melted snow, and the river experiences runoff, or excess water which causes the river to rise and flow from the river and onto the surrounding land. As the region continues to thaw, runoff from the southern portions meets with fresh snow meltoff from the northern portions, causing floods. Additionally, as ice melts and begins to move north, chunks of ice from the south may meet ice from the north that create "ice jams" which dam the flow of the water. Finally, the Red River valley is—geologically speaking—very young. It is situated in

one of the flattest expanses of land in the world, is only a few thousand years old, and therefore has yet to carve out a substantial valley and floodplain[6] (Schwert, 2003).

The location of the Red River makes the surrounding area agriculturally inviting with its rich and fertile soil. The larger cities in the valley were developed in the 1800s at the point where railroad lines running east-west met with the north-south river route, and those cities and towns have continued to thrive to the present day. Though the Red River valley is attractive to human inhabitants, the area cities and towns are seen as "mistakes" from a geological viewpoint due to the vulnerable, often dangerous situations area residents are placed in (Schwert, 2003). Area residents are frequently on flood alert and/or fall victim to Red River flooding during the spring thaw.

3.2.2.2.1 2009 Red River Flood

On February 5, 2009 the National Oceanic and Atmospheric Administration (NOAA) alerted residents of the Red River Valley that there would be a 50-75% chance for significant flooding in the spring. NOAA cited several factors leading to the projection, including precipitation between 200-300% above normal levels beginning in September 2008; the region experienced twenty-three days of snowfall in December 2008 alone. The large amounts of precipitation resulted in 170-300% higher than normal water content levels (National Oceanic, 2009b).

On March 6, 2009 the City of Fargo issued an emergency declaration in response to the projected floods, and five days later on March 11, the city of Grand Forks followed suit (Federal Emergency, March 18, 2009). On March 13, 2009 the

---

[6] Most rivers in the United States are millions to tens of millions years old (Schwert, 2003).

National Weather Service (NWS) issued a Spring Flood Outlook stating that record flooding would be possible in the Red River Valley and that major flooding was expected for nearly all of the Red River Basin (Federal Emergency, March 15, 2009). Above-average precipitation and below-normal temperatures were predicted, further exacerbating the potential for floods. The governor of North Dakota took action on March 13 and issued an executive order declaring a flood emergency, activated state resources and requested help from the US Army Corps of Engineers (Federal Emergency, March 15 & 18, 2009). On March 18, formal response agencies began to actively monitor river levels, and the state Emergency Operations Center (EOC) was activated.

FEMA first reported that flooding had begun on March 21. North Dakota and Minnesota were already on flood watch and flood preparation activities such as levee construction were underway. On March 23 it was predicted that the area would experience heavy precipitation which would fall onto frozen ground and likely cause further flooding (Federal Emergency, March 23, 2009). The following day, several communities along the Red River ordered evacuations, and shelters opened between 70-300 miles away from the most flood-prone cities and towns (Federal Emergency, March 24, 2009). On the evening of March 26, all area hospitals save one were evacuated. On March 28, the Red River crested[7] in Fargo at a record height of 40.84 feet (North Dakota State, 2011). The passenger train service Amtrak suspended travel, many throughout the region were without power and water boil orders were issued. (Federal Emergency, March 28, 29, 2009). On April 5, the threat of flood had

---

[7] The crest of a river is "the highest stage or level of a flood wave as it passes a particular point": http://news.blogs.cnn.com/2011/05/10/what-it-means-when-a-river-crests/

passed, all shelters were closed, and evacuation orders were lifted (Federal

Emergency, April 5, 2009). The extent of damage to the region was estimated to be

$2 billion (National Oceanic, 2009c).

3.2.2.2.2 2010 Red River Flood

On February 23, 2010 the National Weather Service (NWS) issued a forecast

predicting significant spring flooding in the Red River Valley (National Oceanic,

2010A). Less than a month later, on March 15, FEMA issued a report stating that the

Red River Basin would likely experience flooding similar to that of 2009 (Federal

Emergency, March 15, 2010). The snowpack in 2010 was greater than that of 2009

and cold temperatures had limited snowmelt. In addition, precipitation levels in

December were four times above average, and the ground was frozen to a depth of

three feet (National Oceanic, 2010b). That same day, the North Dakota National

Guard was activated to begin levee construction and traffic control in the area.

Members of the surrounding communities also came together to help with flood

mitigation efforts (Federal Emergency, March 19, 2010). On March 19, the North

Dakota state Emergency Operations Center (EOC) was activated, and the city of

Grand Forks, ND declared a state of emergency.

The Red River crested in Fargo at 36.99 feet on March 21, below the record

level of 2009 (North Dakota State, 2011). However, area residents were on alert for

several days, and members of the public as well as formal response agencies took part

in preparation efforts such as sandbagging and levee building. The 2010 floods did

cause damage in the area, primarily to roads and bridges. Damage in North Dakota

totaled $6.2 million (Federal Emergency Management-DR-1907, 2010) and in

Minnesota damage was estimated at $13 million (Federal Emergency Management-DR-1900, 2010). The floods of 2010 were not as destructive as those of 2009, yet they were a significant event for the residents of the Red River Valley, who had also experienced disastrous impact from the previous year's floods.

3.2.2.3 2010 Haiti Earthquake

On January 12, 2010 a magnitude 7.0 earthquake struck the island nation of Haiti at 16:53 local time; the epicenter of the earthquake was ten miles from the densely populated Haitian capital, Port-au-Prince (United States Agency, 2010; Unites States Geological, 2010). This catastrophic event caused approximately 222,000 deaths and over 300,000 injuries. 1.3 million people were displaced, 97,294 houses were destroyed and 188,383 homes were damaged in both Port-au-Prince and surrounding areas, as well as in southern Haiti (Pan American Health, 2011; Unites States Geological, 2010). Public infrastructure such as telephone lines, electricity and transportation all ceased to operate (Romero and Lacey, 2010); the total cost of the destruction is estimated at between $8 and $14 billion (New York Times, 2011).

The Haiti earthquake of 2010 spawned a large international response. Organizations from various countries led search and rescue (SAR) operations, and provided resources to those in need (Cable News, 2010). The United Nations took control of relief efforts on January 22, 2010 and one day later, the Haitian government officially called off the search and rescue operation on January 23 (Conflict, Humanitarian and Security, 2010). However, recovery operations led by foreign governments and outside organizations continued into February (Federal Emergency February 5, 2010).

In July 2011—one and a half years after the earthquake—it was estimated that 634,000 people remained living in displacement camps and the construction of roads, buildings and other basic structures had barely begun (New York Times, 2011). In addition, many Haitians suffer from disease (particularly cholera) caused by lack of access to clean water; as well as malnutrition, especially among infants and children. (Roberts, 2011; Steinlechner, 2011). The Haiti earthquake of 2010 incited an ongoing disaster that continues to adversely affect the Haitian population; the effects of this catastrophe will continue to be felt for years, if not decades, to come.

3.2.3 Data Sampling and Datasets

This section provides a description of the datasets for each of the four disaster events. The following subsections describe sampling methods and the resulting datasets that were analyzed by two external annotators and me.

3.2.3.1 Focalized Disasters

The 2009 and 2010 Red River floods and the 2009 Oklahoma Fires datasets were sampled using the methods described in the following sections because they represent Twitter user behavior during focalized mass emergency events. The floods and fire both had limited duration and an anticipated end; affected populations had awareness of the disaster agents and were cognizant of the threat. The total impact area for each of these disasters had boundaries indicating where the hazard could impact area communities and properties. Though these boundaries were by no means precise, in both the Red River floods and the Oklahoma fires, response agencies remained at the local, regional and state levels, with the exception of Canadian

agencies being involved in the Red River events. They differ in that residents of the Red River Valley had substantial forewarning of the floods, and were equipped to prepare, while residents of Oklahoma had varied experiences with warning. Some Oklahomans had warning, and were aware of the possibility of evacuation and property loss, while others had no warning and had to quickly vacate their homes or other properties.

Sampling decisions reflect the focalized nature of each of these events. Those who were directly impacted by the events broadcast situational awareness information.

3.2.3.1.1 2009 Red River Flood Data

The data collection period for the 2009 Red River flood took place from March 8 – April 27. March 8 is when residents of the Red River Valley began to operate under flood threat conditions, and by April 27 threat of flood had passed. My colleagues in Project EPIC collected tweets during these periods using search terms that were determined based on an initial qualitative examination of the public Twitter stream. This examination involves a human analysis whereby my colleagues read tweets that mention the events, and determine which search terms and hashtags appear most frequently. For Red River 2009 (abbreviated as "RR09"), my colleagues determined these terms for research that would produce a representative sample of data: `red river` and `redriver`.

This initial search took place using the Twitter search API, and yielded 13,153 tweets. My colleagues then identified how many unique users broadcast those tweets, which resulted in 4983 Twitterers. The next step was to collect the entire tweet stream

(or "contextual stream") for each unique user in these datasets, which returned 4,592,466 tweets. The purpose for collecting entire tweet streams was to include those tweets sent during the flood, but which did not contain one of the identified search terms.

My colleagues then set a threshold, and narrowed the dataset to only those users' tweet streams that include three or more occurrences of a search term. This sampling decision was made to concentrate efforts on Twitterers who were actively tweeting about the events during the data collection period, not those who tweeted about either event only once or twice. This provided a unique user set of 358 Twitterers.

The next decision was to narrow the dataset to those who were *local individual* users. The assumption is that situational information was most likely to come from area residents because each event was focalized; those who live in the affected areas would have the most up-to-date and relevant information. For the 2009 Red River floods, there was a federal response that included North Dakota, South Dakota and Minnesota (Federal Emergency, March 30, 2009), as well as a Canadian response ("New ice jam," 2009). Though the Red River floods spawned an international response in that they required action on behalf of both American and Canadian governments, they were still focalized events specific to a particular region surrounding the Red River.

The decision to threshold the datasets to local individuals resulted in a dataset of 19,162 tweets with 49 unique, local individual users. For the RR09 dataset, the label *geographically local* refers to users who were less than a 6-hour drive from

flood-prone areas. *Individual* means those who post to Twitter as themselves, and not

as representatives of an organization or group (Starbird et al., 2010).

3.2.3.1.2 2009 Oklahoma Fires Data

Data collection and sampling for the Oklahoma grassfires (abbreviated as

"OK09") was similar to the process used for the 2009 Red River floods. The

collection period was from April 8 – 13, which is one day before the fires began, and

three days after they were extinguished. For the Oklahoma fires, my colleagues

identified these search terms: `okfire`, `oklahoma`, `grass fire` and `grassfire`.

Again, this initial search took place using the Twitter search API, and yielded 6,674

tweets; identification of unique users resulted in a set of 3852 Twitterers. Once the

contextual stream for each unique user was collected, the result was a dataset of

1,986,091 tweets. The same rationale was used to limit the dataset to only those users

whose tweet streams included three or more occurrences of a search term, and we

arrived at a unique user set of 221 Twitterers.

Like the RR09 dataset, the OK09 dataset was narrowed to those users who

were local and individual. During the Oklahoma fires, though both Oklahoma and

Texas state emergency operations centers were activated, and Federal Emergency

Management Agency (FEMA) officials sent out situation reports, no federal or

international response was needed (FEMA, April 11, 2009). The decision to threshold

the datasets to local individuals resulted in a dataset of 2,779 tweets and 46 unique

local individual unique users. In this dataset, *geographically local* applies to users

who live in Oklahoma, and whose Twitter profiles either list them as residents of an

area that had a fire, or who indicated via Twitter communications that they were

53

experiencing a fire or under threat of a fire (i.e. seeing flames, smelling smoke, watering yards). *Individual* implies the same definition used in the 2009 Red River flood dataset.

3.2.3.1.3 2010 Red River Flood Data

For the 2010 Red River flood dataset (abbreviated "RR10"), I took a different approach. My software engineer colleagues collected Twitter data related to the 2010 Red River floods from March 15 – April 3, 2010 using the following search terms: `fmflood`, `flood10`, `red river`, `redriver`, `ccflood`, and `fargoflood`. They then determined that 370 unique users tweeted one of the search terms during the data collection period.

Sampling these data involved first determining which of the 370 users who tweeted a search term during the data collection period tweeted more than three search terms, to ascertain which Twitter users were actively tweeting about the event. This produced 101 Twitter users. My colleague Mossaab Bagdouri assembled a dataset of those 101 users' tweets spanning the data collection date range. The next step was to visit Twitter profiles to determine who the users were and perform a cursory analysis of the information contained in their tweet streams. At this point in the analysis, the focus was not on ascertaining only those users who were individuals and local to the flood event—rather, the focus was to determine which users were tweeting information that could potentially contribute to a state of situational awareness for those affected by the flooding. This step in the analysis resulted in the identification of one hundred and one unique Twitter accounts—sixty-nine local individuals, ten local organizations/educational institutions, seven local media outlets,

seven local businesses, three individuals with an unknown location, two non-local

individuals, one non-local media outlet, one federal organization and one

international organization. These one hundred and one users tweeted 11,879 tweets

during the data collection period.

3.2.3.2 Diffused Disaster

The 2010 Haiti Earthquake dataset was sampled using the methods explained

below based on its status as a diffused disaster. Tweets sent in the aftermath of the

2010 Haiti earthquake tell a different story that the previous three datasets. The severe

impact of the earthquake and the subsequent loss of life and property as well as the

many injured prompted a different sort of response on Twitter. People from around

the world tweeted about the Haiti earthquake, with many tweets offering condolences

and sympathy, expressing shock and disbelief and offering prayers to Haitians and

their loved ones. While this type of message is understandably a part of post-impact

behavior, particularly when dealing with an event as devastating as the Haiti

earthquake, this type of tweet is not the focus of this dissertation. With so many

tweets sent in the aftermath of the Haiti earthquake, and knowing that in general most

people who tweeted about it sent only one or a few tweets regarding the earthquake

(Sarcevic et al., 2012) a different approach to choosing the sample to annotate for

purposes of situational awareness is necessary.

3.2.3.2.1 2010 Haiti Earthquake Data

For the 2010 Haiti earthquake dataset, I begin with approximately 4 million

tweets sent between January 11 and February 2, 2010. This dataset was collected by

my software engineering colleagues Professor Ken Anderson, Aaron Schram and Matt Novinger using the Twitter search API by searching for the following search terms determined during an initial examination of the public Twitter stream performed by my EPIC colleagues and me: `haiti`, `earthquake`, `quake`, `shaking`, `tsunami`, `ouest`, `Port-au-Prince`, `tremblement` and `tremblement de terre`. This search resulted in approximately 4 million tweets that contained one or more of the search terms, sent by hundreds of thousands of users from various countries.

My colleague Sudha Verma then helped to narrow the Haiti dataset to a more manageable size; the narrowing process involved several steps. Verma first filtered out all tweets from the original set of 4 million that included any of the terms or strings of characters listed below. These terms and strings were selected based on preliminary analysis of the Haiti dataset. Annotators[8] and I had previously analyzed about 10,000 tweets from the Haiti dataset, and found that many did not include information relevant to situational awareness. Those tweets that did not contribute to situational awareness often included one or more of the words or strings listed below in Table 3.1.

The list of non-case-sensitive terms and strings used to filter Haiti dataset tweets follows (profane words are spelled with "*" to replace one letter):

---

[8] The annotators referred to throughout this dissertation are Jenette Preciado and Sarah Vollmann — both work for Project EPIC and have a background in linguistics.

1	*Pray (and any word that includes the string "pray," e.g. prayer, praying.)*
2	*Donat (any word that includes the string "donat," e.g. donate, donation.)*
3	*RT (an abbreviation for "retweet," a convention in Twitter whereby Twitter users re-post previously posted tweets.)*
4	*Ass*
5	*Bless*
6	*Bullsh*t*
7	*Charity*
8	*Fu*k*
9	*God*
10	*Hell*
11	*Hope for Haiti*
12	*Jerk*
13	*Money*
14	*Sh*t*
15	*Telethon*
16	*Wtf*

Table 3.1: List of Terms in Tweets Removed from the Haiti Dataset

In addition, initial findings indicate that a considerable portion of the Haiti dataset is comprised of retweets, and/or duplicate tweets. So Verma also removed any duplicates. This left us with 1,767,451 tweets. The next step was for Verma to use the natural language processing classifier which was previously developed to locate tweets that contribute to situational awareness based on the style, register and tone of the tweet text (Verma et al., 2011). The classifier is trained on Twitter data from the same four different disaster events I examine in this dissertation: the 2009 and 2010 Red River floods, the 2009 Oklahoma Fires, and the 2010 Haiti earthquake. Verma ran the situational awareness classifier on a random sample of 300,000 tweets taken from the ~1.7 million tweet filtered dataset; 300,000 tweets (approximately 17% of the ~1.7 million) were selected because it is an amount the classifier was able to process in a reasonable period of time. Each tweet was classified as either containing information that contributes to situational awareness, or *not* containing information

that contributes to situational awareness. This classification process resulted in a dataset of 230,409 tweets automatically classified as containing situational awareness information.

Professor Martha Palmer and I determined that a suitable number of tweets for annotators to code at the most detailed levels of analysis would be approximately 1,000, which is the average number of tweets between all three previously sampled datasets. To obtain the dataset, I randomly selected sets of 1,000 tweets from the ~230,000 tweet dataset. Annotators annotated each tweet in these 1,000 tweet subsets as containing information that contributes to situational awareness or not.[9] After the first 1,000 tweet subset was annotated, results showed that 233, or approximately 23% of tweets contained information relevant to situational awareness. Using this ratio as a benchmark, I randomly selected an additional 3,005 tweets from the ~230,000 Haiti dataset to comprise the final Haiti dataset of 4,005 tweets to be hand-annotated.

3.2.3.3 Dataset Overview

A chart detailing Twitter users, number of tweets at various stages of sampling and date ranted for each of the datasets is presented in Table 3.2 below.

---

[9] This annotation process is explained in more detail in Chapter 5.

	2009 Red River Flood	**2009 Oklahoma Fires**	**2010 Red River Flood**	**2010 Haiti Earthquake**
Twitter Users	Local individual users	Local individual users	Geographically dispersed individuals, news media representatives, organizations	Geographically dispersed individuals, news media representatives, organizations
Number of Unique Users	49 geographically local individuals	46 geographically local individuals	101 geographically dispersed users	Not Applicable
Tweets in sample coded by annotators	19,152 contextual stream tweets	2,766 contextual stream tweets	11,879 contextual stream tweets	4,005 tweets randomly selected from a 230,409 tweet dataset
Dataset Selection Date Range	Mar. 8 – Apr. 27, 2009 (51 days)	Apr. 8 – Apr. 13, 2009 (6 days)	Mar. 15 – Apr. 3, 2010 (20 days)	Jan. 11 – Feb. 2, 2010 (23 days)

Table 3.2: Dataset Features

Answering macro-level questions with micro-level research methods is an endeavor that involves careful, thoughtful data sampling. The motivation behind the sampling techniques presented above reflects the attributes of each disaster event; focalized and diffused disaster events are sampled differently based on their characteristics and background knowledge of how each event unfolded on the ground. A discussion of the discourse analysis and three phases of annotation performed on each of these datasets is presented in the following chapters.

3.2.4 Discourse Analysis of Tweets

The phase of research that focuses on identifying situational awareness information addresses my first research question. To identify the specific types of information that Twitter users communicate in mass emergency situations, I perform discourse analysis on tweets to reveal the meaning of tweet content. In chapter 4, I

detail each information type that emerged from the discourse analytic process, provide an example tweets from each dataset, and explain tweet content from both a linguistic and behavioral perspective, taking into consideration mass emergency context and background knowledge.

3.2.5 Coding for Situational Awareness Content

An additional aspect to my first research question involves a qualitative coding process that includes three phases. The annotators and I read and closely examine tweets, and assign qualitative codes that describe tweet content at three levels of granularity, which are described below.

3.2.5.1 Pass One: Determining if Tweets Contain Situational Awareness Information

The purpose of the first pass of qualitative coding is for annotators to read every tweet in each dataset and determine if tweets contain information that contributes to situational awareness. Annotators read tweets, and decide if the tweet contains information relevant to situational awareness based on their knowledge of each emergency event. Tweets may contain no information relevant to situational awareness, in which case they are considered off-topic. Tweets that do mention the emergency event are coded with one of two possible codes—either they contain situational awareness information, or they contain information about the emergency event, but that information is not tactical or actionable in nature, i.e. it does not contribute to situational awareness. The three mutually exclusive codes annotators use to code tweets during this pass of analysis are:

O: Off-topic
R: On-topic and relevant to situational awareness
N: On-topic and not relevant to situational awareness

3.2.5.2 Pass Two: Determining if Tweets Communicate Information about the Social, Built and/or Physical Environments

The purpose of the second pass of qualitative coding is to determine if tweets that contain information relevant to situational awareness include information about the social environment, the built environment and/or the physical environment. The codes used in this pass are not mutually exclusive; annotators may assign one, two or all three codes to a single tweet.

During this phase of research, annotators read all tweets within each dataset that contain information that contributes to situational awareness. The annotators code a tweet for social environment if the tweet includes any mention of human action regarding the emergency event. The annotators code tweets for the built environment if they contain information about building damage, road closures, the status of flights in and out of an airport or similar information. The annotators code tweets for the physical environment if they contain information about weather, the landscape of the affected area, any geological or geographical information or any information about the hazard agent. The codes used in this pass of analysis are:

S: Social Environment
B: Built Environment
P: Physical Environment

3.2.5.3 Pass Three: Determining Specific Types of Information in Tweets

For the third and final pass of qualitative coding, annotators read the same tweets they coded in phase two, and assign up to five specific "information types" identified through the performance of discourse analysis of tweets. My previous

analysis has shown that the most information types found in one tweets is five, which

has since been confirmed across the examination of multiple datasets. Having a fixed

number of potential information types assigned to one tweet is important for the

calculation of inner-annotator agreements, which I present in chapter 5. If a tweet

contains information about the social and physical environments, then the annotators

may assign only those information types that fall under social and physical

environment as listed below. Each information type in this phase describes a specific

aspect of the social, built or physical environment.

Social Environment Information Types:

Advice – Information Space
Animal Management
Caution
Evacuation
Fatality
General Population Information
Injury
Missing
Offer of Help
Preparation
Recovery
Report of Crime
Request for Help
Request for Information
Rescue
Response – Community
Response – Formal
Response – Miscellaneous
Response – Personal
Sheltering
Status – Community/Population
Status – Personal

Built Environment Information Types:

Damage
Status – Infrastructure
Status – Personal Property
Status – Public Property

Physical Environment Information Types:

General Area Information
General Hazard Information
Historical Information
Prediction
Status – Hazard
Weather

3.2.6 Linguistic Analysis Focusing on Verbs

The phase of research focuses on building the foundation for computational linguistic methods that can predict whether tweets contain situational awareness information. During this phase, I examine tweets that include information relevant to situational awareness from all four datasets to determine how that information is linguistically constructed; this phase addresses my second research question. The specific linguistic aspect of tweets I focus on is to identify which verbs Twitterers use to express situational awareness. I focus on verbs because verbs communicate the event taking place in a sentence, and are the central, organizing element; generally speaking, the verb organizes all elements of the sentence that syntactically depend on it (Manning & Schütze, 1999). VerbNet is a lexical resource that provides detailed semantic and syntactic information for English verbs. Within VerbNet, verbs are organized into classes; each verb in a class has a similar meaning, and also displays similar syntactic and semantic behavior. The goal of this phase of research is to identify VerbNet classes with high representation in tweets that contain situational awareness information. These VerbNet classes can then be used as a feature to help machines identify tactical, actionable information in mass emergency situations.

3.3 Summary

In this chapter, I present the research questions which drive this dissertation, the background on each disaster event I examine, the details of data sampling methods and final dataset details and an overview of each phase of my research process. In the following chapter, I discuss the discourse analytic process that led to the identification of specific information types communicated by Twitter users that may lead to situational awareness in mass emergency.

Chapter 4

An Organization of Situational Awareness Information
in Mass Emergency

Part 1 of this chapter presents an analysis of tweets that communicate

information relevant to situational awareness, and how thirty-two specific

"information types" were identified through a ground-up, inductive analysis of

Twitter communications. Part 2 describes additional ways in which the

organizational components of disaster are approached by sociologists of disaster;

by government agencies; and by other researchers who focus on the use of Twitter

in mass emergency. I describe how analyses of these additional approaches

resemble or are distinguished from the results of the analysis presented in part 1.

4.1 Part 1: Discourse Analysis of Tweets that Contain Situational Awareness Information

Understanding the information communicated via Twitter during mass

emergency involves a detailed analysis of tweet content, or the use of micro-level

research methods as a way to uncover and understand macro-level phenomena

(Herring, 2004). This research relies on *discourse analysis* (DA), which Johnstone

(2008, p. xi) treats as "a systematic, rigorous way of suggesting answers to research

questions posed in and across disciplines throughout the humanities and social

sciences and beyond." Johnstone goes on to say that discourse analysts "approach

their tasks by paying close and systematic attention to particular situations and

particular utterances or sets of utterances" (ibid, p. xii). We can think of discourse analysis as "the identification of patterns of structure and meaning in language use" (Herring, 1996).

Many scholars have adapted discourse analysis or one form of it (e.g. conversation analysis) to computer-mediated communication (CMC). Herring (1992; 1995) applies discourse analytic methods to her studies of gender identity in online settings, Panyametheekul and Herring (2003) perform conversation analysis on transcripts of an online Thai chat room to uncover how gender affects turn allocation, and del-Teso-Craviotto (2006) uses conversation analysis to show how people partake in sexual activities in online environments. More recently, Honeycutt and Herring (2009) have applied discourse analytic methods to understand how Twitter can be used as a collaborative tool.

I take a similar approach to those outlined above. I use discourse analysis in this dissertation by reading and analyzing tweets and categorizing them according to the information they convey. This ground-up, empirically-based analytic process uncovers not only the message(s) communicated in tweets, but also the linguistic details of how they are constructed; it provides insight into what information Twitter users deem important to communicate. Overall, the discourse analysis I perform on tweets is a type of *computer mediated discourse analysis*.

4.1.1 Computer Mediated Discourse Analysis

Computer mediated discourse analysis, or CMDA as it is commonly referred, is defined by Herring (2004) as:

> [T]he analysis of logs of verbal interaction (characters, words, utterances, messages, exchanges, threads, archives, etc.). In the broadest sense, any analysis of online behavior that is grounded in empirical, textual observations is computer-mediated discourse analysis.

CMDA is an approach to the analysis of computer-mediated communication that includes methods taken from both discourse analysis and content analysis, but extends each to account for the characteristics of studying CMC. The primary focus of CMDA is "language-focused content analysis" (Herring, 2004), which can be qualitative and/or quantitative in nature. Additionally, CMDA differs from other types of discourse analysis in that it takes the technological affordances of CMC systems into account.

Three assumptions are inherent in performing CMDA (Herring, 2004):

1. Discourse exhibits recurrent patterns
2. Discourse involves speaker choices
3. Computer-mediated discourse may be, but is not inevitably, shaped by the technological features of computer-mediated communication systems.

CMDA applies to language at varying levels of granularity. These levels range from the smallest to the largest linguistic unit of analysis: *structure, meaning, interaction* and *social behavior*. This research resides primarily in the first two levels. *Structure* includes sentence or phrase composition, and *meaning*, which includes the meaning of words and utterances. CMDA provides a research paradigm that is specific to CMC, allows for the incorporation of both qualitative (i.e. interpretations of linguistic phenomena that take place via Twitter), and quantitative aspects whereby instances of particular phenomena are counted, and subsequently interpreted.

The possibilities for CMDA lie in researchers' ability to ask macro-level questions that exist at the group or societal level which can be explored and

subsequently answered by performing detailed, micro-level analysis. As Herring

(2004) notes:

> [T]he power of CMDA is that it enables questions of broad social and
> psychological significance, including notions that would otherwise be
> intractable to empirical analysis, to be investigated with fine-grained empirical
> rigor.

Additionally, the questions researchers ask before they implement CMDA must be

"answerable on the basis of textual evidence" (Herring, 2004). This implies that

researchers must ask questions that can be answered based on text alone, which

serves as a direct indicator of *behavior*, as opposed to asking questions about what

people feel or think, what motivates them, or about the knowledge or understandings

they have of a subject or a situation.

The ability to repeatedly read tweets, think about the context within which

they occur, and come to conclusions regarding the meaning and intent of tweet

authors is an unprecedented research opportunity regarding human activity during

times of disaster. In this chapter I provide insight into the information Twitter users

communicate during the four mass emergencies examined in this dissertation.

In the following sections, I present tweets from each of the mass emergency

events examined in this dissertation. Below each tweet is an analysis of the content as

it was communicated in a particular event. The analysis uses CMDA to explain the

content of each tweet, and how readers may use the content to attain situational

awareness during the time- and safety-critical situations. However, not all tweets

contain accurate information for the disaster situation they describe, and these are

noted when possible. The purpose of the analysis of tweet content is to explicate the

amount and detail of information expressed in Twitter communications, not to verify

or refute that information.

This chapter describes the nuance and detail of Twitter communications about

disaster situations that include information relevant to situational awareness, which

also reveals the inductive, ground-up analytic process that led to the development of

the qualitative coding scheme discussed in chapter 5. Each category of information

presented in this chapter arose from the empirical analysis of tweet content. At least

one example from each dataset is represented in the analysis of each information type.

Exceptions are for those information types not represented in datasets, e.g. no

Sheltering tweets were coded in the RR10 or Haiti datasets.

4.1.1.1 Social Environment

In this section, I provide an analysis of each information type that falls under

the category of *social environment*. Tweets that contain information about the social

environment describe human actions during the mass emergency event. Each tweet is

labeled with the dataset it is from, though not all information types include tweets

from each dataset. Below each tweet, the segment(s) which provides the specific

information type is extracted and explained.

Information Type: Advice – Information Space

This information category labels those tweets that advise readers how to

obtain information from sources outside of Twitter, how to navigate Twitter, and/or

which tell readers where to go to get information regarding the mass emergency event

(e.g. television and radio stations to watch or listen to, or hotline numbers to call).

Tweets labeled with *Advice – Information Space* also provide the reader with

information about online activity taking place that relates to the disaster situation.

1. OK09: `@fitnewbie I am not.  search #OKfires #MWCfires #OKCfires lots of info.`

   Segment: `search #OKfires #MWCfires #OKCfires lots of info`

   In this segment, the tweet author instructs the reader to "search," which refers

   to the Twitter convention of "searching," or locating information based on

   particular terms. In this segment, the reader learns that by searching within

   Twitter for the three listed hashtags, one can find "lots of info," i.e. a lot of

   information, about the Oklahoma grassfires.

2. RR09: `WDAZ says the predicted crest of the Red River is now 52 feet. Follow @egffloodstage to get hourly updates of the river level.`

   Segment: `WDAZ says the predicted crest of the Red River is now 52 feet.`

   In this segment, the reader learns that WDAZ, which is a television station in

   Grand Forks, ND is reporting on river predictions. For those who are

   interested in staying updated on the river level, forecasts and additional

   aspects of the flood situation, they now know that this particular station is

   providing news and updates.

   Segment: `Follow @egffloodstage to get hourly updates of the river level`

   This segment instructs readers to "follow" a particular Twitter account. The

   practice of following is a convention of Twitter users whereby one user

   chooses to add a public Twitter account to his/her list of accounts followed,

and thereby sees all of the tweets of the followed Twitterer. Here, the reader learns of a particular Twitter account that is providing "hourly updates of the river level." If a user follows this account, then regular reports of the river level will show up in the user's tweet stream.

3. RR10: `Lots of new folks joining the North Dakota flood info center today. Join the online community at www.ndfloods.com #flood10 #flood #nd`

   Segment: `Lots of new folks joining the North Dakota flood info center today`

   This segment alerts the reader to the fact that many are joining the "North Dakota flood info center." The phrase "info center" is an ambiguous description; however, given the context of the 2010 Red River flood, and the preceding phrase "North Dakota flood," the reader can surmise that the "info center" is some sort of repository or collection of information regarding flooding situations that most anyone can contribute to or join.

   Segment: `Join the online community at www.ndfloods.com`

   This segment invites the reader to join an online community, and lists the URL for the community. In addition to the previous sentence providing information regarding the purpose of this online community, the URL itself is a clear indication that the community is focused on the North Dakota floods, or "ND floods." Additionally, though the Red River Valley is well known for flooding and residents of the area are on flood watch every spring, other areas of North Dakota also experience flooding. This particular community may not

be focused solely on the Red River floods, but have information about other floods or flood watches in North Dakota.

4. Haiti: `State Dept number for Americans seeking info of family in #Haiti - 888-407-4747 @xirclebox @carriep @hadorable @ndn #twitterhelp`

   Segment: `State Dept number for Americans seeking info of family in #Haiti - 888-407-4747`

This segment instructs the reader on which phone number to call to potentially receive information about family members in Haiti, since post-earthquake, many people were concerned about the safety of loved ones. The segment specifies that the phone number is for Americans—meaning American citizens—who have family in Haiti. The "State Dept" is a shorthand expression for the United States Department of State. Based on the information in this tweet as it was broadcast in the aftermath of the Haiti earthquake, it expresses that the State Department (as it is commonly known) set up an emergency information repository for those in the US searching for information on "family" in Haiti. The specification of "family" implies that close friends, concerned others nor strangers would not be privy to information the State Department may have on those in Haiti. What is not clear in this tweet is if Haitian-Americans who have family in Haiti can retrieve information about their loved ones, or if the State Department information center is for those non-Haitian-Americans who have family in Haiti, i.e. those Americans with loved ones working with aid organizations, church groups or other agencies in Haiti when the earthquake occurred.

Information Type: Animal Management

This information type is used for tweets that include information about pet or other animal care, transportation and/or evacuations due to mass emergency.

Note: No *Animal Management* tweets were coded in the RR10 dataset.

1. OK09: `RT @cityofokc OKC Animal Control is on the scene at 149th and Hiwasee with a response trailer. #okfire`

   Segment: `OKC Animal Control is on the scene at 149th and Hiwasee with a response trailer`

   This segment informs the reader that "OKC Animal Control," or the Oklahoma City Animal Control department, has team members at a particular location with a "response trailer." A response trailer is a large trailer equipped with animal crates, food and water bowls, liter boxes, cleaning supplies and administrative forms. Learning that one of these trailers is set up at a particular location informs the reader of where s/he can take pets in need, or potentially find displaced pets.

2. RR09: `RT @therealjosh: there's an emergency animal shelter set up at the Fargo fairgrounds as well, RT if you can #flood09`

   Segment: `there's an emergency animal shelter set up at the Fargo fairgrounds as well`

   In this segment, the reader is told about an "emergency animal shelter" that has been established at the Fargo fairgrounds. Fairgrounds are frequently the site of county and/or state fairs, which are venues for livestock shows and competitions. They tend to be areas with a lot of open space, as well as the infrastructure necessary to house many animals—including large livestock—which can be difficult to move and house when they are evacuated.

73

An additional point of interest in this segment is the final phrase "as well." This phrase indicates that some other form of sheltering has been set up, or that additional types of aid are available to those affected by the floods; whether this segment refers to another animal shelter, to a shelter for humans, or to a different type of help is not clear. However, it gives the reader an indication that there is more information regarding flood response efforts, for which the reader can search.

3. Haiti: `#News #World Group tackles animal welfare in Haiti after quake http://bit.ly/4FYQT2` (Associated Press)

   Segment: `Group tackles animal welfare in Haiti after quake`

   In this segment, the reader learns that an unspecified "group" is addressing the issue of animal welfare after the earthquake. The lack of a specified group—as well as lack of specific actions taken by the group regarding animal welfare—may cause concerned readers to do further research to find out who exactly is dealing with animal issues in the earthquake aftermath and what they are doing to help the animal population of Haiti.

Information Type: Caution

This information type labels tweets that include information about warnings, watches, declarations and other communications meant to inform readers about threats or potential consequences that may arise due to a hazard. *Caution* tweets may include information about warnings issued by government agencies. These tweets may also include suggestions that area residents take care, or perform certain actions in relation to the disaster, but are not official mandates or orders.

1. OK09: `#okfire  MWC area of Oakwood Drv near Douglass Blvd-Residents asked to avoid SE 29 and SE 15 Street, from Westminster to N Anderson Rd`

   Segment: `Residents asked to avoid SE 29 and SE 15 Street, from Westminster to N Anderson Rd`

Here, the reader learns that residents of a particular area of Midwest City, abbreviated to "MWC," are "asked to avoid" a particular one-mile area defined by bordering streets. The cautionary reference to avoiding an area implies that there is a reason for people to stay away, but the reader does not learn exactly why this particular square mile is to be avoided.

Another point to note in this segment is the use of passive voice, in which the recipient—the element of the sentence addressed in the asking event—acts as the subject of the sentence. In this case, "residents" are the recipient because they are "being asked" to perform a certain action. Additionally, there is no explicit agent, or "initiator" of the action. The reader does not know who is asking residents to avoid this square mile of space, and in using passive voice, the tweet author has removed him/herself from the agentive role. Though it is often possible to find who the agent of an action is by referring to context, this tweet does not provide sufficient background. However, the specificity of the square mile to be avoided, as well as the correct description of the bordering streets, lends the request credibility; it is likely that whoever is asking residents to avoid the area is familiar with the geography of Midwest City.

2. RR09: `RT @mikeoncrime State of emergency declared in East St Paul north of WPG due to major ice jam. No word on evacuations. Snow still falling`

Segment: `State of emergency declared in East St Paul north of WPG due to major ice jam`

This tweet contains much information that contributes to situational awareness, but it is the above segment that conveys information that is cautionary in nature. The reader learns that a "State of emergency" has been declared in a particular area. A state of emergency can only be declared by the government; during a state of emergency, normal functions of the government, and typical behaviors of members of the public are often suspended or changed due to the emergency. This particular state of emergency is in effect because of an ice jam on the Red River in East St. Paul, which is a rural municipality in the greater Winnipeg area of Manitoba, Canada. The reader learns that this particular ice jam is very serious; it may spur evacuations in the area.

3. RR10: `ChastityWalberg: #Fargo city leaders say neighborhoods should be checking dikes every hour to make sure there aren't any problems. #fmflood #Flood10`

Segment: `#Fargo city leaders say neighborhoods should be checking dikes every hour to make sure there aren't any problems.`

Unlike the segment in the previous tweet example, this segment explicitly states who is initiating the action. "Fargo city leaders" are taking a cautionary stance in telling "neighborhoods"—which is a metonym for "the people who live in neighborhoods"—to regularly check dikes to ensure the dikes are holding up, are not leaking and are not compromised in any other way.[10]

---

[10] A metonym is a word or expression used as a substitute for another word or expression with which it is closely associated. For example, "Hollywood" as a substitution for the movie industry, or in the

4. Haiti: `Earthquake experts are warning that the devastating quake that struck Haiti on Tuesday could be the first of several.` http://bit.ly/6Kdhlo

   Segment: `Earthquake experts are warning that the devastating quake that struck Haiti on Tuesday could be the first of several.`

   This is a straightforward warning on behalf of "Earthquake experts" which tells the reader that the devastating earthquake that struck Haiti may be followed by other earthquakes. It is not uncommon for a major earthquake to be succeeded by one or more aftershocks, which are also earthquakes, but are referred to as aftershocks because they occur after the initial, strongest earthquake. This information informs the reader that those in Haiti may experience further destruction, and implies that more response activities may become necessary.

Information Type: Evacuation

This information type describes tweets that indicate an evacuation has been officially ordered by emergency management or government agencies, where evacuation implies the transportation of people from a threatened or impacted area to a safe location. This information type labels tweets that indicate evacuations are taking place, have taken place or that evacuations have yet to be ordered.

1. OK09: `Good news: Evacuations have been ceased and residents are being allowed back in the area. Earlier tweets were hung up in cyberspace I guess`

   Segment: `Evacuations have been ceased`

---

above example, "neighborhoods" as a substitution for people who live in neighborhoods (Hudson, 2000 p. 317).

This segment tells the reader that evacuations have ceased, i.e. evacuations were once in effect, but they are no longer taking place.

Segment: `residents are being allowed back in the area`

The second phrase of this segment "residents are being allowed back in the area" implies permission, or that some authority figure or organization has to permit residents to return to their homes. Overall, this segment informs the reader that evacuations were ordered by some formal response or government agency, but the order to evacuated has been lifted, and additionally, people are permitted back into the evacuated area.

2. RR09: `Retweeting @chriswelle: They have called to evacuate Moorhead south of 94 and West of 8th street. Crest upped to 41.5.`

   Segment: `They have called to evacuate Moorhead south of 94 and West of 8th street.`

In this segment, the reader learns that "they" —which refers to a formal response agency—have "called to evacuate" a particular area of Moorhead, Minnesota. The reader can ascertain that "they" are formal responders or another government organization because of the authority needed to make an evacuation order.

The topic of "they" has arisen in previous crisis informatics research. In their 2008 paper on citizen response to the October 2007 Southern California wildfires, in which the authors conducted in-person interviews with those affected by the fires as well as online surveys, Shklovski et al. (2008) point to the need for residents of an affected area to rely on themselves for

protection, as there was no "they," i.e. there was no higher authority on which residents could rely. One survey respondent wrote:

> What we learned in [an earlier wildfire] is that there is no "they." "They" won't tell us if there is danger, "they" aren't coming to help, and "they" won't correct bad information. We (regular folks) have to do that amongst ourselves.

The reference to "they" is significant, because it is a phrase that is used throughout Twitter communications during mass emergency. It is always possible that there is a specific, anaphoric reference to "they," and that the reader can locate who exactly "they" refers to in a tweet. However, when an anaphoric reference is not readily available, or it simply does not exist, the reader can often surmise that "they" refers to formal response agencies or organizations.

3. RR10: `ChastityWalberg: Cass Electric to get Aquadams today to protect some substations... if you plan to evacuate, make sure to call your utility to shut off power`

Segment: `if you plan to evacuate`

This segment implies that evacuations may take place, but they do not appear to be mandatory, i.e. formal response and/or government agencies have not ordered residents to evacuation. However, the content of this segment indicates that some may be planning to evacuate, which tells the reader that the flood situation is serious enough to warrant this type of behavior on behalf of the affected population.

Segment: `make sure to call your utility to shut off power`

Though this segment does not contain a direct indication of evacuation, it

relates to evacuation behavior. When people choose or are forced to leave

their homes due to impending or predicted flood, there is a possibility their

homes will be inundated with water. A common household injury which

occurs in flooding situations is that of electrocution; electric current can travel

through water and harm anyone around it if the water is also in contact with

electric wires. If residents return to flooded homes that also have working

electric units, electrocution is a threat. Therefore, turning off the main power

supply in a home is advised.

4. Haiti: `CNN's Sanjay Gupta left to care for quake victims as hospital staff ordered out for safety. http://bit.ly/7HZNe4`

Segment: `hospital staff ordered out for safety`

This segment informs the reader that "hospital staff" have been "ordered out

for safety." The use of "order out" in the passive voice indicates that some

external authority has forced the hospital staff to leave a specific building or

area. The reference to "safety" in relation to the ordering out event indicates

that the event is an evacuation put in place for the well being of those who

may be affected by some potentially unsafe situation.

Information Type: Fatality

This information type is used to label tweets that convey information about

people who have died or are suspected to have died due to the mass emergency.

Fatality also labels tweets that provide information about the location and state of the

bodies of the deceased.

Note: Only the tweets in the Haiti dataset were labeled with the *Fatality*

category.

1. Haiti: `Breaking #news #tcot Hundreds feared dead in massive earthquake in Haiti: Port-au-Prince/Wa.. http://twurl.nl/5tnqr1`

   Segment: `Hundreds feared dead`

   This segment informs the reader that "Hundreds," meaning hundreds of

   people, are "feared dead." The reference to "hundreds" indicates that this

   tweet was sent soon after word of the Haiti earthquake was released, as the

   final number of dead was considerably higher. Stating that people are "feared"

   dead is a common occurrence in the Haiti dataset. Using the qualifier "feared"

   in front of "dead" allows Twitter users to hedge, or not commit to a statement

   of how many people have actually perished in the earthquake. During

   disasters, when accurate information is not available but it is known that the

   situation is dire, people use phrases like "feared dead" to express the severity

   of the event, but not commit to a statement of fact.

2. Haiti: `Watching #haiti on cbs and my heart is seriously crushed...I feel like I can't do enough...the bodies the chaos`

   Segment: `the bodies`

   This segment tells the reader that people have died in the Haiti earthquake.

   Read in context, "the bodies" implicates dead, human bodies. In the aftermath

   of this devastating earthquake which the tweet author is watching on television

   or is streaming online, the use of the phrase "the bodies" indicates that corpses

   are publically visible, and available for the media to videotape.

Information Type: Feeding/Hydration

This information type labels tweets that provide information about the need

for or provision of food and/or water to those in need.[11]

Note: Only tweets in the OK09 and Haiti datasets were labeled with the

Feeding/*Hydration* category.

1. OK09: REMINDER: Bethel Christian in Choctaw has food and
   supplies for victims of the #OKFires starting at 9am this
   morning. DM me for more info

   Segment: Bethel Christian in Choctaw has food and supplies for
   victims of the #OKFires

   In this segment, the reader learns that a particular church in Choctaw,

   Oklahoma, has "food and supplies" available for fire victims. This segment

   indicates that the fire left some people without routine access to food, and that

   victims are in need of basic resources.

2. Haiti: the coupon system seems to be working quite fine in the
   food distribution #haiti

   Segment: the coupon system seems to be working quite fine in
   the food distribution

   This segment indicates that food distribution is taking place during the

   aftermath of the Haiti earthquake. Food shortages were serious enough to

   require a coupon system, which meant that earthquake victims were provided

---

[11] This information type was not identified during the linguistic analysis of Twitter data, but was revealed in later analyses of government documentation and the sociology of disaster literature. Subsequent analysis of my datasets revealed that tweets which communicate information about providing food and/or water to those in need were present, and that this is a type of information Twitter users attend to and which contributes to situational awareness. The reference to the nutritional needs of populations affected by disaster in government documents and in academic research, and my identification of this type of information in my datasets, show that *Feeding*/*Hydration* is a category of information to examine in future research on situational awareness information broadcast via Twitter. Further discussion of this analysis is presented in Part 2.

with coupons, or vouchers, that they were able to trade for food and water being distributed by aid and government organizations.

Information Type: General Population Information

This information type labels those tweets that provide information about the population of a region or country. Examples of this type of information include number of people in a particular area, average age, and other basic general demographic information.

Note: Note: Only the tweets in the Haiti dataset were labeled with the *General Population Information* category.

1. Haiti: `Respond: Haiti needs help: poorest nation in our Hemisphere hit by strongest earthquake to hit Caribbean in decades http://bit.ly/4YH7sv`

   Segment: `poorest nation in our Hemisphere`

   In this segment, the reader learns that Haiti is the "poorest nation in our Hemisphere." It is not clear if the reader is referring to the Northern or the Western hemisphere. Regardless of which hemisphere the tweet author means to designate, having the status as the poorest nation in any hemisphere speaks to the dire circumstances Haiti was in before the earthquake struck, and allows the reader of the tweet to surmise what the pre-earthquake circumstances mean with regards to the post-earthquake catastrophe.

2. Haiti: `The average age of Haiti's citizens is 18.5 (That's before the quake) #Haiti #omgfacts`

   Segment: `The average age of Haiti's citizens is 18.5 (That's before the quake)`

This segment informs the reader that the average age of a Haitian person is eighteen and a half. The median age of a Haitian is approximately twenty-one and a half,[12] so it is possible that this information claiming the average, or mean, age is eighteen and a half is accurate. By contrast, the median age of a citizen of the United States is approximately forty. Regardless of whether the tweet author has precise information on the average age of Haitian citizens, he or she is conveying information that the average age of a Haitian is young; approximately half the average age of an American or other citizen of a developed country.

A young average age for a population indicates low life expectancy and/or high fertility rates, which are characteristic qualities of underdeveloped countries.[13] Given that this tweet was broadcast in the context of the Haiti earthquake, and the phrase which follows the age information "That's before the quake," it is likely the author is expressing astonishment and/or concern regarding the potential long-term effects of the earthquake on a population of people who were already experiencing higher than average mortality rates.

Information Type: Immediate Recovery

This information type labels tweets that communicate information about efforts to return to a normal state soon after the disaster event has occurred. Immediate recovery efforts may take place at the individual, community, state, national or international level.

[12] https://www.cia.gov/library/publications/the-world-factbook/geos/ha.html
[13] http://www.un.org/esa/population/publications/worldageing19502050/

1. OK09: `Velma and Meridian homeowners have been allowed to return to their homes after being evacuated to the Stephens County Fairgrounds. #OKFires`

   Segment: `Velma and Meridian homeowners have been allowed to return to their homes`

   Here, the reader learns that residents of two towns in Oklahoma are allowed back into their homes. The residents had been evacuated for an unknown period of time; now, the reader can assume wildfire threat has past and residents are able to begin the process of returning to their normal lives. This tweet segment also implies that residents have homes to return to, i.e. their homes were not destroyed by the wildfires.

2. RR09: `House is mostly all put back together for the first time since the flood.  Now on to other things...`

   Segment: `House is mostly all put back together for the first time since the flood.`

   This segment indicates that the Twitterer was affected by the flood and his/her home was unsettled or disrupted in some way. The phrase "since the flood" implies that the flood has past, and that the tweet author is now able to return to a sense of normalcy by putting his/her home back in order.

3. RR10: `#Fargo announces sandbag removal program. If you'd like to do it, you can apply online: http://ow.ly/1sBy2 (via @inforum)`

   Segment: `#Fargo announces sandbag removal program.`

   In this segment, the reader learns that the city of Fargo is ready for area residents to help remove sandbags, which indicates that danger of flooding has past. Once threat of flood is over, all of the temporary levees and dikes that

have been built around the area must be deconstructed and removed; this recovery process requires volunteers just as the preparatory activities do.

4. Haiti: `St. Louis relief agency rebuilding orphanage in Haiti: ST. LOUIS --- The president of St. Louis-based Internationa... http://bit.ly/b0xgpL`

   Segment: `St. Louis relief agency rebuilding orphanage in Haiti`

   In this segment, the reader is informed that an American relief agency—based in St. Louis, Missouri—will be or is rebuilding an orphanage in Haiti. The reader cannot know if the rebuilding efforts are currently underway, or if they are planned for some future date due to the use of present continuous tense in this segment. In this context, the "rebuilding" may be happening at that moment, or it may be a future activity. Regardless of the timing, rebuilding efforts indicate that recovery efforts are taking place, either tangibly or administratively.

Information Type: Injury

Injury labels tweets that communicate information about persons who have suffered bodily harm due to a hazard. These tweets also include information about lack of injuries, or low numbers of injuries.

Note: No *Injury* tweets were coded in the RR10 dataset.

1. OK09: `The fires are being blamed for 52 injuries throughout the state, said Patrice Greenawalt, trauma director for the Dpt of Health. #OKFires`

   Segment: `The fires are being blamed for 52 injuries throughout the state, said Patrice Greenawalt, trauma director for the Dpt of Health.`

   In this segment, the reader learns that according to the trauma director at the

   Department of Health, 52 injuries have been sustained by people in Oklahoma

86

due to the fires. Whether this number implies that 52 different people suffered from injuries, or an unknown number of people suffered 52 injuries is unknown. However, the reader can safely assume that this information is verified as the tweet author cited his/her source as a spokesperson for a government agency.

2. OK09: `But, as usual, Okies know when to get out of the way. Only one person injured in yesterday's wildfires.`

   Segment: `Only one person injured in yesterday's wildfires.`

   This segment provides contrasting information to that of the tweet in the above example. According to the author of this tweet, there was only one injury due to the fires; there is no mention of 52 injuries. In any disaster situation, precise information is simply not available; no one can know exactly how many injuries were suffered. However, the discrepancy between the two numbers is considerable enough to warrant further investigation on behalf of the reader if he or she is concerned about injury reports. The tweet in example 1 above provides a source, whereas the tweet in this example does not. The reference to a source in example 1 is one indication of trustworthiness. In addition, knowing the context within which these tweets were sent is important; the Oklahoma fires caused damage to over 200 structures and over 100,000 acres; a fire of this magnitude is likely to cause more than one injury. However, it is the responsibility of the reader to ascertain for him/herself if the information in the first or second tweet is correct. The reader can choose to do further research and verify either injury number through additional sources, or the reader can choose to believe either piece of information based on

background knowledge, tweet content and context of the fire situation.

3. RR09: `feeling very inadequate because I put my back out very shortly after starting to fill sandbags--can barely move now.`

   Segment: `I put my back out`

   This segment indicates that the tweet author hurt his/her back during while filling sandbags. The preparatory activity of filling sandbags during a flood threat is physically taxing, and involves bending at the waist, shoveling heavy sand, and placing the sand into burlap or polypropylene bags. If those filling the sandbags are not experienced, or fail to protect their backs by using correct form and method, it is likely injury will occur.

4. Haiti: `Watching 60 minutes on Haiti doctors cutting off a Childs leg with a saw... Jesus help these people`

   Segment: `doctors cutting off a Childs leg with a saw`

   This segment tells the reader about a harrowing event that occurred in the aftermath of the Haiti earthquake. It is estimated that as many as 200,000 Haitians lost a limb in the earthquake (Phillips, 2010). The Haiti earthquake caused great amounts of building damage; those who were inside the damaged buildings as they fell often experienced "crushing injuries," many of which then required amputation (ibid.). The segment above indicates that because doctors had to remove a child's leg, it was likely because the child experienced a crushing injury and required his/her leg to be removed to prevent further injury, infection or death.

Information Type: Medical Attention

This information type labels tweets that provide information about the health and medical situations of those affected by a mass emergency situation. These tweets may include information about giving immediate care; medical teams traveling to the site of the disaster; special attention needed for those with medical conditions; and needs specific to the health and wellness of those in an area affected by mass emergency.[14]

Note: Only tweets in the RR09 and Haiti dataset were labeled with the *Medical Attention* category.

1. RR09: `VALLEY CITY, N.D. (AP) - Voluntary evacuation of people who need medical help or who are elderly or who live in low-lying areas.`

   Segment: `Voluntary evacuation of people who need medical help`

   This segment informs the reader that "people who need medical help" are specifically advised to evacuate from Valley City, North Dakota. This segment indicates that the flood is serious enough that residents of the region are concerned about those who have medical needs, as they are a vulnerable population who require special attention in disaster situations.

---

[14] This information type was not identified during the linguistic analysis of Twitter data, but was revealed in later analyses of the sociology of disaster literature, government documentation and additional research on the use of Twitter in mass emergency. In particular, the research of Sarcevic et al. (2012) provided source material that indicates medical attention is a category of information Twitter users communicate about in mass emergency. Sarcevic et al.'s paper (ibid.) provides many examples of tweets that communicate about medical needs and the coordination of medical care in the aftermath of the Haiti earthquake. The research of Sarcevic et al., the reference to medical needs in the sociology of disaster literature and government documentation, and my subsequent identification of tweets from my datasets that include information about medical attention, reveal that Medical Attention is a category of information that needs to be considered in future research on situational awareness information broadcast via Twitter. Further discussion of this analysis is presented in Part 2.

2. Haiti: `Slow medical care is one more thing for quake victims to survive: Port-au-Prince, Haiti (CNN) -- Nearly a th...` `http://tinyurl.com/ycovllw`

    Segment: `Slow medical care is one more thing for quake victims to survive`

    This segment informs the reader that medical care in Haiti is "slow," which implies that victims are in need of medical attention, but are unable to access or receive it. The implication of this tweet is that slow medical care is also a potential cause for further fatalities, because it is "one more thing for quake victims to survive," meaning that people may die due to the lack of medical attention and care.

3. Haiti: `Medical services in Haiti "overwhelmed" by quake: Geneva, Jan 14 (DPA) A medical aid organisation s..` `http://bit.ly/6bUkz4`

    Segment: `Medical services in Haiti "overwhelmed" by quake`

    This segment tells the reader that "Medical services" are "overwhelmed" in Haiti. The reader does not know if the phrase "Medical services" refers to services that existed in Haiti before the earthquake struck, to services that arrived in response to the earthquake, or to both. However, the reader learns that the current supply of medical services is not enough to meet the needs of those who require care in the aftermath of the earthquake.

Information Type: Missing

This information type labels those tweets that offer information or ask questions about missing persons in the aftermath of a mass emergency. Communication of information about a missing person does not always imply that the

missing person is in a situation of distress, only that concerned others do not know the whereabouts of a person or people.

Note: Only tweets in the Haiti dataset were labeled with the *Missing* category.

1. `Haiti: Heroes' actor looking for parents - Haitian native Jimmy Jean-Louis learned that a house he had grown up in Haiti ... http://ow.ly/16kJ6G`

   Segment: `actor looking for parents`

   In this segment, the reader is informed that an actor who lives in the United States (Jimmy Jean-Louis) is "looking for" his parents. When an individual "looks for" a person or people, it is because the person's whereabouts are unknown. In the aftermath of an earthquake, when people "look for" others, it is because they are missing, i.e. their location is not known.

2. `Haiti: Officials Vow To Find Lynn Students, Others From Hotel Montana in Haiti (source: Big News Network): PORT-AU-PRINCE... http://bit.ly/aJA6eV`

   Segment: `Officials Vow To Find Lynn Students, Others From Hotel Montana`

   In this segment, the reader learns that officials are working to "find" particular people. Like the above tweet, "find" as it is used in this tweet is similar to "look for." This segment indicates that people are missing because they must be found, and unknown "Officials" are working to locate the whereabouts of students and other residents of a specific hotel in Haiti.

3. `Haiti: 9 Aussies safe in Haiti, others still missing http://bit.ly/4vx7l2`

   Segment: `others still missing`

   Here, the reader learns that "others" are "still missing." The first phrase of this tweet indicates that nine "Aussies," or Australian citizens, are accounted for

91

and safe in Haiti. The segment which follows tells the reader that "others,"

which likely implies other Australian citizens, have not been accounted for,

because they are "still" missing. The use of "still," which implies "up to this

point" indicates that the missing status of "others" is not new, i.e. that the

others have been missing for some period of time.

Information Type: Offer of Help

This information type labels tweets that communicate Twitter users' and

others' ability or willingness to help those affected by a mass emergency in any

number of ways which may include provision of supplies or shelter, offers of

transportation or offers of services.

1. OK09: `RT @rhukill: PLEASE RETWEET. If anyone affected by the #okfires needs baby clothes please let @jeditanner know!`

   Segment: `If anyone affected by the #okfires needs baby clothes please let @jeditanner know`

   In this segment, the tweet author is relaying a message regarding the provision

   of clothing for infants. The Oklahoma fires destroyed 228 homes, leaving

   many without basic needs such as clothing and household supplies. This

   tweets informs the reader that a particular Twitter user is either offering baby

   clothes, or has access to baby clothes for those in need.

2. RR09: `Red River Logde offers 25 rate for people evacuated (mandatory) from homes in Fargo Moorhead area and $35 rate for people coming to help out`

   Segment: `Red River Logde offers 25 rate for people evacuated (mandatory) from homes in Fargo Moorhead area`

   In this segment, the reader learns that a hotel in Fargo is offering rooms at a

   discounted rate for residents affected by a mandatory evacuation. By tweeting

   this information, the tweet author is indicating this is a guaranteed price on

behalf of the hotel. From the text of this segment, it appears that hotel

management is charging $25 per room per evening, which is well below the

usual daily charge of more than $50 per room per evening.

Segment: `$35 rate for people coming to help out`

This segment tells the reader similar information to that above—a special,

discounted rate is being offered by a local hotel for people traveling to the

Fargo-Moorhead area to help with flood efforts.

3. RR10: `NDSU Can Help Cope With Flooding: NDSU offers practical, research-based information to help deal with flooding and... http://bit.ly/cGPT3N`

Segment: `NDSU Can Help Cope With Flooding`

Here, the reader learns that "NDSU," which stands for North Dakota State

University, is offering help regarding the flood.

Segment: `NDSU offers practical, research-based information to help deal with flooding... http://bit.ly/cGPT3N`

The reader continues to learn that the help offered by NDSU includes

"research-based information" that affected people can use to deal with

flooding. The segment continues to a link, which the reader can assume

contains the information.

4. Haiti: `UPS will ship up to 50 Pounds Free to Haiti, Pick up a case of something tomorrow and drop it off at any UPS Location, Address to Haiti`

Segment: `UPS will ship up to 50 Pounds Free to Haiti`

This segment informs the reader that UPS, or the United Parcel Service, is

offering their transportation services to ship items to Haiti free of charge. Due

to the often expensive costs associated with international shipping, the offer of

93

free transport of goods to those in need in Haiti might encourage people to donate necessary items.

This tweet is notable because the information it communicates is false. At no time during the earthquake aftermath did UPS officials offer to ship items to Haiti free of charge. However, regardless of truth value, the text of this tweet expresses an offer of help, and it is the duty of the reader to ascertain if the information is accurate.

Information Type: Preparation

This information type labels tweets that communicate information about activities and actions meant to guard against or ready people/property for an impending hazard event. These tweets may convey information about risk reduction and/or mitigation as it relates to the hazard. Tweets in this category mention specific measures taken to reduce the threat of an impending hazard agent, e.g. dike building in a flood situation or wetting property in a fire situation.

Note: No tweets in the Haiti dataset were coded with *Preparation*.

1. OK09: `It is amazing how people are coming together and helping to water neighbors yards...`
   `http://www.koco.com/news/19140975/detail.html`

   Segment: `people are coming together and helping to water neighbors yards`

In this segment, the reader learns about preparatory measures people take when under threat from wildfire. One way residents go about preparing their homes for potential threat of fire is to water the surrounding grass, brush and trees. By dampening any material that may act as fuel for a fire, residents hope to deter flames from starting or spreading. Additionally, this segment tells the

reader that there is a community effort afoot, and people are helping each other by watering neighbors' yards. By preparing all homes in an area, risk of fire spreading is hopefully abated.

2. RR09: `Flood fight leaders have goal to produce 500,000 sandbags "today" - stock currently on hand is being used for leaking dikes.`

   Segment: `Flood fight leaders have goal to produce 500,000 sandbags "today"`

   In this segment, the reader learns of a goal to produce 500,000 sandbags in one day. Much of the preparatory activity during predicted floods involves filling bags made of heavy material (usually polypropylene or burlap) with sand. The bags are then stacked together to create levees, which are walls meant to keep water from overflowing into surrounding areas.

   Segment: `stock currently on hand is being used for leaking dikes.`

   This segment mentions "stock," which refers to the reserve of sandbags on hand. The current stock of sandbags is being used to repair previously constructed dikes which are leaking. This tweet implies that a lot of preparatory work is necessary to fight the flood, as many sandbags are needed, and current flood fighting measures are insecure.

3. RR10: `ChastityWalberg: City of #Fargo planning to build levees to protect up to 40' -- expecting a 38' crest on Saturday...`

   Segment: `City of #Fargo planning to build levees to protect up to 40'`

   Here, the reader learns that Fargo city management plans to build levees that will protect the city up "up to 40'." This means that the levees will be high enough to stop water from overflowing into surrounding areas as long as the

river crests at 40 feet or lower. In 2009, the Red River crested at 40.84 feet, and though the prediction included in this tweet was not as dire, city management was taking every precaution and preparing for a crest two feet higher than the prediction of 38 feet.

Information Type: Report of Crime

This information type labels tweets that include reports or mentions of criminal activity, security precautions in case of criminal activity, or which advise against criminal activity such as theft, traffic violations and other illegal actions.

1. RR09: `Red River at Fargo: 37.37. 3 and a half feet below record crest. Garage Fires & Burglaries reported in flood areas Details on CH 4`

   Segment: `Burglaries reported in flood areas`

   This segment tells the reader that burglaries have been reported in "flood areas," or locations that are flooding or under threat of flood. The reader does not know if any burglaries actually took place, only that the police have been alerted and the crime is reported.

2. RR09: `Cass County Sheriff Paul Laney says people caught cutting into roads to divert water from their yards will face charges.`

   Segment: `Cass County Sheriff Paul Laney says people caught cutting into roads to divert water from their yards will face charges.`

   This segment, which is the entire tweet, informs the reader that the sheriff is warning residents of Cass County—a county in North Dakota situated along the Red River—not to "cut into roads to divert water from their yards." A common tactic used to reroute water surplus is to dig ditches that can collect water and divert it away from the surrounding area. Residents of Cass County

are being told not to partake in this particular practice, and that doing so will

result in criminal charges.

3. RR10: `ChastityWalberg: Sheriff Bergquist: 30+ roads washed out`
    `in Clay County, MN -- don't drive around barricades... you`
    `could face charges....`

    Segment: `don't drive around barricades... you could face`
    `       charges....`

This segment informs readers they should not "drive around barricades." In

flood situations, a common measure that is taken by police to keep drivers off

dangerous roads that are flooded or will likely flood is to create barriers that

vehicles cannot drive around. However, it is not uncommon for some drivers

to ignore the barricades and to circumvent them, which is illegal.

4. Haiti: `BBC news criticizes people in #Haiti are looting food –`
    `what's their alternative ?   - nip to Tesco ?`

    Segment: `people in #Haiti are looting food`

This segment informs the reader that looting, or stealing large amounts of

goods—often by a large number of people—is occurring in the aftermath of

the Haiti earthquake. Accounts of looting in the aftermath of disaster

situations is a contentious issue among sociologists of disaster. Dynes and

Quarantelli (1968) point to various reasons for people to loot when a disaster

or other civil disturbance occurs, and assert that the looting behavior is often

evidence of depravity, i.e. people are simply fulfilling basic needs. The

authors also state that looting reports are often based on misinterpretation of

the situation by formal response agencies and the media. What may be

deemed "looting" is not in fact a criminally-minded act, but a way to simply

survive. In addition, it is not uncommon for goods to be given away in disaster situations, but again, this activity may be called "looting" by those not fully informed (Dynes and Quarantelli, 1968).

The author of this tweet is in effect taking the point of view of Dynes and Quarantelli in questioning the fact that looting behavior is being criticized. The author does this with his/her sarcastic, rhetorical question at the end of the tweet: "What's their alternative? Nip to Tesco?"[15] The tweet author is challenging the idea that others condemn looting by those who are in great need, as the people of Haiti have few, if any, options to meet daily requirements.

However, I include the above tweet, and others that mention looting behavior, in the *Report of Crime* category. Regardless of the frequent misinterpretation of "looting" in disaster situations, looting is a crime. It is the responsibility of the reader of those tweets that mention looting to interpret them for him/herself, and decide if the "looting" is in fact the illegal seizure of goods, or if it is a case of people trying to survive.

Information Type: Request for Help

This information type labels tweets in which Twitterers are either asking for help directly, or relaying messages about others' request for help or aid of some sort. This may include requests for others to create technological tools (e.g. map mashups) or to provide services, among other appeals. Additionally, *Request for Help* tweets

_____

[15] Tesco is a popular chain of grocery and dry goods stores found in the UK and other European countries.

may indirectly ask for help by stating a need.

1. OK09: RT @NeighL: OKLAHOMA CITY - The American Red Cross is seeking donations to help with wildfire relief.

   Segment: The American Red Cross is seeking donations to help with wildfire relief.

This segment informs the reader that the local chapter of the American Red Cross is asking for donations to help victims of the Oklahoma wildfires; the segment works as an implicit request for help by passing along this type of information. The text of this segment does not specify if material donations in the form of clothing, blankets and other goods are requested, if the money is requested, or either/both types of donations.

The request for donations in the 2009 Oklahoma wildfire disaster is considered a request for help. The wildfires were a focalized disaster, meaning they did not garner national nor international attention nor response. Though the fires exacted considerable damage to structures and land and caused many injuries, they did not create worldwide nor nationwide awareness. Since the fires were a regional emergency, and were damaging enough to require the need for donations and other forms of help from local aid organizations, the above tweet is considered a request for help; by asking for donations, the reader is made aware of how serious the fires were. This is not necessarily the case in other mass emergencies e.g. the 2010 Haiti earthquake, which quickly spurned an international response and massive requests for donations. In the case of a massive earthquake that quickly topples entire cities, the need for

financial aid is obvious; in the case of a local wildfire, the need for financial

aid signifies the gravity of the situation.

2. RR09: `RT @inforum: Volunteers still needed at Nemzek in Moorhead and Dilworth. In Dilworth, report to the east end of the old Walmart parking ...`

   Segment: `Volunteers still needed at Nemzek in Moorhead and Dilworth`

   Here, the reader learns that volunteers are "still needed" at "Nemzek," which

   is short for Nemzek Hall, an athletic complex on the campus of Minnesota

   State University in Moorhead, Minnesota. The use of "still" in the first part of

   the segment is telling. By saying volunteers are "still" needed, the tweet

   author indicates that volunteers are needed at the present moment; they have

   also been needed up until the present moment. The reader could surmise that

   the need for volunteers is quite grave given the circumstances of the 2009

   flood, and linguistic construction of this segment. After the mention of

   Moorhead, the tweet author adds "and Dilworth," which indicates that

   volunteers are also needed in another town in Minnesota. Though this segment

   is a statement, its content points to an implicit request for help.

   Segment: `In Dilworth, report to the east end of the old Walmart parking ...`

   This segment refers to the need for volunteers expressed in the previous

   segment. It provides specific directions for where volunteers are to report in

   Dilworth, a town east of Moorhead in the Red River Valley.

3. RR10: `RT @fdipays: Received Code Red call requesting volunteers to sandbag in Moorhead on Tuesday..spread the word! #fmflood #fargoflood`

Segment: `Received Code Red call requesting volunteers to sandbag in Moorhead on Tuesday`

This segment informs the reader that the tweet author received a "Code Red Call" from an unspecified source. The term "code red" implies a very serious or dangerous situation and/or an emergency. The purpose of the call is to request help by directly asking people to volunteer to sandbag in Moorhead, Minnesota.

Segment: `spread the word!`

This segment is also a request for help; the tweet author commands other Twitter users to "spread the word," or notify others that volunteers are needed to sandbag in Moorhead.

4. Haiti: `As more aid arrives, fuel is emerging as a key issue. Aid flows can only operate fully for 2-3 days w/o more fuel in #Haiti`

Segment: `fuel is emerging as a key issue`

This segment is not a request for help, but it does important discursive work in establishing the context within which the following segment is to be interpreted. The tweet author tells the reader that fuel is a "key issue," meaning that fuel is a necessary component to successful aid efforts.

Segment: `Aid flows can only operate fully for 2-3 days w/o more fuel in #Haiti`

This segment informs the reader that "aid flows," which are deliveries of supplies, will suffer if more fuel is not delivered to Haiti. The way aid distribution works in a situation like the Haiti earthquake is for large shipments to arrive at major ports or airports. Smaller vehicles such as

helicopters, small plans and trucks then transport the supplies to outlying areas and deliver them to hospitals, aid centers or area residents.

In the aftermath of the Haiti earthquake, there was a dearth of fuel for small vehicles. Regardless of how many supplies were located at the temporary distribution centers (e.g. airports), if those supplies could not be transported to those in need, they were rendered useless. This segment is making an implicit request for help, i.e. it does not directly ask for fuel to be sent to Haiti. Instead, the tweet author informs the reader of the situation, and explains that fuel is necessary for aid efforts to continue. In doing so, the author makes a request for help in the form of replenishing fuel supplies.

Information Type: Request for Information

This information type labels tweets in which Twitter users ask for updates about the mass emergency situation. These may include requests regarding the wellbeing and safety of others, requests for the status of the hazard, and requests for information about specific property (private or public) areas of land. Overall, this is a category of information that covers a wide range of questions Twitter users ask in relation to the mass emergency event. Many of the tweets in this category are also coded with additional information types which further describe the information being requested.

1. OK09: `@mrscdonaldson how's gma betty's house?`

   Segment: `how's gma betty's house?`

   In this segment, though the tweet author is asking a question directly to another Twitter user, other readers of the tweet can also gain useful, tactical

information. The author asks about the state of a specific home, that of "gma betty," whom the reader can assume is known by both the tweet author and the Twitter user to whom the tweet is directed. The author uses the abbreviated form of "grandma" to refer to a woman the reader can presume is called "Grandma Betty." If the reader knows Grandma Betty, or knows where she lives, the reader can ascertain that Grandma Betty's house is under threat of fire, since the tweet author is asking about the house. Asking about property in the midst of a wildfire indicates that the asker is concerned about a home, vehicle or other possession because those possessions are in potentially dangerous situations due to the present or future state of the fire.

2. RR09: `Anyone else find it weird that the whole city #Winnipeg is getting brown outs? It's not all the same grid right?`

   Segment: `Anyone else find it weird that the whole city #Winnipeg is getting brown outs?`

In this segment, the reader learns that Winnipeg, Manitoba is experiencing brownouts, which are drops in electric voltage that cause lights to dim and may have other effects on electric devices and tools. Brownouts may be intentional, meaning that electricity suppliers purposely drop voltage to reduce the load placed on a particular electricity generation system. Brownouts may also happen unexpectedly, due to overuse by consumers, a mechanical failure or other disruption to the electric system. The fact that brownouts are occurring is significant; though the reader does not know the source of the brownouts, knowing they are taking place provides insight into the infrastructural situation in Winnipeg. In addition, the Twitterer makes a

request to other Twitter users, and wants to know if others think it odd that an

entire city of over 600,000 residents is experiencing brownouts.

Segment: `It's not all the same grid right?`

In this segment, the reader uses a *tag question* to ask about the city of

Winnipeg's power grid system. A tag question is a statement which includes

an auxiliary component, or tag, at the end. The tag acts as an orienting feature

in that it conveys a positive or negative position on behalf of the speaker. Tag

questions are often used when the speaker presupposes an answer, i.e. the

speaker has an expectation of what the answer to his/her question is or should

be. In the above segment, the tweet author uses the tag "right?" at the end of

the preceding negative statement, indicating that the author does not believe

that the entire city of Winnipeg is on the same power grid. In addition, the

author writes "It's not all on the same grid." The reader can make the

anaphoric reference to the previous sentence in the tweet, and know that "it"

refers to all of Winnipeg. After this segment in addition to the previous one,

the reader knows that all of Winnipeg is experiencing brownouts, yet the

tweet author does not think the entire city is on the same power grid, which is

confusing and potentially worrisome.

3. RR10: `hmmm. anyone else think the tap water has been tasting funny lately? #fargoflood`

Segment: `hmmm. anyone else think the tap water has been tasting funny lately?`

The author of this tweet begins this segment with a *contextualization cue* in

the form of "hmmm." The term contextualization cue "refers to any aspect of

the surface form of utterances which, when mapped onto message content, can

be shown to be functional in the signaling of interpretive frames" (Gumperz, 1977). Given the context of the 2010 Red River flood, and the sentence which follows "hmmm," the use of a contextualization cue in this instance indicates that the tweet author is pondering the connection between the funny-tasting tap water, and the ongoing flood situation. Overall, this segment is a request posed to others in the Red River Valley area; it is also a suggestion that the flood situation is potentially serious enough to affect the water filtration system.

4. `Haiti: why can't we send the pre-mixed formula (with clean water) to Haiti?  Why the call for breast milk?`

   Segment: `why can't we send the pre-mixed formula (with clean water) to Haiti?`

   In this segment, the tweet author is asking why "we" cannot send pre-mixed formula—which is a manufactured food for infants—to those in need of infant nourishment in Haiti. This segment indicates that a request has been made for people to provide food suitable for infants to those in need in Haiti in some form other than manufactured formula. The tweet author is asking the Twitter audience at large why formula made with clean water is not acceptable.

   Segment: `Why the call for breast milk?`

   In this segment, the reader learns that instead of formula, there has been a request for people to send breast milk to Haiti. The tweet author is again asking a question to the Twitter audience at large, wanting to know why breast milk is preferred over manufactured formula for those infants in need of nourishment in Haiti.

105

Information Type: Rescue

This information type labels tweets that provide information about the need for or the performance of search and rescue efforts for people trapped or otherwise incapacitated due to the emergency event.

Note: Only tweets in the Haiti dataset were labeled with *Rescue*.

1. Haiti: `IDF Rescue team in Haiti, rescued a 52 year old man after he smsed themhttp://bit.ly/7zQzaz`

   Segment: `Rescue team in Haiti, rescued a 52 year old man after he smsed them`

   This segment informs the reader that a 52-year-old man was rescued by a particular rescue team, the members of which are likely trained to perform such operations. It also includes mention that the man who needed rescuing sent a text, or SMS message, to the rescue. This segment indicates that the SMS message is what alerted the rescue team to the man's situation, which is when they took action and were able to rescue him.

2. Haiti: `hmm "A lot of Haitians are wondering why so many Americans seem to be getting rescued, and not so many Haitians."-Kate Snow, ABC News #Haiti`

   Segment: `hmm "A lot of Haitians are wondering why so many Americans seem to be getting rescued, and not so many Haitians."`

   Once again, a segment begins with the contextualization cue of "hmm." In this case, the use of "hmm" before a direct quote from a mainstream media news anchor is a form of commentary on behalf of the tweet author. The information that more Americans are being rescued in the aftermath of the Haiti earthquake and "not so many" Haitians are being rescued is perhaps questionable and/or troubling.

Information Types: Response

The following four information types are all used to code tweets that describe some form of response effort in relation to a mass emergency event. Tweets labeled with one of the *Response* categories communicate information about *who* is performing a response activity. The *Response* categories include: *Response – Community, Response – Formal, Response – Miscellaneous* and *Response – Personal*. Many of these tweets are labeled with additional information type categories that further describe the response activity, but not all. For example, this tweet is labeled with only *Response – Community*:

```
RR09: @art4thehomeless We have many groups of volunteers from
all over helping with the flood in Fargo ND right now. :)
```

This tweet provides information that contributes to situational awareness by informing the reader that many people, e.g. "many groups of volunteers from all over" are in Fargo, North Dakota and are helping with flood efforts. No additional information is provided to tell the reader exactly what these volunteers are doing, but the reader can ascertain from the content of this tweet that the flood situation is concerning enough to prompt a response from many people. An explanation of each *Response* category follows.

Information Type: Response – Community

This information type labels tweets that convey information about responses to the threat or impact of a hazard agent that may cause a mass emergency by community members working in groups, such as faith-based organizations or non-profit agencies. *Response – Community* also labels tweets that indicate any group of members of the public are performing tasks related to the mass emergency. These

groups may be referred to by a specific name or in reference to a specific

commonality (e.g. university students) or they may be referred to implicitly, or with

general terms such as "volunteers."

1. OK09: `Bethel Christian will have 400 boxes of food, drinks, and hygiene kits to give to fire victims after 9am tomor. DM me for more info. Plz RT`

   Segment: `Bethel Christian will have 400 boxes of food, drinks, and hygiene kits to give to fire victims after 9am tomor.`

   Here, the reader learns that Bethel Christian, which is a large church in

   Choctaw, Oklahoma, will have basic supplies on hand "after 9am tomor." for

   fire victims. A sizable amount of supplies—400 boxes of food, drinks and

   hygiene items—will be offered to those in need. This amount of necessities

   the church has available speaks to the gravity of the fires and provides the

   reader with an indication of how many people were affected.

2. RR09: `Going to check out a neighborhood where residents have created their own dike patrolling system. I'll be live there at 10!`

   Segment: `a neighborhood where residents have created their own dike patrolling system.`

   In this segment, the reader learns that "residents" of a particular neighborhood

   have banded together to patrol the dikes in the area. During the 2009 Red

   River flood, many areas were in need of protection from dikes that were built

   in preparation for the flood. While diking is an effective way to prevent flood

   damage, it is also possible that the dikes will become damaged by springing

   leaks or allowing water to seep through sections that are not watertight.

   Therefore, members of the community are careful to monitor the integrity of

   the dikes in their neighborhood to make sure they are all functioning as

expected. This is an example of community members responding to disaster needs of their own volition.

3. RR10: `RT @DaronSelvig: Flood fight volunteer turnout is light in Moorhead. Volunteers encouraged to report to Nemzek Hall at MSUM. http://bit. ...`

Segment: `Flood fight volunteer turnout is light in Moorhead.`

In this segment, the reader learns that "volunteer turnout is light in Moorhead," which implies that there are not enough people in the area to help with the flood situation; most likely, with sandbagging. Learning that turnout is "light" indicates that flood fight organizers have a target number of volunteers they would like to perform tasks, i.e. organizers may want to fill a certain number of sandbags, and figure they need a certain number of people to complete that task in a particular time period. Not having enough volunteers may also indicate that area residents are not willing or able to help with flood fight efforts, or that they do not think it is necessary to help. In addition, this segment tells the reader that volunteer numbers are low *in Moorhead*. This may imply that area residents are in fact helping with flood fighting efforts elsewhere, and that Moorhead is being neglected in favor of other locations along the Red River.

Segment: `Volunteers encouraged to report to Nemzek Hall at MSUM.`

In this segment, the reader is told where volunteers, who are often members of the community, are encouraged to "report," or to show up in order to help with the flood fight. Nemzek Hall acted as a volunteer center during the 2010

floods; people would report there, and then be transported to various locations in the area to help construct sandbag dikes wherever they were needed.

4. Haiti: `U.S. Churches Search for Workers, Missionaries in Haiti: As aid workers frantically rescue survivors of Haiti's ma... http://bit.ly/8gTqjB`

   Segment: `U.S. Churches Search for Workers, Missionaries`

   This segment informs the reader that U.S. churches are responding to the Haiti earthquake by searching for workers and missionaries. The churches are likely searching for Americans or other non-Haitian people, since missionaries and church workers are often people who are citizens of one country and travel to foreign countries for any number of reasons.

Information Type: Response – Formal

Response – Formal labels tweets that convey information about response, lack or response or need of response to mass emergencies by local or state emergency responders, governmental agencies, non-governmental organizations, the military and similar groups. This category also includes recommendations and advice from official community, state and national leaders. The agents, or "doers" of response efforts described in these tweets may be referred to explicitly, e.g. "the police." There may also be implicit reference to an external authority that likely indicates a formal response agency.

1. OK09: `Police blocked off SE 149 between Anderson & Westminster Rds b/c of fires. Firefighters working to contain fires by Lake Draper. #OKFires`

   Segment: `Police blocked off SE 149 between Anderson & Westminster Rds`

Here, the reader learns that "Police," the local agency responsible for crime prevention, law enforcement and maintenance of public order have "blocked off," or refused access, to a particular stretch of road. Not only does the reader learn that a certain section of road in Oklahoma is not available for use, he or she also learns that the fire situation is serious and/or dangerous enough to warrant road closures by formal response agencies.

Segment: `Firefighters working to contain fires by Lake Draper.`

In this segment, the reader learns that "Firefighters," which refers to local or outside agencies responsible for extinguishing fires, are "working to contain fires" in a particular location. The use of the present continuous tense—*are working*—indicates that the actions of firefighters are ongoing and happening at the moment the tweet was sent.

2. RR09: `@JanetStewartCBC can they blow it up?  explosives? #redriver #flood09`

Segment: `can they blow it up?  explosives?`

This tweet was presented as an illustrative example in chapter 2. Here, I delve into greater detail regarding its meaning. This segment makes two anaphoric references—*they* and *it*—which the reader can interpret if he or she is aware of the context of the 2009 Red River floods, and has background knowledge about flood prevention techniques. However, if the reader does not have background knowledge, it is still possible to surmise that *they* refers to formal response agencies, the military or a similar organization. As explained earlier, the unmarked use of *they* in tweets sent during mass emergency tend to refer to formal response agencies, unless there is clear anaphoric reference to other

111

people, a group or an organization. In addition, *they* is mentioned in this tweet as the agent in a *blowing up* activity. People who are not members of formal response or military agencies do not usually have access to materials used to blow up or explode things—the tweet author refers to explosives, which indicates how the blowing up activity is expected to be carried out.

The second anaphoric reference in this tweet is to *it*. Those who are familiar with the Red River region and how flooding often unfolds understand that a common cause for flooding is ice jams. One way to cope with ice jams is for formal response agencies or the military to use explosives to break up the ice jams and allow the smaller pieces of ice to flow downstream. Though ice jams are not explicitly mentioned in this tweet, they are what require the use of explosives during a flood. Therefore, the reader can interpret this segment as a question which asks "can formal response agencies/the military blow up the ice jam with explosives?"

3. RR10: `Well, they've closed one of the 5 bridges in town. How many more before we're done?`

    Segment: `they've closed one of the 5 bridges in town`

    In this segment, the reader is once again presented with *they*. The reader learns that "they," meaning members of a formal response agency, likely the local police force, have closed a bridge in an area under flood threat to prevent traffic flow which could potentially lead to accidents and/or injury.

4. Haiti: `http://bit.ly/8PHx8w UN Peace Keepers Shoots At Hungry Mob in Haiti haiti palace`

    Segment: `UN Peace Keepers Shoots At Hungry Mob`

112

In this segment, the reader is presented with information about formal response to the Haiti earthquake by international forces. UN (United Nations) peacekeepers are military and police personnel from 115 different countries who are member states of the United Nations.[16] They serve to protect and provide security for populations affected by disaster, war or other situations of human discord or conflict. UN peacekeepers have had a presence in Haiti since 2004, and in the aftermath of the 2010 earthquake, the UN Security Council made the decision to send more troops to Haiti.[17]

In the above segment, the reader learns that peacekeepers shot at a hungry mob outside of the presidential palace. What the reader does not learn from the above segment is that this information was close to two years old at the time it was broadcast on Twitter in the aftermath of the 2010 earthquake.[18] The information presented in this tweet refers to a situation that took place in 2008, during which food distribution was underway, and UN peacekeepers shot at people hoping to obtain food and other supplies. I present this tweet in the *Response – Formal* analysis to show that information is not always presented in the correct context, and that Twitter users must be vigilant and check sources and perform further research to understand information accuracy during disaster situations. Information broadcast on Twitter during mass emergency can lead to situational awareness, but not all information is timely nor accurate.

[16] http://www.un.org/en/peacekeeping/
[17] http://www.un.org/en/peacekeeping/missions/minustah/
[18] http://news.bbc.co.uk/2/hi/7337792.stm

113

Information Type: Response – Miscellaneous

Response – Miscellaneous labels tweets that communicate about response to a mass emergency situation, but which do not specify who the response is performed by; the response activities communicated in tweets labeled with this information type may be performed by community groups/organizations, formal response agencies, individuals or small groups. These tweets do not contain any information that allows the reader to assess who is responding, but which nonetheless is communicating about response.

Note: Only tweets in the Haiti dataset were coded with *Response – Miscellaneous*.

1. Haiti: Rapichat.com Updates Haiti workers hand out 'women-only' aid coupons (AP) http://ow.ly/16sN9O

   Segment: Haiti workers hand out 'women-only' aid coupons

   In this segment, the reader learns that "workers" are distributing coupons intended only for women during the aftermath of the Haiti earthquake. The vague reference to "Haiti workers" does not inform the reader what type of organization, service group, military operation or other agency the workers are affiliated with, hence the label of *Response – Miscellaneous*.

   Workers handed out women only aid coupons—which are vouchers women use for food and other supplies—because of the unbalanced food distribution that occurred after the earthquake. Frequently, young men would rush to the head of food distribution lines, which were first-come, first-served.

This left women and weak people without food or supplies.[19] The provision of coupons that only women could redeem was an attempt to balance food distribution among all members of the population.

2. Haiti: `Aid pours into Haiti airport as relief workers struggle to distribute it: Los Angeles TimesBush, who was widely c...` `http://bit.ly/5WC9Hw`

   Segment: `Aid pours into Haiti airport as relief workers struggle to distribute it`

In this segment, the first phrase indicates to the reader that aid is arriving in Haiti, though the origins of the aid are not mentioned. The tweet author writes that aid "pours into" Haiti, which indicates there is an abundance, or large quantities, of aid being sent in response to the needs of the Haitian people. The second phrase indicates that unspecified relief workers are struggling to distribute the aid. Overall, this segment informs the reader that response efforts are not straightforward. Though large amounts of aid are getting to Haiti, challenges exist regarding how to distribute it to those in need.

Information Type: Response – Personal

Response – Personal labels tweets that convey information about how individuals or small groups of people without any community or formal affiliation respond to the mass emergency situation.

1. OK09: `Holy cow, #OKFire wildfires are burning down neighborhoods - at least 50 houses, so far - not in Edmond, but I'm watering the yard, anyway.`

   Segment: `but I'm watering the yard, anyway.`

---

[19] http://www.guardian.co.uk/world/feedarticle/8924414

This segment informs the reader that the tweet author is watering his/her yard as a response to fire threat. The segment begins with "but," which indicates that the Twitter user is contrasting the succeeding phrase to the preceding phrase. The phrase before this segment indicates that fire threat is a considerable distance from the Twitter user's location, yet despite the 50 mile gap, s/he is wetting the yard to deter potential flames from starting a fire. The segment ends with "anyway," which is used to confirm the tweet author's preceding point. The use of "but" and "anyway" in this tweet work together to indicate that the tweet author is not overly concerned with fire threat, but threat of fire does exist, and the Twitter user is responding as s/he sees fit.

2. RR09: `@jaisonf I feel like the captain of a ship. If it goes down I'm going with it. But wife, kid & pets: different story.`

   Segment: `If it goes down I'm going with it.`

In this segment, the reader learns that the tweet author is responding to flood threat by staying in his home for the duration of the 2009 Red River flood.[20] The Twitter user employs a metaphor by referring to his home as a ship in the preceding phrase, and himself as the captain, or owner, of that ship. His claim "if it goes down I'm going with it" indicates that he will stay in his home even if it floods.

During flood threats, many area residents choose to remain on their property, despite warnings and even orders to evacuate.[21] Residents know that during a record-breaking event like the 2009 flood, protecting personal

---

[20] I use the possessive "his" due to the reference to "wife" and the assumption that the term "wife" indicates a heterosexual marriage.

[21] http://minnesota.publicradio.org/display/web/2009/04/22/flood_staying_put/

116

property is largely their own responsibility. Therefore, many stay close to

home in order to create levees and/or ring dikes to protect their homes. They

also monitor water levels in the area and watch for the beginnings of flood

damage by constantly checking for water in basements.

3. RR10: `People I know are packing up their houses and moving out to avoid the flood. Last year they were helicoptered out. Trying to avoid that.`

    Segment: `People I know are packing up their houses and moving out to avoid the flood.`

In this segment, the tweet author informs the reader that people s/he knows are

packing and moving out of their homes in response to the flood threat of 2010.

The phrases following this segment indicate that the previous year, in the

flood of 2009, these same people had to be taken out of their homes by

helicopter. In 2010, they were responding to the potential that the same course

of events may take place by choosing to leave before the flood situation

became dire and the use of helicopters became necessary.

4. Haiti: `Wow! He flew a PLANE not a small plane, filled with food, water etc to haiti.. Wow! Thanks John Travolta... Thank YOU! #cultivated`

    Segment: `He flew a PLANE not a small plane, filled with food, water etc to Haiti`

This segment indicates that an individual flew a plane of supplies to Haiti. The

"he" in this segment is a cataphoric reference to the following phrase which

names who "he" refers to: the American actor John Travolta.[22] The Haiti

earthquake spawned response from an international audience, and many

---

[22] Cataphora occurs when a linguistic unit such as a pronoun refers forward to another unit. It the
contrast to anaphora, which occurs when a linguistic unit refers back to another unit (Hudson, 2000 p.
55).

celebrities and famous personalities took personal measures to help with response efforts. In the case of John Travolta, that response involved using his plane flying skills to pilot a plane of supplies to Haiti.

Information Type: Sheltering

Sheltering labels tweets that include information about shelter information in general (e.g. that they exist or are needed) and where shelters for people are located in areas under threat or suffering from effects of a disaster.

Note: No *Sheltering* tweets were coded in the RR10 or Haiti datasets.

1. OK09: `Fire Warning for Midwest City.  Fire threatening Turtlewood, Windsong, and Oakwood additions should evacuate to Community Center #okfire`

   Segment: `Turtlewood, Windsong, and Oakwood additions should evacuate to Community Center`

   This segment is part of a tweet that contains many types of information. The sheltering information communicated in this tweet is provided when the tweet author writes that three areas of Midwest City, Oklahoma should evacuate, and specifies that they should "evacuate to Community Center." The order to "evacuate to" a particular place is an implicit indication that the location to which people are to evacuate is a shelter or other safe place. In this segment, the reader learns that "Community Center"—which likely refers to the Midwest City Community Center, as this is the only community center in the area—is serving as a shelter.

2. RR09: `Flood Shelter Open at Alerus Center in Grand Forks, ND @ 4:00 pm  Wed - RedCross (218)773-9565 #redriver #fargoflood #ndfloods #flood09`

   Segment: `Flood Shelter Open at Alerus Center in Grand Forks`

This segment tells the reader that a shelter has been set up at the Alerus Center, located in Grand Forks, North Dakota. The Alerus Center is a large event center used for concerts, athletic games, conventions and other events that draw large numbers of people. In addition to informing the reader that a shelter has been established during the 2009 Red River flood, this segment may also indicate the severity of the flood by informing readers that the situation has reached a point where a shelter is necessary, and that a large facility is the site of the shelter.

Information Type: Status – Community/Population

Status – Community/Population labels tweets that provide information about the overall state or condition of a community or population. Information in this category may be positive, i.e. that people are safe and unharmed, or that people are in need, in a situation of duress or similar unfortunate circumstances.

1. OK09: `Residents allowed to return to homes. Firefighters adddressing hotspots: http://tinyurl.com/cf5p4s #OKfire`

   Segment: `Residents allowed to return to homes`

   This segment informs the reader that residents of areas either under threat or damaged by fires were "allowed" to return to their homes. The use of "allowed" indicates that residents were either ordered to leave their homes or opted to leave their homes, and that the area where the homes are located was cordoned off or otherwise blocked from entry. The permission to return to homes also indicates that residents' homes are either undamaged or are damaged but still standing, i.e. they were not destroyed by the fires.

119

2. RR09: `close to home now: RT @cbcmanitoba: Residents on Kingston Crescent, off of Osborne Street, may be at risk for flooding. #flood09 #winnipeg`

   Segment: `Residents on Kingston Crescent, off of Osborne Street, may be at risk for flooding.`

   This segment informs the reader that residents of a particular neighborhood in Winnipeg, Manitoba may experience flooding. In a situation like the 2009 Red River flood, threat is not isolated to one or two homes or buildings. When threat arises, it applies to large areas such as entire neighborhoods or other tacitly defined regions within a particular area.

3. RR10: `I heard on the news that we're having a flood disaster. Really? A flood fight for sure. But a disaster? C'mon people knock it off.`

   Segment: `I heard on the news that we're having a flood disaster`

   In this segment, the tweet author conveys information s/he heard on the news, which is that "we're" experiencing a flood disaster. The use of "we're" in this context is a reference to the community or region as a whole. According to the news, i.e. the media, the region is experiencing a disaster.

   Segment: `A flood fight for sure. But a disaster? C'mon people knock it off.`

   In this segment, the tweet author is referring to the previous segment, and indicating that "we." or the community or region of which s/he is a resident, is not in fact experiencing a disaster. The Twitter user states that the area is fighting the flood, but that no disaster has taken place. She goes on to imply that "people" should stop sensationalizing the event by referring to it as a disaster. The referent of "people" is not clear – the author could be referring to the media, or s/he could be referring to his/her fellow community members

who may be describing the event as a disaster. Overall, this tweet author indicates that s/he does not think the area is experiencing a disaster; the situation is less serious and should not be considered disastrous.

4. Haiti: `New adoptions from Haiti have been stopped for good reason, yet every other story out of Haiti is about the extreme suffering of the kids.`

Segment: `New adoptions from Haiti have been stopped for good reason`

This segment tells the reader that "new adoptions" of Haitian children have been stopped "for good reason." The implication of "new" as a description for adoptions is used as a contrast to adoptions that were already underway before the earthquake happened. After the earthquake struck, many children were left orphaned. However, many children also had parents and other family members to care for them, but had become separated from their families during the earthquake. There was concern among aid groups that infants and children who were not orphans could be mistaken for orphans and adopted by foreign families if new adoptions were permitted to take place in the immediate aftermath of the earthquake. These new adoptions would break up existing families and cause great anguish and heartache for all involved, which is why new adoptions were halted, i.e. this is the "good reason" to which the tweet offer refers.[23]

Segment: `yet every other story out of Haiti is about the extreme suffering of the kids.`

[23] http://edition.cnn.com/2010/WORLD/americas/01/21/haiti.adoptions/

This segment tells the reader that kids are experiencing great suffering in Haiti. The tweet author begins this segment with "yet," which indicates that the phrase which follows "yet" is in contrast to the previous phrase. The previous phrase informs the reader that new adoptions have ceased in Haiti, and for good reason. However, the author of this tweet appears to be expressing frustration at the overall state of the Haitian people find themselves in; others want to help Haitian children, but the circumstances of the earthquake and its devastating effects create a situation rife with obstacles to providing help or to alleviating suffering.

The two segments in this tweet confront two separate issues, but the author attempts to link the issues. The first segment informs the reader about the new adoption halt because it is possible that people may adopt children who have families. Strictly speaking, the adoption issue has little to do with the fact that children are suffering. While one could argue that adopting orphaned children would alleviate suffering among some children, the fact remains that Haitian kids are in need of food, medical attention and basic care, which is a separate issue that requires its own attention.

Information Type: Status – Personal

Status – Personal labels tweets that provide updates or information about individuals or small groups of people (e.g. families). Information may be provided in the first person, i.e. people may self-report, or they may offer information on behalf of others.

122

1. OK09: `Just to let my friends know, I'm OK, and no fires are near me, at the moment`

   Segment: `I'm OK`

   In this segment, the author informs readers that s/he is okay, i.e. not injured or experiencing other ill effects of the Oklahoma fires. Using Twitter or social media applications as a means to broadcast this "I'm okay" message is not uncommon in mass emergency situations (cf. Palen et al., 2007; Vieweg et al., 2008). The capabilities of Twitter allow this user to simultaneously send the message to many people that s/he is safe.

   Segment: `no fires are near me, at the moment`

   In this segment, the author informs readers that s/he is currently out of harm's way. However, the author qualifies the statement that "no fires are near me" with "at the moment." The use of "at the moment" serves as a caveat; the tweet author is indicating that fires may spread near him/her, but at the present time, fires are not in the immediate vicinity.

2. RR09: `Brushing my teeth with bottled water feels a like being out camping`

   Segment: `Brushing my teeth with bottled water`

   This segment informs the reader that the tweet author is in a situation that requires bottled water for everyday use. In the context of a major flood, this segment reveals that the author is in a location, likely his/her home, that has either lost access to public water, or in which the tap or well water is potentially contaminated and unsafe for human consumption. In such situations, people turn to bottled water for basic needs.

123

3. RR10: @happyhero Nah, I'm good..used to live right next to the dike but I moved after the last flood! This year I'm all set.

Segment: Nah, I'm good

In this segment, the tweet author is answering a question, likely a question which asked how s/he was doing with regards to the 2010 flood. The author answers with "Nah, I'm good." The "nah" is a negative response to whatever question s/he is answering, followed with "I'm good," which implies that the personal situation of the Twitter user is fine and s/he is out of harm's way.

Segment: used to live right next to the dike but I moved after the last flood! This year I'm all set.

In this segment, the tweet author tells the reader that s/he used to live next to "the dike," which is an unmarked reference to a dike, indicating that the author assumes the Twitter user to whom s/he is directing the tweet understands the reference to "the dike." Living near a dike is an indication that a person lives in an area that floods often, or that is under risk for flooding often. The tweet author goes on to say s/he "moved after the last flood." In the context of the 2010 Red River flood, the "last flood" refers to the 2009 Red River flood. The tweet author chose to relocate from a location that was in an area highly prone to flooding to his/her present location which is in a less risky area. At the time this tweet was sent, the author was "all set," indicating that s/he was in a safe location and not concerned about falling victim to the flood.

4. Haiti: My friend's mother, who'd been missing in Haiti, has been found alive and well!! http://bit.ly/4y3hDp

Segment: My friend's mother, who'd been missing in Haiti, has been found alive and well!!

124

In this segment, the tweet author relays information to the reader that a particular individual with whom the author has a personal connection, has been "found alive and well." The segment indicates that the person was previously missing, but has now been located and is safe and unharmed.

4.1.1.2 Built Environment

In this section, I describe tweets that fall under the *Built Environment* category, which includes information about structures, facilities, public infrastructure and various forms of property. Each tweet is labeled with its corresponding dataset, and below each tweet, the segment(s) which provides the specific information type is highlighted and described.

Information Type: Damage

Tweets labeled with the *Damage* category convey information about property or some component of the built infrastructure that has suffered from the disaster. These tweets may mention damage to public or private structures, and property of many varieties.

1. OK09: `OK Emergency Mgt reports at least 140 buildings destroyed, not counting many in Carter and Payne Counties. More info: http://bit.ly/13YKwG\`

   Segment: `at least 140 buildings destroyed, not counting many in Carter and Payne Counties`

   This segment indicates that the grass fires in Oklahoma have destroyed "at least" 140 buildings that the tweet author is aware of. The use of "at least," in addition to the phrase "not counting many in Carter and Payne Counties" indicates that more buildings have been destroyed, and it is likely that further information on damaged buildings will be forthcoming.

125

2. RR09: `Erosion under/along dike near Oak Grove caused a`
   `floodwall to buckle. School was inundated, and likely lost to`
   `floodwaters. SAD Red 40.27`

   Segment: `a floodwall to buckle`

   A floodwall is a temporary or permanent structure built in areas that are prone

   to flooding. The purpose of a flood wall is to hold water back from seeping

   into surrounding areas as a river or other body of water rises. The damage

   indicated in this tweet is that a floodwall has buckled, or collapsed, due to

   erosion of the earth supporting it.

   Segment: `School was inundated`

   The second segment in this tweet that tells the reader about damage refers to a

   school building that was "inundated," meaning that after the flood wall was

   damaged, water poured over the buckled wall and flooded the nearby school.

3. RR10: `ChastityWalberg: In '09, #Fargo restored the backyards of`
   `237 homes damaged by levees, claims averaged $1,700 – this yr.`
   `not as much damage #fmflood #flood10`

   Segment: `this yr. not as much damage`

   This tweet segment communicates to the reader that there is damage from the

   2010 Red River floods. However, the phrase "not as much" indicates that

   though there is damage, it is not as widespread as damage from a previous

   year—likely the previous year of 2009—which was a record-breaking flood

   year. From this tweet, the reader learns that 2010 flood had a negative impact

   on buildings and infrastructure, but the extent of the damage includes fewer

   than 237 homes and less than an average of $1,700 in insurance claims.

4. Haiti: `Big Haiti quake topples buildings, many casualties (Reuters)http://bit.ly/84Fc3G`

   Segment: `Big Haiti quake topples buildings`

   In this segment, the reader learns that the earthquake toppled buildings. The use of the verb "topple" indicates that buildings were knocked over and are likely demolished and no longer functional or available for use.

Information Type: Status – Infrastructure

The *Status – Infrastructure* category includes information about the state or condition of various physical, human-made elements found in the built environment such as roads, airports, electricity and similar items.

As the analysis developed, one aspect of this category that arose is the difference between permanent and temporary infrastructure. In the RR09 and RR10 datasets in particular, there are mentions of flood prevention measures which usually mention construction of dikes and/or levees. It is not always clear if Twitter users are communicating about permanent or temporary dikes, as the Red River Valley is home to both types of structures during flood season. Therefore, I include any mention of dike or levee in this category, regardless of whether it is a permanent part of the infrastructure or if it is a temporary structure constructed in response to imminent or predicted flooding. I provide more than one tweet per dataset in this category, as some *Status – Infrastructure* tweets include a lot of information specific to particular emergency events.

1. OK09: `#okfire I-35 reopened in Payne County, HWY 51 still closed`

   Segment: `I-35 reopened in Payne County, HWY 51 still closed`

Here, the reader is made aware of the status of highways in an area affected by grass fires; one highway is open for travel, the other remains closed. This information is useful for those living in or traveling to the area so they can plan which route to take.

2. RR09: `If the floodway can't open to divert water around the`
`city when the crest hits, levels could reach 25 feet inside`
`Winnipeg. Scary  #flood09`

Segment: `If the floodway can't open to divert water`

This segment tells the reader that there is concern about water diversion in the Red River floodway, referred to by its unmarked form of "the floodway." The Red River floodway is a human-made channel used to shift part of the river's flow during a flood and divert it back into the Red River at a point further downstream. If the floodway is not functioning, water flows right through the city of Winnipeg, potentially causing flooding and damage. In this tweet, there is concern about the floodway and whether it will function as expected to divert water away from Winnipeg.

3. RR09: `Horace, small suburb of Fargo, just lost water b/c lift`
`stations failed.`

Segment: `just lost water b/c lift stations failed`

This segment mentions the loss of water due to the failure of lift stations, which are part of the public sewage system. The lift stations work by taking raw sewage from underground pipes, storing it in pits or wells, and eventually moving it to treatment plans, which are at a higher elevation. When lift stations fail during a flood, it is because too much water has flowed into the pits or wells, causing raw sewage to back up and leak into the surrounding

128

environment. This tweet tells us that this particular aspect of the public infrastructure is not functioning due to flooding.

4. Haiti: `CNN International Desk: landlines and local cell phones are still down in #Haiti but internet seems to be working (via @vhernandezcnn)`

   Segment: `landlines and local cell phones are still down in #Haiti`

This segment informs the reader that both landline telephone infrastructure and cellular phone infrastructure are not functional in Haiti. The downed phone lines make communication difficult, as many rely on landline and cellular phones.

Segment: `but internet seems to be working`

This segment begins with "but," which serves to contrast the succeeding phrase to the preceding phrase. The previous phrase tells the reader that phone use of any sort is not possible in Haiti, but there is the possibility of communication via the Internet. However, the tweet author qualifies his/her claim about the functioning Internet with "seems." By using the phrase "seems to be working" to describe the state of the Internet, the tweet author is not claiming that Internet communication is functional beyond question.

Information Type: Status – Personal Property

The *Status – Personal Property* category labels those tweets that offer information about the condition of property that is privately owned by individuals or families. This type of property is separate from facilities such as schools, community centers, businesses and other publically-accessible establishments.

129

1. OK09: `Taking a deep breath after another afternoon of fires threatening homes. Luckily, no homes lost in this round. I'm glad rain is coming!!`

   Segment: `fires threatening homes`

   This segment indicates that homes are currently at risk for damage/destruction

   from fires. The reader learns that fires are ongoing, and that they are

   potentially impinging upon personal property.

   Segment: `no homes lost in this round`

   This segment tells the reader that though fires are a threat to the integrity of

   personal property, thus far, no homes are "lost." It is important to note that not

   being "lost" does not mean homes or other property are not damaged, it

   simply means no homes were rendered uninhabitable or completely destroyed.

2. RR09: `None of the homes in the voluntary evacuation areas have been flooded yet. #fargoflood #pressconf`

   Segment: `None of the homes in the voluntary evacuation areas have been flooded yet`

   This segment conveys information about the status of personal property in the

   form of homes. It tells the reader that though areas have been evacuated due to

   the Red River flood, none of those homes has suffered any damage. This

   information serves as an indication that the flood is serious enough to have

   caused evacuations (albeit voluntary evacuations), but it has not gotten to the

   point of exacting damage on any homes. However, the tweet author qualifies

   the tweet with the use of "yet," indicating that flooding is a possibility.

3. RR10: `@thebroadbroad Thanks darling! My house is safe but my parents live by the river so we have been sandbagging it's looking better`

   Segment: `My house is safe`

In this segment, the reader learns that the home of the tweet author is safe from flood threat and currently undamaged.

Segment: `but my parents live by the river so we have been sandbagging it's looking better`

The tweet author goes on to tell the reader that his/her parents' house is near "the river," meaning the Red River. The author starts the phrase with "but," indicating that the succeeding information is in contrast to the preceding information; in this case, one house is safe, while another may be in danger. The tweet author has been sandbagging his/her parents' home due to its potentially precarious state as it's located near the river. The tweet author ends this segment with "it's looking better." The use of "it's" may imply a few things. It could mean "the flooding situation as a whole" is looking better, or it could mean "the author's parents' home" is looking better. Regardless of what "it" refers to in this segment, the author ends the tweet on a positive note, indicating that conditions of some sort are improving.

4. Haiti: `The earthquake collapsed many houses in Port-au-Prince, Haiti. I wish to go there, so I think to help something to do at this time.`

Segment: `The earthquake collapsed many houses in Port-au-Prince, Haiti.`

This segment informs the reader that houses were "collapsed" due to the Haiti earthquake. The use of the verb "collapse" indicates that homes were demolished and are no longer functional nor useful. The tweet author also qualifies his/her statement with "many," implying that more than a few homes have been destroyed by the earthquake.

Information Type: Status – Public Property

This *Status – Public Property* category labels tweets containing information about the condition of public property, such as schools, hospitals, libraries and other property that is available for public use.

1. OK09: @cnn fire department in carter county oklahoma burned. documents, and a firetruck

   Segment: fire department in carter county oklahoma burned. documents, and a firetruck

   This segment conveys to the reader that a fire department building, a fire truck and some documents were burned. Public property extends to items such as vehicles and paperwork, though they are not available for use by any member of the public and/or citizen. This tweet communicates the effect of the fire in telling the reader that the fire department in Carter County, Oklahoma has experienced damage that extends beyond damage to its structure; from this, the reader can postulate that the fire department is likely not able to function in its normal capacity.

2. RR09: Retweeting @inforum: Dike fails at Fargo's Oak Grove Lutheran School, two buildings lost http://is.gd/puWd

   Segment: two buildings lost

   This segment tells the reader that two buildings, which are associated with a school, are "lost," meaning it is likely they are destroyed or damaged beyond repair. It is important to note that though the buildings mentioned in this tweet are part of a private school campus, this tweet is labeled with *Status – Public Property* to avoid confusion; mention of buildings such as schools are placed in this category, regardless of their strict classification as public or private.

132

3. RR10: `RT @redriveratfargo: Nothing to worry about, but there is now a waterfall filling up the el zagel golf course.`

   Segment: `a waterfall filling up the el zagel golf course`

   This tweet lets the reader know that a public golf course—which is near the Red River—is experiencing effects of the flooding river. The tweet author specifies that the situation at the golf course is "nothing to worry about." This type of tweet conveys information about the status of public property in stating that the property is essentially okay. This tweet is also letting the reader know that this particular area is "filling up," that the condition of the golf course is not cause for immediate concern. This type of information—which communicates that property is undamaged and in tact—provide the reader with information about what areas are under threat and what areas are not cause for concern.

4. Haiti: `FTR NEWS: Haiti: Presidential palace and hospital among buildings hit`

   Segment: `Presidential palace and hospital among buildings hit`

   This segment informs the reader that two structures have been "hit." The presidential palace is the home of the Haitian president, but for purposes of this analysis, I treat it as public property as it is a facility often used for large public gatherings and other civic events. In the context of an earthquake, "hit" implies that the buildings were damaged by the earthquake. This segment also tells the reader that these two facilities were "among" buildings hit, meaning that there were more buildings that suffered damage.

4.1.1.3 Physical Environment

In following section describes tweets that detail information about the *Physical Environment*. This includes geographical information about where each disaster occurred, information about hazard agents, and weather updates. Each tweet is labeled with the dataset to which it refers, and under every tweet example, the segment(s) that indicate information type is explained.

Information Type: General Area Information

The *General Area Information* category labels tweets that communicate geographic or logistic information about the region under threat or in which a disaster has occurred. This type of information helps receivers of information gain an understanding about the physical aspects of the area, which can lead to greater awareness of the disaster situation overall.

1. OK09: `Midwest City is just east of Oklahoma City, Tinker AFB is located there.`

   Segment: `Midwest City is just east of Oklahoma City`

   This segment provides the reader with a geographic fact—that Midwest City, which is an area affected by the grass fires, is east of Oklahoma City, which is also affected by the fires. By reading this segment, the reader can infer that the fires are spreading east from the more well-known and populated Oklahoma City (the fires may also be spreading in other directions, but this tweet tells the reader of at least one direction.)

   Segment: `Tinker AFB is located there`

134

The second phrase in this tweet also provides the reader with basic geographic information. We read that an air force base is located east of Oklahoma City, in Midwest City. Alternatively, the tweet author may be using the air force base as a reference for locating Midwest City. Though the reader of this tweet cannot know the exact motivation for the tweet author to point out the location of the air force base, learning that there is a major military center in the area under threat of fire tells the reader about the affected population.

2. RR09: `@skangus GF built permanent levees after 97 flood of the Red. Difference was they declared entire neighborhoods greenspace to do it.`

   Segment: `GF built permanent levees after 97 flood of the Red.`

This segment tells the reader that there are permanent levees in "GF" which stands for Grand Forks, one of the more populated cities situated along the Red River. A levee is a structure whose purpose is to stop water from flowing into surrounding areas when the river floods; in Grand Forks, there are permanent levees, meaning they are long-term structures and not temporarily built each flood season. From this tweet, the reader also learns that there are permanent *levees*, i.e. more than one, in Grand Forks, which indicates it is a region which regularly undergoes flood threat.

Additionally, the tweet author refers to the Red River as "the Red." Shortening the title of the river from "Red River" to "the Red" is a form of *clipping*. Clipping is a process in which words or phrases are shortened or "clipped" from their usual, longer form (Hudson, 2000, p. 242). The tweet author is using the *marked* form of the Red River, but is clipping the marked form into a shorter fragment.

135

Segment: `they declared entire neighborhoods greenspace to do it`

In this segment, the tweet author refers to the previous sentence in the tweet. "They" refers to "GF," meaning the city of Grand Forks, and "it" refers to the building of permanent levees. This segment provides the reader with information about the Grand Forks area by explaining that the city—or city officials—likely exercised some form of eminent domain to take neighborhoods, or an area where people had homes, and declared them "green space," or government-owned property. This action permitted the city of Grand Forks to then build permanent levees on the land.

3. RR10: `@femaregion8 ND and MN and border the Red River which is flooding now.`

   Segment: `ND and MN and border the Red River`

   In this segment, the reader is informed of the location of the Red River, as well as of two states that border it. The reader learns that both "ND," or North Dakota, and "MN" or Minnesota, are adjacent to the Red River. If the reader is unfamiliar with this particular region of the United States, but has a basic understanding of American geography, s/he learns that the Red River lies between these two states which abut one another.

4. Haiti: `Haiti is only 1/2 of the island it's on. #Haiti`

   Segment: `Haiti is only 1/2 of the island it's on.`

   This segment informs the reader that Haiti is an island nation. It additionally informs the reader that Haiti is "only 1/2 of the island," which tells the reader that there is another country on the other half of the island. Haiti's status as an island nation, and its proximity to another country could both be useful

136

information for those interested in understanding more about the 2010 earthquake situation.

Information Type: General Hazard Information

General Hazard Information labels those tweets that provide information about the hazard agent, but are not specific to the particular mass emergency situation; i.e. readers of these tweets learn about fire, flood or earthquakes in general. These tweets may be about the particular region in which a hazard is exacting threat or damage, but the information is such that it is not specific to the particular year or time the disaster is taking place.

1. OK09: @Leosulli firestorm is a fire. Wind is Hi n fuels fire. It spreads n wind carries it. fire 'jumps' from house to house. Cedars, grass catch

   Segment: `firestorm is a fire.`

   Here, the reader learns that a "firestorm" and a "fire" are the same thing. Both words were used to describe the 2009 grass fires in Oklahoma, and the tweet author is clearing up any confusion about a difference between the two descriptors.

   Segment: `Wind is Hi n fuels fire.`

   This segment specifically describes the Oklahoma fires, and in doing so, it also provides information about wildfires in general. The reader first learns that at the moment the tweet was sent, "Wind is Hi," meaning that the area is experiencing strong winds. Following the first phrase, the Twitterer tells the reader that the strong wind "fuels fire," which is a comment about wild fires in general.

   Segment: `It spreads n wind carries it`

Similar to the previous segment, this segment provides information about the relationship between wind and fire, namely that wind serves as a catalyst for the spread of fire. From this segment, the reader can also infer that the direction of the wind relates directly to the direction the wildfire will spread.

Segment: `fire 'jumps' from house to house.`

In this segment, the reader learns that fire can "jump from house to house," which is again a comment about how fire functions. If a fire is blazing in one house, those houses around it are also at risk for damage if the fire "jumps."

Segment: `Cedars, grass catch`

The final segment of this tweet also offers information about the general behavior of a fire. The reader learns that "Cedars, grass catch," meaning that area flora are at risk for catching fire, which can be extremely damaging, as the trees and grass act as kindling for the flames that can quickly spread.

2. `RR09: @skangus Red River is like a pencil scratch in kitchen table. Plus it flows north. So downstream tends to backup.`

   Segment: `downstream tends to backup`

   Here, the reader learns that downstream, or the section of a river that is situated away from the source, is likely to "backup," or experience flooding. The previous phrases in the tweet provide background information regarding why downstream is prone to flooding, and the reader learns that rivers located in very shallow regions are also likely to experience flooding.

3. `RR10: @scurker Ever since I moved here ;-) Yeah, it floods yearly - spring melt + still frozen ground + flat as a sheet of paper = flooooooooooood.z`

   Segment: `it floods yearly`

This segment explicitly and clearly tells the reader that the area undergoes flooding every year.

Segment: `spring melt + still frozen ground + flat as a sheet of paper = flooooooooood`

The second segment in this tweet provides information about why flooding occurs in the Red River Valley. The reader learns that the "spring melt" happens when surrounding ground is "still frozen" in an area that is "flat as a sheet of paper," and that these three characteristics together create conditions for flooding to take place.

4. Haiti: `@GottaLaff The earthquake happened because Haiti is on top of a major fault. Like LA and SFO are built on top of a major fault.`

Segment: `The earthquake happened because Haiti is on top of a major fault.`

This segment provides information specific to the Haiti earthquake, and in doing so, informs the reader about earthquakes in general. The segment states why the Haiti earthquake occurred, because it is situated "on top of a major fault." A fault is a fracture, or volume of loose rock situated between larger, more stable blocks of the earth's crust.[24] Earthquakes occur along faults, and because Haiti is on a "major fault," a high magnitude earthquake took place.

Information Type: Historical

Those tweets labeled as *Historical* refer to information comparing present hazard conditions to past hazard conditions of the same type. For example, during the Red River flooding events, some tweets offer information comparing the 2009 event

---

[24] http://www.fema.gov/hazard/earthquake/why.shtm

to the 1997 floods. And in the 2010 Red River floods, some tweets compare to the previous year's floods.

1. OK09: "This is probably the worst disaster in Midwest City's history, even worse than the May 3rd tornadoes", PD Chief Brandon Clabes.

   Segment: This is probably the worst disaster in Midwest City's history

   This segment informs the reader that the 2009 Oklahoma fires are one of the worst, if not the worst, disaster the region has experienced. This statement conveys to the reader the grave situation area residents are facing.

   Segment: even worse than the May 3rd tornadoes

   In this segment, the reader learns that the region had experienced a previous mass emergency in the form of a tornado. In May 2008, Oklahoma was one of three states to experience a deadly and destructive tornado.[25] Though the 2009 fires did not cause as many deaths as the 2008 tornado, the Midwest City, OK police chief claims the fires are worse than the tornado, which may imply that more damage took place, and/or more residents are displaced.

2. RR09: @Petiesperson i'm about 70miles north of fargo. nd. in grand forks. flooded here in 1997. lost my house. some deja vu. weird.

   Segment: flooded here in 1997.

   This segment tells the reader that the tweet author experienced flooding during the Red River flood of 1997. Until 2009, the 1997 flood had been the most devastating flood in the region in recent history.

---

[25] http://www.cbsnews.com/2100-201_162-4088493.html

Segment: `lost my house.`

In this segment, the reader learns that the tweet author has experience with the flood in the form of property loss; his/her home was damaged beyond repair due to the 1997 flood. This segment additionally tells the reader about the severity of the 1997 flood, and provides insight into the threat the area is facing in 2009.

Segment: `some deja vu.`

Déjà vu is a feeling of having already experienced something. This tweet author is referring to this particular feeling in comparing the 1997 flood to the 2009 flood. As I stated above, the 1997 flood had exacted the most damage to the Red River Valley until the flood of 2009; those residents who experienced both floods frequently made comparisons between the two flood events as a way to convey the seriousness of the 2009 flood.

3. RR10: `I don't remember it going from "look at all the snow" to "OMG where did all this water come from?" So rapidly last year.`

   Segment: `I don't remember it going from "look at all the snow" to "OMG where did all this water come from?" So rapidly last year.`

In 2010, Red River Valley residents were again in a position to compare flooding events, but this time around, most were using the previous year as a reference point. The author of this tweet experienced the Red River flood of 2009, referenced by "last year," and is remembering that event one year later. The Twitterer indicates that the Spring thaw happened more quickly in 2010 than it had in 2009, which may be cause for worry, considering the severity of the 2009 flood.

4. Haiti: `The World Series quake of '89 (SF Bay area) was of a similar strength to the one in #Haiti, but killed only 63. Big diff: building codes!`

   Segment: `The World Series quake of '89 (SF Bay area) was of a similar strength to the one in #Haiti, but killed only 63.`

This segment makes a historical comparison between the 2010 Haiti earthquake, and an earthquake that occurred in San Francisco, California approximately twenty years previously. The 1989 San Francisco earthquake was magnitude 6.9, very similar to the magnitude 7.0 2010 Haiti earthquake.[26] However, the tweet author points to a major difference between the two earthquakes, namely that the 1989 earthquake resulted "only 63" deaths. The use of "only" indicates that sixty-three is a much lower amount than the number of deaths in the Haiti earthquake, which is in fact the case.

Information Type: Prediction

Tweets labeled with *Prediction* provide information about forecasts or even guesses of what will happen with regards to the hazard, e.g. how high a river will rise or where a fire will move.

Note: No *Prediction* tweets were coded in the Haiti dataset.

1. OK09: `Well this fire here in oklahoma it lloks like its going to get bigger than the 12 miles it is now`

   Segment: `Well this fire here in oklahoma it lloks like its going to get bigger than the 12 miles it is now`

In this segment (which is the entire tweet), the tweet author informs the reader that it appears the fire in Oklahoma is going to grow, and spread beyond the twelve-mile area it currently incorporates. The use of "lloks like" (sic) to

---

[26] http://pubs.usgs.gov/dds/dds-29/

142

mean "looks like" implies a prediction, i.e. the tweet author does not know for certain if the fire will grow and become bigger, but from his/her perspective, that appears to be the case.

2. RR09: `Saw an article last night predicting 2 seperate crests in Grand Forks. We could be bridgeless for awhile. #flood09`

   Segment: `article last night predicting 2 seperate crests in Grand Forks.`

   This segment tells the reader that as of the previous evening, the Red River was expected to experience two separate crests. Two crests means that at two different points in time, the river will hit a maximum level before it recedes.

   Segment: `We could be bridgeless for awhile.`

   In this segment, the tweet author implies that "We," meaning the community of Grand Forks, "could be bridgeless for a while." The use of "could" implies a prediction of a future state regarding the transportation infrastructure in the city.

3. RR09: `@fpbowen they are talking '79 levels - will most likely lose road access :(`

   Segment: `they are talking '79 levels`

   This segment is particularly interesting because it references '79, meaning 1979. The most devastating flooding experience in the Red River Valley until 2009 was in 1997. However, the area also experienced severe flooding and considerable damage in years prior to 1997. The Red River flood of 1897 was the largest on record until 1997, but in 1979, the second largest flood as compared to 1897 struck the valley.

143

There are a few ways to interpret the year reference in this segment. One possibility is that the tweet author accidently transposed two numbers, and mistakenly wrote "'79" instead of "'97." Another possibility is that the Twitterer has historical knowledge of the Red River Valley floods, and at the time when this tweet was written, the severity of the 2009 flood was not yet known nor predicted to be as bad as the 1997 flood; therefore, the tweet author referenced a previous flood that was also damaging, but which was not as bad as that of 1997 nor 1897. Regardless of the intent of the author, knowing that "they," which likely refers to formal response agencies and/or weather forecasters, are making predictions based on past large floods is cause for concern.

Segment: `will most likely lose road access`

In this segment, the tweet author is making the prediction that "road access" will likely be lost, meaning certain roads will no longer be accessible and residents will have to find alternate routes.

4. RR10: `River Projections Looking Positive: The river level is up, but so are spirits tonight. http://bit.ly/cZ80H6`

Segment: `River Projections Looking Positive`

This segment mentions that the projections, or predictions, are "Looking Positive." This segment appears straightforward; predictions regarding the state of the river are favorable or good. However, it is important to note that for a prediction (or anything) to be "positive," or advantageous, the reader must understand that point of view of the deliverer of the message. What does "positive" imply in this tweet segment? Based on the context of the 2010 Red

144

River flood, and the norms that exist around the yearly threat in the region, "positive" many have any number of detailed meanings. Regardless of what the tweet author specifically means by "positive" in this context, the reader learns that those in a position to predict the future state of the river are hopeful and optimistic that flooding and subsequent damage will not occur.

Additionally, in this tweet, the author refers to "River" and "The river." The reader can assume it is the Red River to which these phrases refer, as the sampling method used to extract the tweets in the RR10 dataset specify finding mentions of the Red River. Twitter users write "the river" to refer to the Red River throughout this and the RR09 dataset, because the Red River is the *unmarked* river in the region. In the above example, the tweet author mentions the Red River by relying on the practice of "unmarking" common linguistic forms.

Information Type: Status – Hazard

This information type labels tweets that provide information about the location and/or state of a hazard, such as where a fire is burning, the current level of a river or other body of water in relation to flood levels or the strength of an earthquake and its aftershocks.

1. OK09: `RT @rustysurette: BIG fire south of downtown OKC.  Thick smoke covering the metro.  It's a grass fire in Newcastle`

   Segment: `BIG fire south of downtown OKC`

   Here, the reader learns that there is a "Big fire" and the location of the fire, which is "south of downtown OKC." This information provides the reader

with an understanding of the fire size, and what area of Oklahoma is experiencing the fire.

Segment: `Thick smoke covering the metro.`

This segment tells the author about "Thick smoke," which serves as an indication of the severity of the fire; the larger the fire, the more smoke is created. The smoke referred to in this segment is "covering the metro," which likely means it is covering the "metro area" of Oklahoma City. Additionally, because we know that the fire is located south of Oklahoma City, and the smoke is covering the surrounding area, it is likely that wind is playing a factor and is additional cause for concern.

Segment: `It's a grass fire in Newcastle`

Newcastle, Oklahoma is a city located approximately 20 miles south of Oklahoma City. In this segment, the reader learns that there is a fire in this city. When the reader considers this tweet segment in the context of the rest of the tweet, s/he can infer that the fire is quite serious, as it is causing thick smoke to spread and cover a large area.

2. RR09: `hugh ice jams form on the red river in winnipeg, lower lying homes flooded. Emerson manitoba has there creast`

   Segment: `hugh ice jams form on the red river in Winnipeg`

   This segment tells the reader that "hugh ice jams" (sic) have formed on the Red River in Winnipeg, Manitoba—a city with a population of almost 700,000—located downstream on the Red River. Ice jams are particularly concerning, as they block the flow of the river and can cause water to spill onto the banks and into the surrounding area.

Segment: `Emerson manitoba has there creast`

In this segment, the reader learns that Emerson, a city located just over the North Dakota border, has experienced a river crest. Emerson is a city approximately 70 miles south of Winnipeg. Knowing that Emerson has seen the river crest provides the reader with information about the status of the river flow and the flood situation overall.

3. RR10: `Just got home. 50 min drive turned into 2 1/2 hour odyssey cuz Sheyenne river is flowing over the southbd I29! Fun drv thru flood waters!`

    Segment: `Sheyenne river is flowing over the southbd I29`

    Here, the reader learns that the Sheyenne River, which is a tributary of the Red River, has overflowed and is running over a north-south interstate in Fargo, North Dakota. This segment tells the reader about the flood situation by conveying that a particular road is covered in water; if the reader is aware of or learns about the location of the Sheyenne River in relation to the highway, he or she gains more detailed information about where exactly flooding is taking place.

4. Haiti: `Wow! The Earthquake only lasted about 30 seconds.`

    Segment: `Wow! The Earthquake only lasted about 30 seconds.`
    This segment tells the reader the length of time of the 2010 Haiti earthquake, which "lasted only about 30 seconds." The use of "only" indicates that 30 seconds is a short period of time, though the average length of an earthquake is ten to thirty seconds.[27] In addition, the Twitter user begins the segment with

---

[27] http://www.seis.utah.edu/qfacts/genfaq.shtml#gq2

"Wow!" which is an indication that the succeeding information is surprising or unexpected.

This segment is written in the past tense. The use of past tense can be viewed as indicative of the type of disaster event. Earthquakes are not predictable, and last for seconds. It is unlikely that during a high magnitude earthquake like the 2010 Haiti earthquake that people would be providing hazard status information while the hazard is occurring. Instead, information about the hazard agent was provided after impact.

Information Type: Weather

Weather labels tweets that communicate information about weather conditions; specifically, how those conditions may effect the current state, progression or outcome of a mass emergency event.

1. OK09: #OK Fires #MWCFIRES Winds coming back up. gust to 40 MPH now. Fires being refueled.

   Segment: Winds coming back up. gust to 40 MPH now. Fires being refueled.

This segment informs the reader that winds are "coming back up," which implies that winds had died down, but are now gaining strength. The next phrase tells the reader that gusts are "to 40 MPH" which means the highest clocked wind speed is forty miles an hour. The last phrase indicates that fires had died down, but now they are being "refueled," by the winds. The use of "refuel" indicates the fires had died down, but are now growing again.

2. RR09: `@Drifter0658 The river has been receding, but with the melt of snow its possible it could go back up to 40', hope not. That's wearing!`

   Segment: `the melt of snow`

   This segment indicates that weather conditions have warmed up, causing snow to melt. Previous conditions were cold enough to warrant snowfall, which accumulated; enough snow accumulation was present to bring about concern that the melting would cause the river to rise.

3. RR10: `Hayley LaPoint: Weather really helping out the flood worries... cold and no precip. Red River crest little lower now because of this! Tune i`

   Segment: `Weather really helping out the flood worries... cold and no precip.`

   In this segment, the reader learns that weather conditions are positive with regards to the flooding situation. Cold weather is favorable during the spring flooding season, as it keeps accumulated snow from melting too quickly and causing the river to overflow. In addition, there is "no precip." or no precipitation, which is additionally positive information, as precipitation may also cause river levels to rise.

4. Haiti: `@leashless re cold weather, that's not the case in #haiti with 17-31ºC today in capital area; maybe for kids and frail, only`

   Segment: `#haiti with 17-31ºC today in capital area`

   This segment informs the reader that the temperature in the capital city of Port-au-Prince is between 17-31 degrees Celsius, which is 63-88 degrees Fahrenheit. In the aftermath of a catastrophic event such as the Haiti

earthquake, weather conditions have an effect on how survivors cope and how response efforts transpire.

4.2 Part 2: Additional Studies Of and Approaches Toward Mass Emergency

The analysis presented in part 1 describes Twitter behavior of particular samples of data during the four mass emergency situations examined in this dissertation. To do this type of analysis, I carefully selected samples from much larger datasets with the goal of creating datasets that are representative of each emergency situation. However, it is possible that these sampling techniques omitted tweets that contain situational awareness information, or that additional types of information that contribute to situational awareness are not present in my datasets. To investigate where omissions occurred in my analysis of Twitter data, I turn to research from the sociology of disaster, and government procedures that outline approaches to take toward mass emergency. This analysis of additional sources of information regarding human behavior in mass emergency serves as a way to scrutinize the linguistic analysis of Twitter communications, and see how it compares to related resources.

4.2.1 Disaster Processes in the Sociology of Disaster Literature

One perspective regarding how information is organized during mass emergency situations is offered by Dynes (1970), who puts forth a scheme of processes that communities go through when faced with disaster. These processes are: (1) the warning process, (2) preventative action, (3) significant damage, (4) inventory of damage, and (5) organizational mobilization. The information types I uncovered in

150

my inductive analysis of Twitter data are more specific than the processes listed

above, and it is helpful to see where particular information types correlate to Dynes'

scheme.

Twitter - Information Types	Dynes' Social Processes in Disaster
Caution Response – Formal	Warning Process
Evacuation Preparation	Preventative Action
Fatality Injury Missing Status – Community/Population Status - Personal	Significant Damage
Damage Status – Infrastructure Status – Personal Property Status – Public Property	Inventory of Damage
Immediate Recovery Rescue Response – Community	Organizational Mobilization

Table 4.1: Correspondence of Twitter Information Types and Dynes' Social Processes

The warning process involves awareness of the possibility of danger, and

communicating and receiving the message. Warning messages should include two

parts: a description of the event, and an explanation of how that event will impact

people. People need enough information about the predicted hazard to understand

what the hazard is and how it may affect them. The warning should include the

location of the hazard and how much time people have before impact is expected

(Mileti & Peek, 2000). Warning messages also need to take into consideration the

social influences that influence peoples' interpretations of such messages (Dynes,

1970).

Preventative action occurs only in those events that allow for warning or prediction. It involves actions such as boarding windows in the case of high winds, building temporary levees in the case of a flood, and evacuating people and animals if necessary. *Significant damage* refers to the period after a disaster impact, when injuries are assessed, and affected populations make efforts to communicate with family and friends. The *inventory of damage* period occurs when an initial, quick assessment of area damage is made. The *organizational mobilization* process refers specifically to organizational response to a disaster, where organizations are local businesses, media groups, and community associations. Organizations in an area affected by disaster often quickly take on a role to respond to the situation.

Dynes' assessment of how people organize behavior in disaster situations provides a point of reference with which to compare the information types I identified in the first part of this chapter. The information types are more specific, but many are situated within the social processes Dynes outlines; the comparison serves as support for the results of the linguistic analysis of Twitter communications.

4.2.2 Governmental Approaches Toward Mass Emergency

In the United States, the Department of Homeland Security (DHS) is a cabinet department of the federal government. Several agencies, including the Federal Emergency Management Agency (FEMA) are governed by the DHS (Tierney, 2007). In September 2011, the DHS published the National Preparedness Goal (Department of Homeland, 2011), which outlines procedures for how to attend to many types mass emergency situations. The National Preparedness Goal (NPG) details "mission areas," defined as prevention, mitigation, protection, response and recovery of mass

152

emergencies. Each mission area includes "core capabilities" (e.g. "On-scene Security and Protection," and "Fatality Management Services") that generalize across various types of hazards. The purpose of the core capabilities is to outline the critical elements required for a state of "success" as success relates to a secure and resilient nation. The National Preparedness Goal additionally points to the integral role of community participation of all levels (e.g. individuals, private and non-profit sectors, faith-based organizations) in confronting mass emergency situations.

The NPG outlines the range of threats that may cause a mass emergency situation, which include: natural hazards (e.g. flood), pandemic illness (e.g. influenza outbreak), technological and accidental hazards (e.g. chemical substance spills), terrorist attacks (e.g. explosive detonation) and cyber attacks (e.g. financial system failures). This range of threats serves as the foundation for the development of the core capabilities presented in each mission area. In the following sections, I consider core capabilities relate to mass emergencies brought about due to the impact of a natural hazard agent from two mission areas: *Protection* and *Response*, as these are the two mission areas that correspond to the emergency phase of the disasters examined in this dissertation. Since government sources regarding mass emergency policy have an all-hazards focus, many of the procedures that are in place focus on acts of terrorism, the threat of weapons of mass destruction or similar events. These procedures are not considered for this research, as this dissertation focuses only on mass emergencies caused by natural hazards.

4.2.2.1 Comparing Governmental Approaches to Mass Emergency and Microblogged Communications

Regarding core capabilities in the *protection* mission area, the National

Preparedness Goal (Department of Homeland, 2011) states:

> Protection includes capabilities to safeguard the homeland against acts of terrorism and man-made or natural disasters. It is focused on actions to protect the citizens, residents, visitors, and critical assets, systems, and networks against the greatest risks to our Nation in a manner that allows our interests, aspirations, and way of life to thrive.

The National Preparedness Goal defines the core capabilities regarding *response* as:

> [T]hose capabilities necessary to save lives, protect property and the environment, and meet basic human needs after an incident has occurred.

Each of these mission areas includes core capabilities that describe actions to take in

relation to the goals of the mission area, which encompass approaches to take toward

all types of mass emergencies. In my analysis of the National Preparedness Goal, I

identified descriptions of actions to take which correspond to the information types I

identified in the first part of this chapter. In addition, I also identified descriptions of

actions which I did not identify in my analysis of Twitter communications; these

descriptions serve to point out the limits of my empirical analysis.

4.2.2.1.1 Government Procedures and Twitter Communications: Corresponding Categories of Information

Many of the information types that lead to situational awareness identified in

my analysis of tweets correspond with the information in manuals and guidelines

concerning mass emergency put forth by the United Sates government. These

information types correlate to core capabilities in the *protection* and *response* mission

areas in the National Preparedness Goal, to working drafts of the National Protection

Framework (Presidential Policy, 2012a); and to the National Response Framework (Presidential Policy, 2012b). The correspondence of information types to government approaches serves as support for the information types I identified in my analysis of Twitter communications.

The information types listed in the left column of Table 4.2 correspond to the core capabilities in the protection and response mission areas in the right column. The core capabilities from the protection mission area are: *Public Information and Warning*, which involves the delivery of reliable, actionable information to populations under threat of disaster; *Intelligence and Information Sharing,* which involves providing relevant, timely information about the disaster agent and the impact of the disaster on human populations; and *Physical Protective Measures,* which involves reducing risks that may occur due to a disaster agent by implementing physical protections. The core capabilities from the response mission area are: *Critical Transportation,* which involves evacuation of people and animals; *Fatality Management Services,* which includes body recovery and victim identification; *Mass Care Services*, which involves delivering resource and sheltering to those in need; *Mass Search and Rescue Operations*, which involves finding and rescuing missing persons, and *Public and Private Services and Resources*, which involves the mobilization of governmental, non-governmental, and private sector resources both near and external to the affected region to assess the situation post-impact, meet basic needs and stabilize the area.

Twitter Information Types	Government Core Capabilities
Caution	Public Information and Warning
Caution Historical Prediction Response – Community Response – Formal Response – Miscellaneous Response – Personal Status – Community/Population Status – Hazard Status – Personal Weather	Intelligence and Information Sharing
Preparation Response – Community Response – Formal Response – Miscellaneous Response – Personal	Physical Protective Measures
Animal Management Evacuation	Critical Transportation
Fatality	Fatality Management Services
Feeding/Hydration Injury Medical Attention Sheltering	Mass Care Services
Missing Rescue	Mass Search and Rescue Operations
Damage Immediate Recovery Report of Crime Response – Community Response – Formal Response – Miscellaneous Response – Personal Status – Community/Population Status – Personal Status – Personal Property Status – Public Property	Public and Private Services and Resources

Table 4.2: Correspondence of Twitter Information Types and Government Procedures

4.2.2.1.2 Information in Government Procedures Not Included in Twitter Communication Analysis

Government resources have also helped to further refine the definitions of the

information types I explain in section 4.1, and have provided indications of

information types that are not present in the four datasets I examine, or that I failed to

identify. Information presented in government documents which indicate information that may be communicated during mass emergency situations via Twitter are:

Environmental Impact
Feeding/Hydration
Medical Attention

Environmental Impact describes information about environmental health and safety actions that should be performed in the aftermath of a disaster impact, such as how to deal with hazardous materials, air quality, waste management, and similar issues. *Feeding/Hydration* describes the provision of nutritional resources to those in need in the aftermath of disaster impact. *Medical Attention* describes lifesaving actions taken via emergency medical services, as well as actions taken to ensure overall public health and wellness in a mass emergency situation.

The categories of information presented above expand the list of information types that I identified in my linguistic analysis of Twitter data. Since these information types were not a part of the schema used to identify information communicated during the four mass emergencies examined in this dissertation and presented in the following chapter, I returned to each dataset to determine if these information types detailed above were in fact communicated in the datasets.

There were no tweets that indicated *Environmental Impact* information. However, it is reasonable to expect that such information may have been communicated and was not included in the linguistic analysis of the datasets; or that this type of information was not communicated by the Twitter users whose tweets were examined for this research. Given that environmental impact in disaster situations does occur, and can have effects such as water contamination during floods

and air pollution during fires,[28] it is logical to expect that this type of information is of interest to affected populations. Considering the attention paid to environmental concerns by government and environmental agencies, *Environmental Impact* should be included in the list of information types that lead to situational awareness during mass emergency.

The *Feeding/Hydration* category was represented in the datasets I examined. Eight tweets from the OK09 dataset mention food distribution or supplying food for victims of the fire. In the Haiti dataset, thirty-eight tweets included information about hydration and food. Based on these findings, it is apparent that an additional information type of *Feeding/Hydration* is necessary in the information type schema.

In my additional examination of the datasets used for this dissertation, I also uncovered five tweets in the RR09 dataset, and thirty-five tweets in the Haiti dataset that mention medical care or medical situations. It is apparent that a *Medical Attention* category is also necessary to add to the information type schema.

4.2.2.1.3 Information in Twitter Communication Analysis Not Included in Government Procedures

There are also information types identified in the analysis of Twitter communications that are not implied in government guidelines, which are:

Advice – Information Space
General Area Information
General Hazard Information
General Population Information
Offer of Help
Request for Help
Request for Information

---

[28] http://www.epa.gov/naturalevents/wildfires/index.html#prepare, http://www.epa.gov/agriculture/tned.html

The absence of these types of information in governmental frameworks and resources may indicate a number of points, which I illustrate by considering the above information types in two groups: one of "general" information, and one of speech acts, specifically "directives" and "commissives" (Searle, 1975).

The information types *General Area Information, General Hazard Information* and *General Population Information* all describe information that relates to a mass emergency event, but which provide the reader with common, established knowledge. Government guidelines do not mention general information in their procedures and manuals for dealing with mass emergency situations, which may indicate these types of information are assumed or previously known, or that general information is not taken into account during mass emergency situations.

The other four information types not presented in government documentation can be described as *directives* and *commissives*, which are types of illocutionary *speech acts*. A speech act involves the use of language to *do* something. For example, when a minister performs a marriage, and states "I now pronounce you married," in saying this, he or she is also performing an action, in this example, the act of marriage. When Twitter users communicate the information types *Advice – Information Space, Request for Help* and *Request for Information,* we can think of these as directives, a form of speech act that are intended to cause hearers or readers to take a particular action. Examples of directive speech acts are: advising, asking, instructing and suggesting (Bach, 1998). Many of the tweets labeled with one of these three information types contain directive speech acts, as the tweet text works to advise or request. The *Offer of Help* information type is a form of commissive, which

is a speech act that commits the speaker or writer to some action. Examples of commissive speech acts are: agreeing, inviting and offering (Bach, 1998). Those tweets labeled with *Offer of Help* contain information that either commits the tweet author to some action, or relays information that commits another party to some action.

I turn to speech acts as a way to consider what happens on Twitter during disasters, and how Twitter activity diverges from conventional actions taken during mass emergency situations. The analysis of Twitter communications presented in this chapter reveal what types of information people communicate that can contribute to situational awareness, it also reveals how Twitter users harness the capabilities of Twitter to perform certain actions through language. Many Twitter communications that take place during mass emergencies describe a state of affairs, e.g. "Valley City evacuating half the town," or "New fires are popping up at Hiwasee Rd." However, those information types that fall into the directive and commissive speech act categories reveal that in addition to stating information that government agencies also attend to, Twitter users additionally use the microblogging service to ask questions, ask for goods and services, and extend offers of goods and services. The information types *Advice – Information Space, Offer of Help, Request for Help* and *Request for Information* point to ways in which linguistic behavior on Twitter goes beyond stating and describing the state of affairs during disaster, and works in ways intended to commit writers and readers to particular actions in relation to the mass emergency situation.

4.2.3 Discourse Analysis of Twitter Communications, the Sociology of Disaster, and Government Documentation

The comparison of the information types I uncovered in my analysis of Twitter communications to sociological research on disaster situations, and to governmental procedures, serves as a way to gauge the validity of my ground-up, inductive analysis. I found that many of the information types I uncovered correspond to the ways the Department of Homeland Security (DHS) organizes their approach toward disaster situations, and that my analysis failed to uncover some information types detailed by the DHS. I also found that the processes described by Dynes (1970) regarding information organization during disaster at the community level correlate to my information type categories. These comparisons provide a picture of the different ways in which information can be organized in a mass emergency situations, as well as different procedures for assessing the actions people take during such events.

4.2.4 Summary

The analysis presented in section 4.1 explains the information Twitter users conveyed during four mass emergency situations. The description of additional types of information presented in section 4.2 presents further information types that were not identified through the analysis Twitter communications, but which contribute to situational awareness. This chapter provides an overview of specific information that can lead to situational awareness through the inductive analysis of microblogged communications, as well as through considering government policies and procedures and the sociology of disaster literature. Understanding what information is necessary to have in times of mass emergency is beneficial toward contributions to situational

awareness. In the following chapter, I discuss the application of those information

types identified through the analysis of Twitter communications through the

performance of a qualitative coding process.

Chapter 5

Application of Linguistic Behavior Codes

This chapter provides an explanation of the qualitative coding process. This phase of research determined what information Twitter users communicate during mass emergency situations in three separate coding passes, which are described in the following sections.

5.1 First Pass Coding: Coding for On- or Off-Topic Tweets

The initial step in the qualitative coding process is for annotators to assign a first pass code to each tweet in every dataset.[29] The purpose of this phase is to determine if Twitterers are communicating information that contributes to situational awareness. Three mutually exclusive first pass codes may be assigned to each tweet in each dataset: O, R or N. The definition and examples which illustrate each first pass code follow. For the following examples, and examples throughout this chapter, the order of emergency events is 2009 Oklahoma Fires (abbreviated "OK09"), 2009 Red River Flood (abbreviated "RR09"), 2010 Red River Flood (abbreviated "RR10") and 2010 Haiti Earthquake (abbreviated "Haiti").

O: Off-topic

>   Definition: "O" tweets do not contain any information that mentions or relates to the emergency event.

---

[29] The annotators are employed by Project EPIC, and both have a background in linguistics: Jenette Preciado ("Preciado") and Sarah Vollmann ("Vollmann").

Examples:

1. OK09: `Think I'm going to knit while watching FernGully. Such a great message, loved this movie as a kid`

2. RR09: `Only the rest of today and tomorrow and then I can finally check facebook!!!`

3. RR10: `Headed to town soon to have a get together with a bunch of the local amateur radio operators, will get a cup of coffee while I'm at it`

4. Haiti: `These guys r so useless lol -- Terrell Owens & Ludacris Judge A Booty Shaking Contest In Miami http://youtu.be/XB4aLwx5CXY`

R: On-topic, relevant to situational awareness

Definition: "R" tweets are relevant to situational awareness; they contain information that provides tactical, actionable information that can aid people in making decisions, advise others on how to obtain specific information from various sources, or offer immediate post-impact help to those affected by the mass emergency.

Examples:

1. OK09: `Velma OK: traffic restricted/all persons must report in at Town Hall.  Residents are asked to seek shelter outside of the area. #OKFires`

2. RR09: `Highway 18 closed in Wyndmere, water over I-94 @Buff-Alice exit. Lots of roads with water over 1 or 2 lanes in southeast ND and west MN`

3. RR10: `Wow, this is so weird. The sandbags and dikes are in place and now we just wait. Never been this far ahead of the game before.`

4. Haiti: `If you have family in Haiti, the US Govt set up a number for u to call & they will help locate ur family 888 407 4747 (Via @GlobalGrind)`

N: On-topic, irrelevant to situational awareness

Definition: "N" tweets are on-topic, but do not contain information that contributes to situational awareness. These tweet are on-topic because they mention the emergency by including offers of prayer and support in relation to the emergency, solicitations for donations to charities, or casual reference to the emergency event.

1. OK09: `RT @elmofromok: watching oklahoma burn is surreal #oklahomaburning`

2. RR09: `And I thank my friends at the Flag 1100 AM for the opportunity to be on their station during our flood crisis. Nice Studios, great toys!`

3. RR10: `Floods are good for business. About 30 National Guard troops just showed up at the bar.`

4. Haiti: `New blog post coming soon. In the meantime, if you're looking to give directly towards helping people in Haiti go here: http://bit.ly/5n6PK6`

5.1.1 First Pass Coding Kappa Calculations

After annotators assign first pass codes, I calculate their agreement rate using the kappa calculation (Cohen, 1960), because it takes into account the possibility that annotators agree by chance. The kappa values for the first pass coding of each dataset are listed in Tables 5.1-5.4. For the first two datasets (OK09 and RR09), I coded the data with the annotators. Once they gained experience and expertise with the coding process, I acted solely as the adjudicator, i.e. I made a final decision for those tweets on which annotators disagreed.

Annotators	OK09 1st Pass Kappa Values
Vieweg and Vollmann	92.7%
Vieweg and Preciado	90.1%
Preciado and Vollmann	87.3%

Table 5.1: OK09 1st Pass Kappa Values

Annotators	RR09 1st Pass Kappa Values
Vieweg and Vollmann	90.2%
Vieweg and Preciado	90.0%
Preciado and Vollmann	91.6%

Table 5.2: RR09 1st Pass Kappa Values

Annotators	RR10 1st Pass Kappa Value
Preciado and Vollmann	94.0%

Table 5.3: RR10 1st Pass Kappa Values

Annotators	Haiti 1st Pass Kappa Value
Preciado and Vollmann	57.1%

Table 5.4: Haiti 1st Pass Kappa Values

165

The kappa calculations for the OK09, RR09 and RR10 datasets are all above 80%, which is considered good (Manning et al., 2008). The kappa value for the Haiti dataset is lower, which indicates that annotators had difficulty coming to agreement. After meeting to discuss the Haiti dataset, it became clear that one annotator was coding tweets that contained information about fundraising and request for money for Haiti relief efforts as on-topic, and relevant to situational awareness ("R"), while the other was coding these same tweets as on-topic, but not containing information relevant to situational awareness ("N"). I reiterated that such tweets should be coded as on-topic, but not containing information relevant to situational awareness (i.e. with "N").

5.1.2 First Pass Coding Results

The following sections provide a description of the first pass coding results for each dataset. I provide a table and graph to illustrate results for the OK09, RR09, RR10 and Haiti datasets.

5.1.2.1 2009 Oklahoma Fires First Pass Coding Results

The OK09 dataset contains a total of 2766 tweets. The results of the first pass of coding are illustrated in Table 5.5 and Figure 5.1. Of the entire dataset, 535 tweets, or 19% of the dataset, contain information relevant to situational awareness. Figure 1 displays a chart of the OK09 first pass coding results; for this and all other figures in the following sections, the category type is on the X axis, and the number of tweets on the Y axis. Tweets labeled with "O" are off-topic, tweets labeled with "R" are on-topic, and include information relevant to situational awareness (SA), and tweets

labeled with "N" are on-topic, but do not contain any information relevant to

situational awareness.

OK09 1st Pass Category	Number of Tweets	Percentage of Dataset
Off-topic (O)	2064	74.6%
On-topic, relevant to SA (R)	535	20.0%
On-topic, not relevant to SA (N)	167	5.4%

Table 5.5: OK09 1st Pass Tweet Totals
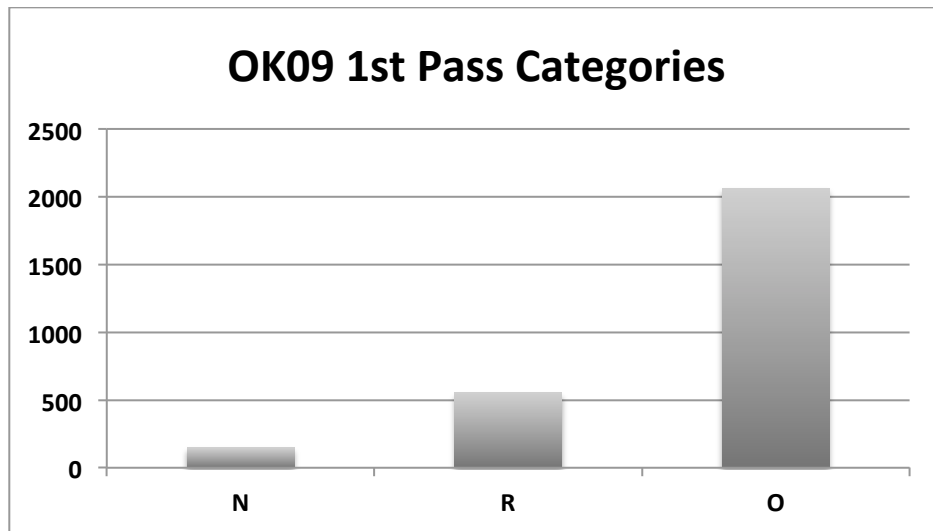
Figure 5.1: OK09 1st Pass Categories

5.1.2.2 2009 Red River Flood First Pass Coding Results

The RR09 dataset includes a total of 19,152 tweets. The results of the first

pass of coding are displayed in Table 5.6. Of the entire dataset, there are 1483 tweets

which contain information that may lead to situational awareness, which is 7.7% of

the dataset. Below Table 5.6 is Figure 5.2, which provides a visual representation of

the number of tweets that fall into each first pass category.

RR09 1st Pass Category	Number of Tweets	Percentage of Dataset
Off-topic (O)	17,107	89.3%
On-topic, relevant to SA (R)	1483	7.8%
On-topic, not relevant to SA (N)	562	2.9%

Table 5.6: RR09 1st Pass Tweet Totals
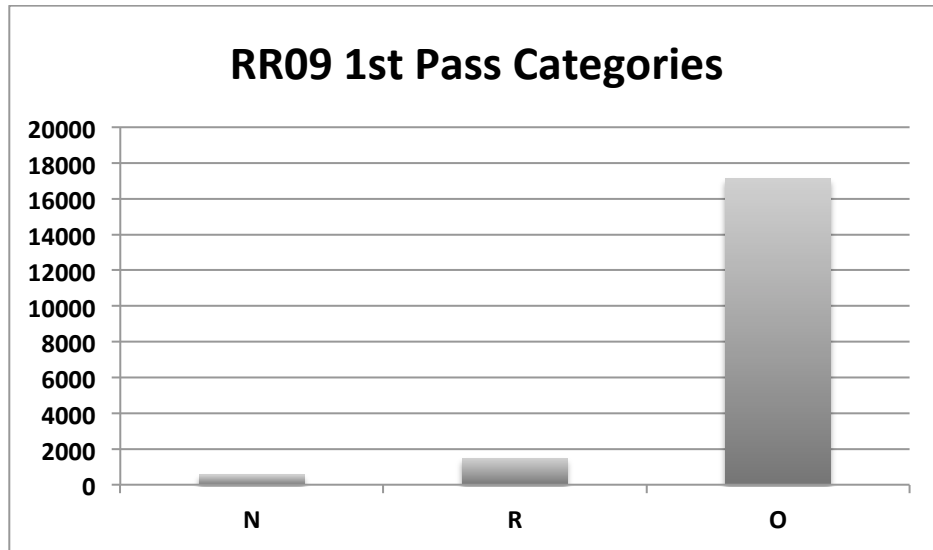
Figure 5.2: RR09 1st Pass Categories

5.1.2.3 2010 Red River Flood First Pass Coding Results

There are 11,879 tweets in the RR10 dataset. Of these, the first pass of coding

shows that 9.2% (1091 tweets) contain information that may contribute to situational

awareness. Table 5.7 provides the breakdown of off-topic, on-topic tweets that

include information relevant to situational awareness (SA), and on-topic tweets that

lack situational awareness information, as well as the percentage of the dataset

represented by each type of tweet. Figure 5.3 provides a chart displaying the

information found in Table 5.7.

RR10 1ˢᵗ Pass Category	Number of Tweets	Percentage of Dataset
Off-topic (O)	10,610	89.3%
On-topic, relevant to SA (R)	1091	9.2%
On-topic, not relevant to SA (N)	178	1.5%

Table 5.7: RR10 1ˢᵗ Pass Tweet Totals
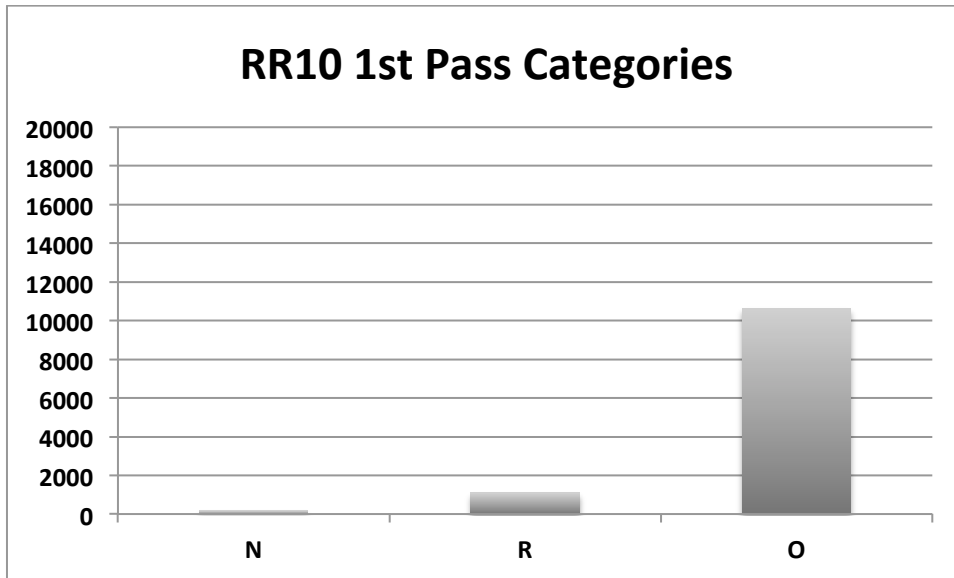
RR10 1st Pass Categories

5.1.2.4 2010 Haiti Earthquake First Pass Coding Results

Annotators started the first pass coding process for the Haiti dataset with a total of 4005 tweets that were arbitrarily selected from the larger Haiti dataset, which resulted in a total of 943 tweets that contain information relevant to situational awareness. The results of the first pass of coding are provided in Table 5.8 and Figure 5.4 below.

Haiti 1st Pass Category	Number of Tweets	Percentage of Dataset
On-topic, not relevant to SA (N)	1883	47.0%
Off-topic (O)	1179	29.5%
On-topic, relevant to SA (R)	943	23.5%

Table 5.8: Haiti 1st Pass Tweet Totals 1
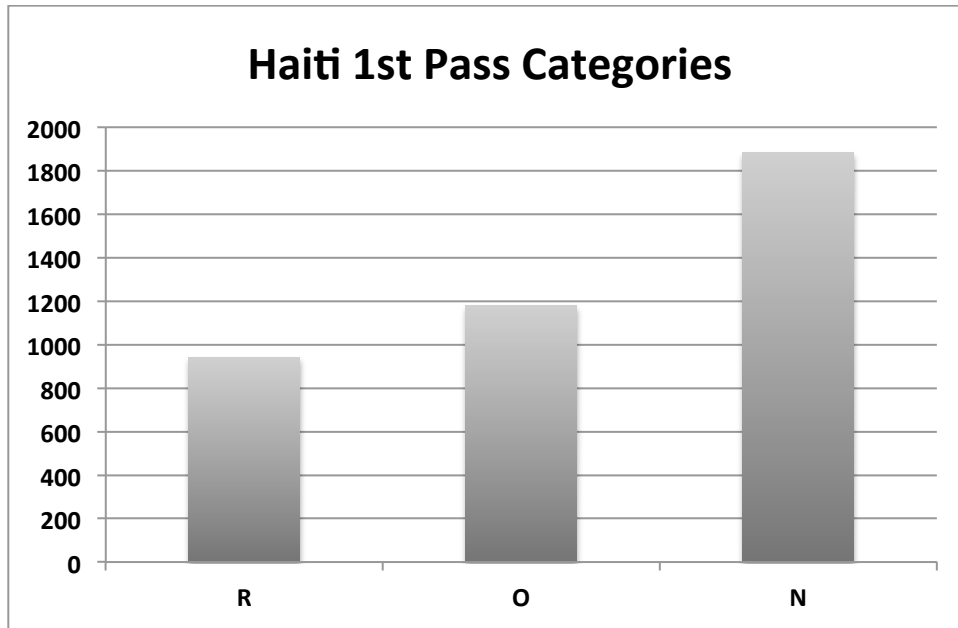
Figure 5.4: Haiti 1st Pass Categories

5.1.3 First Pass Coding Outcomes

The results of the first pass of coding for each dataset are shown in Table 5.9. This table compares the proportion of tweets that were off-topic ('O' tweets), on-topic and relevant to situational awareness ('R' tweets), and on-topic but containing no information relevant to situational awareness ('N' tweets).

	OK09	RR09	RR10	Haiti
'R' Percentage	20.0%	7.8%	9.2%	23.5%
'N' Percentage	5.4%	2.9%	1.5%	47.0%
'O' Percentage	74.6%	89.3%	89.3%	29.5%

Table 5.9: 1st Pass Code Percentages per Dataset

The OK09 and RR09 datasets were originally sampled for previous research for which my colleagues and I were interested in finding those Twitter users who were tweeting "from the ground," as we surmised that those users would be likely to communicate first-hand knowledge of the events. The motivation to understand how affected populations use Twitter during mass emergency is what led to the samples

for the OK09 and RR09 datasets used in this dissertation; these same sampling decisions also produced datasets which contain situational awareness information.

The RR10 dataset was sampled similarly to the OK09 and RR09 datasets, with some exceptions. For this dataset, I was focused not on finding Twitter users who were on the ground, but on tweets that contained situational awareness information regardless of the tweet author's location. It is likely due to the focalized nature of the RR10 event, and its similarity to the RR09 event, that my sampling decisions resulted in a dataset with similar 'R', 'N' and 'O' distributions as the previous two datasets.

The majority of tweets in the OK09, RR09 and RR10 datasets are off-topic. This is because these datasets include contextual streams for each user, i.e. every tweet each user sent during the data collection period. The data collection period for the OK09, RR09 and RR10 datasets spans the emergency period for each event, but during the Oklahoma fires and the Red River floods, Twitter users in each dataset mostly tweeted about topics that had nothing to do with the particular emergency.

The Haiti dataset was sampled differently from the other three datasets and does not include users' contextual streams. Instead, it includes only tweets that include a search term, but do not include any of the stop words listed in section 3.2.3.2.1 of chapter 3. The tweets in the Haiti dataset were also automatically classified using Verma's situational awareness classifier (Verma et al., 2011). These sampling decisions led to a dataset that includes a higher percentage of on-topic tweets than the other three datasets. Yet the results of the first pass coding show that most of the tweets in the Haiti dataset, while on-topic, do not contain information

171

relevant to situational awareness. However, this is an encouraging result, as there is a higher percentage of 'R' tweets in the Haiti dataset than all other datasets, which indicates that the sampling methods used for this dataset are promising.

It is also possible that the larger percentage of on-topic tweets that do not contain situational awareness information in the Haiti dataset as compared to the other three datasets is due to disaster type. The Haiti earthquake caused an international response and many requests and offers of financial donations were made. While information about donations are not coded as relevant to situational awareness for catastrophic events like the Haiti earthquake, many tweets were broadcast regarding fundraising and donations, and many of those tweets ended up in the Haiti dataset used for this dissertation.

5.2 Second Pass Coding: Coding for the Social, Built and/or Physical Environment

After first pass coding was complete on the first few datasets, it became clear that an intermediate pass of annotation was necessary—a level between the first level and the assignment of specific information types. My decision to add this level of annotation was made after attempting to organize a training session for the annotators. At that point, I had independently developed and defined twenty-five specific information type categories through discourse analysis of the OK09 and RR09 datasets. This amount of coding category possibilities was overwhelming for annotators, and they found it very difficult to assign specific information types at this point in the annotation process. Therefore, I grouped the information type categories into three second pass categories with the aim of making the annotation process

clearer, since the conscientious design of coding schemas is an integral part of an annotation task (Bayerl et al., 2003).

I created the second pass categories based on the twenty-five information types which had been identified at that point. This process resulted in three categories: *Social Environment*, *Built Environment*, and *Physical Environment*.[30] I provided annotators with a training guide that defined each second pass category, and also listed the corresponding information type categories. This approach provided annotators with a way to see the detailed information types they would eventually be assigning to tweets. Second pass annotation allows for further familiarity with the data among annotators, as well as opportunity for discussion and mutual understanding of category definition at all levels of annotation. Definitions and examples of the second pass categories are below. It is important to note that in the second pass of annotation, tweets may be labeled with one, two or all three codes.

S: Social Environment

Definition: These tweets include information about how people and/or animals are affected by a hazard, questions asked in relation to the hazard, responses to the hazard and actions to take that directly relate to the hazard and the emergency situation it causes. These tweets all include description of a human element in that they explain or display human behavior.

Examples of "S" tweets:

1. OK09: `RT @dorman340 is watching the news 12 injuries reported due to the fires. Will be requesting the Gov declare disaster areas tomorrow #okfire`

2. RR09: `Go away water  RT @ctvwinnipeg: City of Winnipeg declares State of Emergency http://tinyurl.com/dalc5o`

---

[30] The second pass coincide with the three systems that Dennis Mileti refers to when describing disaster situations: the earth's physical system, human systems, and constructed systems (Mileti, 1999, p. 107).

3. RR10: `Absaraka, ND, on alert for possible evacuation: The`
   `Absaraka Dam north of that North Dakota town has people who`
   `li... http://bit.ly/a1Lasr.`

4. Haiti: `FEMA NYTF-1 Urban Search and Rescue team made up of`
   `members from the NYPD & FDNY are in their way to Haiti. Go`
   `(cont) http://tl.gd/3lrol`

B: Built Environment

Definition: Tweets that include information about the effect of the hazard on the built environment, including updates on the state of infrastructure, such as road closures or bridge outages, damage to property, lack of damage to property and the overall state or condition of structures.

Examples of "B" tweets:

1. OK09: `OK Emergency Mgt reports at least 140 buildings`
   `destroyed, not counting many in Carter and Payne Counties.`
   `More info: http://bit.ly/13YKwG`

2. RR09: `Breach sends water onto school campus in Fargo`
   `http://bit.ly/yBlfO`

3. RR10: `Portion of oak street closed in north Fargo. Elm`
   `Street by 14th Ave closed and flooded over.`

4. Haiti: `UN headquarters in Haiti collapsed in quake`
   `(AP)http://bit.ly/6dMTWB`

P: Physical Environment

Definition: Tweets that contain specific information about the hazard including particular locations of the hazard agent or where the hazard agent is expected or predicted to travel or predicted states of the hazard agent going forward, notes about past hazards that compare to the current hazard, and how weather may affect hazard conditions.[31] These tweets additionally include information about the type of hazard in general (e.g. "this is what happens during a flood" or "wildfires often have these effects"). This category also subsumes any general information about the area under threat or in the midst of an emergency, e.g. "the land is flat and dry, so fires can easily spread."

Examples of "P" tweets:

---

[31] Annotators engaged in lengthy discussion regarding how to define the 'P' category as it relates to the location of a hazard agent. Tweets that mention hazard location in the most general sense, i.e. "fire in Midwest City, Oklahoma" or "flood in Fargo" are not labeled with 'P.'

1. OK09: `RT @rustysurette: BIG fire south of downtown OKC. Thick smoke covering the metro.  It's a grass fire in Newcastle.`

2. RR09: `RT @newsbbq: New blog post: Red River Ice is Finally Breaking Up http://tinyurl.com/c5jy3t`

3. RR10: `RT @redriveratfargo Red River @ Fargo is 23.61'. 5.61' above flood stage. -17.21' record crest. Up 0.53' from last read. #fargo #nd`

4. Haiti: `PORT-AU-PRINCE, Haiti -The largest earthquake ever recorded in the area shook Haiti on Tuesday, collapsing a ho.. http://bit.ly/6U6EMy`

5.2.1 Second Pass Coding Kappa Calculations

The kappa calculation for the second pass of coding involves calculating agreement based on category type; these results are shown in in Tables 5.10-5.13 For example, if Annotator X codes a tweet with S and P, and Annotator Y does the same, then for category S we have 100% agreement, for category P we have 100% agreement, and for category B we have 100% agreement. However, if Annotator X codes another tweet with P and Annotator Y codes the same tweet with S and P, then we have 0% agreement for category S, 100% agreement for category P, and 100% agreement for category B. The second pass kappa values for each dataset are listed below. Again, for the first two datasets, I annotated the data along with the external annotators.

Annotators	OK09 2nd Pass Kappa Values		
	S	*B*	*P*
Vieweg and Vollmann	81.1%	89.5%	76.3%
Vieweg and Preciado	81.4%	92.0%	76.2%
Preciado and Vollmann	77.9%	88.9%	76.9%

Table 5.10: OK09 2nd Pass Kappa Values

Annotators	RR09 2nd Pass Kappa Values		
	S	*B*	*P*
Vieweg and Vollmann	83.6%	91.9%	85.1%
Vieweg and Preciado	81.6%	91.2%	86.9%
Preciado and Vollmann	79.0%	90.0%	85.8%

Table 5.11: RR09 2nd Pass Kappa Values

Annotators	RR10 2nd Pass Kappa Values		
	S	*B*	*P*
Preciado and Vollmann	88.5%	92.7%	90.6%

Table 5.12: RR10 2nd Pass Kappa Values

Annotators	Haiti 2nd Pass Kappa Values		
	S	*B*	*P*
Preciado and Vollmann	93.3%	90.1%	90.8%

Table 5.13: Haiti 2nd Pass Kappa Values

Kappa calculations for all datasets and all second-pass categories were either above 80%, which is considered good agreement, or between 67-80%, which is considered fair agreement (Manning et al., 2008).

5.2.2 Second Pass Coding Results

The following sections describe the results of the second pass of coding. The outcomes for each dataset are explained, and a table and graph provide further illustration of the coding results for each event.

5.2.2.1 2009 Oklahoma Fires Second Pass Coding Results

After the first pass of coding, the 535 tweets on the OK09 dataset that contain information relevant to situational awareness were then coded with second pass categories; tweets can be labeled with one, two or all three second pass categories.

The results of the second coding pass are shown in Table 5.14. Tweets labeled with

"S" include information about the social environment, tweets labeled with "B"

include information about the built environment, and tweets labeled with "P" include

information about the physical environment.

OK09 2nd Pass Category	Number of Tweets	Percentage of all 'R' Tweets
Social Environment (S)	198	37.0%
Physical Environment (P)	101	18.9%
Social and Physical Environment (SP)	82	15.3%
Built Environment (B)	73	13.7%
Social and Built Environment (SB)	33	8.9%
Built and Physical Environment (BP)	32	6.2%
Social, Built and Physical Environment (SBP)	16	3.0%

Table 5.14: OK09 2nd Pass Tweet Totals

To visualize the number of tweets that fall into each second pass category, below is a

chart which depicts the information in the above table. Figure 5.5 shows the second

pass category on the X axis, and the number of tweets coded with that information

category on the Y axis.
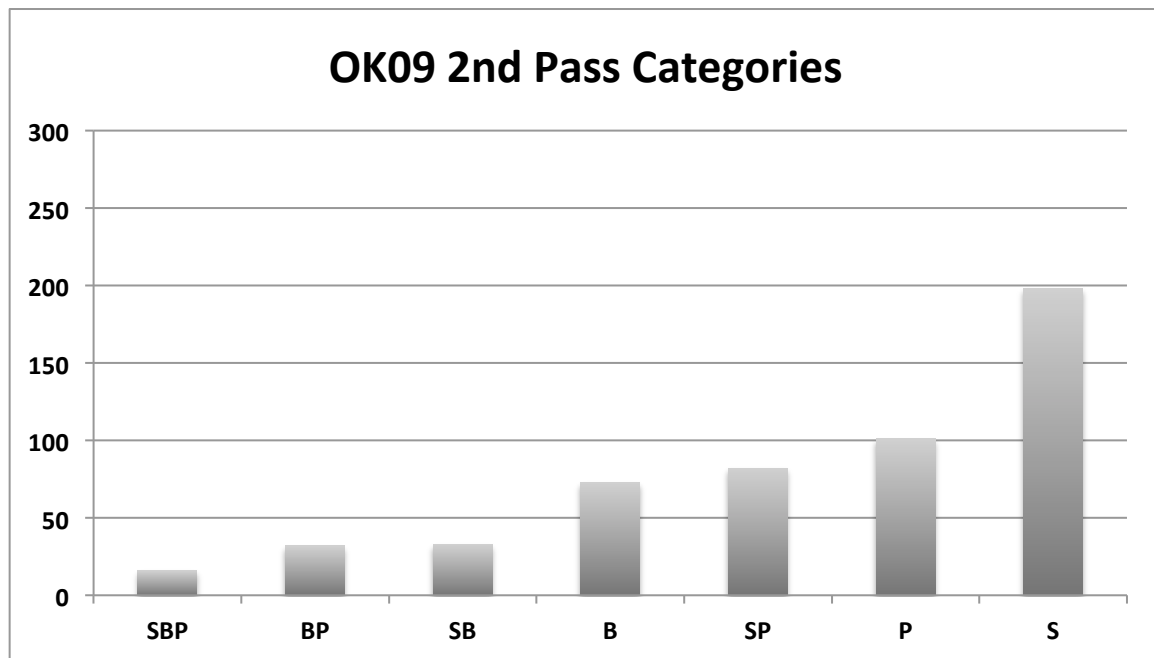
Figure 5.5: OK09 2nd pass categories

177

5.2.2.2 2009 Red River Flood Second Pass Coding Results

The second pass codes for the 1483 tweets in the 2009 Red River flood dataset

that contain information relevant to situational awareness are shown in Table 5.15

and are visually depicted in Figure 5.6.

RR09 2nd Pass Category	Number of Tweets	Percentage of all 'R' Tweets
Social Environment (S)	674	45.4%
Physical Environment (P)	426	28.7%
Social and Physical Environment (SP)	208	14.0%
Social and Built Environment (SB)	68	4.6%
Built Environment (B)	63	4.3%
Built and Physical Environment (BP)	34	2.3%
Social, Built and Physical Environment (SBP)	10	0.7%

Table 5.15: RR09 2nd Pass Tweet Totals
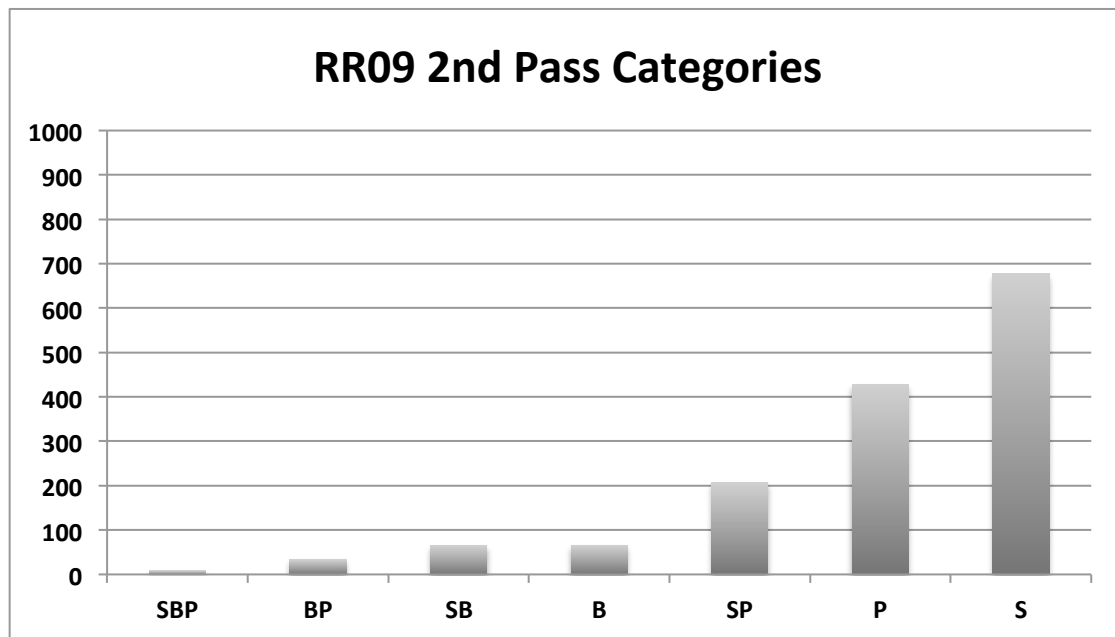
Figure 5.6: RR09 2nd Pass Categories

5.2.2.3 2010 Red River Flood Second Pass Coding Results

The second pass of coding for the RR10 dataset reveals that like the previous

two datasets, most tweets contain information relevant to the social environment, or

are "S" tweets. The results of the second pass of coding are detailed in Table 5.16 and

displayed in Figure 5.7.

RR10 2nd Pass Category	Number of Tweets	Percentage of all 'R' Tweets
Social Environment (S)	596	54.6%
Physical Environment (P)	347	31.8%
Social and Built Environment (SB)	52	4.8%
Social and Physical Environment (SP)	47	4.3%
Built Environment (B)	38	3.5%
Built and Physical Environment (BP)	7	0.6%
Social, Built and Physical Environment (SBP)	4	0.4%

Table 5.16: RR10 2nd Pass Tweet Totals
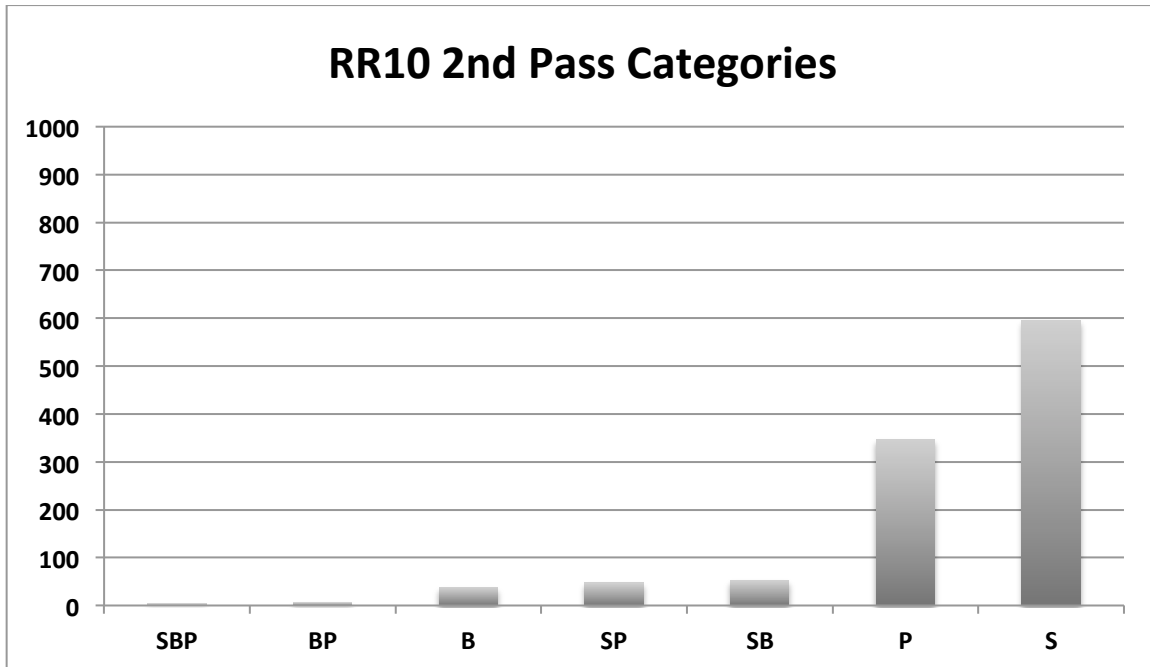
Figure 5.7: RR10 2nd Pass Categories

5.2.2.4 2010 Haiti Earthquake Second Pass Coding Results

The second pass of coding on the 943 R tweets in the Haiti dataset revealed

that like the other datasets, most tweets were social environment, or "S" tweets,

though the disparity between "S" tweets and the other second pass tweets is much

greater in the Haiti dataset. This is likely due to the nature of the Haiti earthquake;

those tweeting about the earthquake situation were primarily concerned with directing

others to useful information, formal response efforts, and the status of the Haitian

people. The status of the built and physical environments in this tragic earthquake

were not communicated about as often, possibly because there was not as much

179

information available, and likely because Twitter users were primarily concerned with the human situation in Haiti. The second pass coding results are listed in Table 5.17, and depicted in Figure 5.8.

Haiti 2nd Pass Category	Number of Tweets	Percentage of all 'R' Tweets
Social Environment (S)	789	83.7%
Social and Built Environment (SB)	63	6.7%
Social and Physical Environment (SP)	28	3.0%
Physical Environment (P)	28	3.0%
Built Environment (B)	24	2.5%
Built and Physical Environment (BP)	7	0.7%
Social, Built and Physical Environment (SBP)	4	0.4%

Table 5.17: Haiti 2nd Pass Tweet Totals
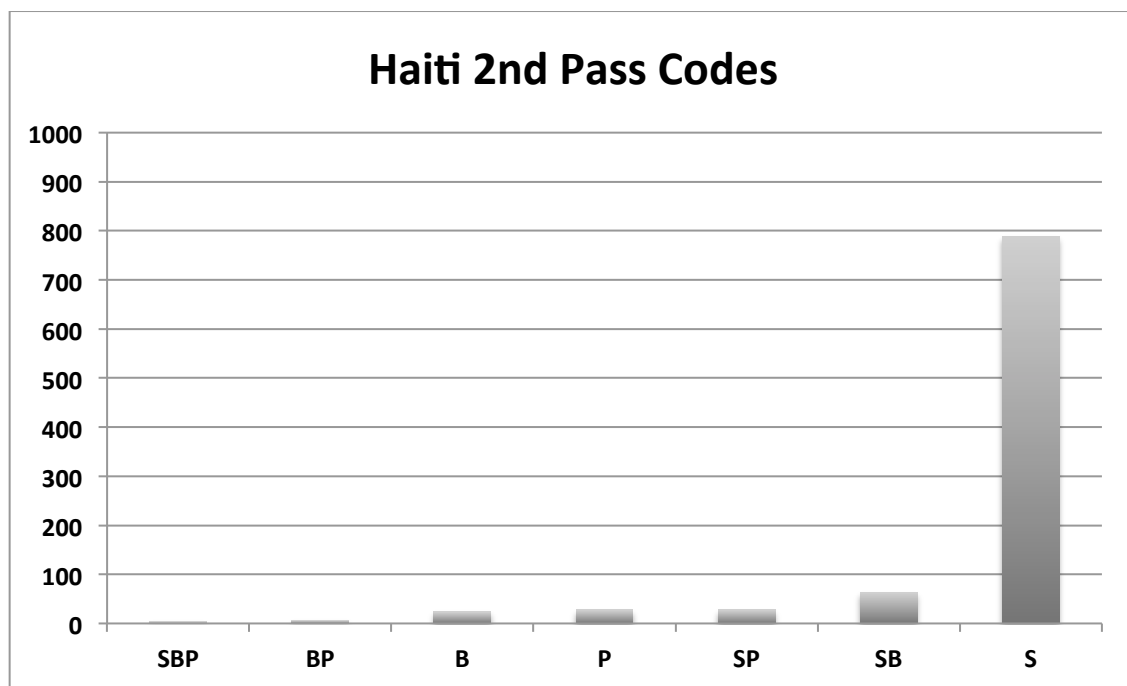
Figure 5.8: Haiti 2nd Pass Categories

5.2.3 Second Pass Coding Outcomes

The results of the second pass of coding begin to tell the story of each emergency event as it unfolded on Twitter; Table 5.18 provides the proportion of each second pass tweet type for all tweets that contain situational awareness information.

	OK09	RR09	RR10	Haiti
'S' Percentage	36.9%	45.6%	54.6%	83.7%
'SB' Percentage	8.9%	4.3%	4.8%	6.7%
'SP' Percentage	14.1%	14.0%	4.3%	3.0%
'B' Percentage	13.7%	4.4%	3.5%	2.5%
'P' Percentage	18.1%	28.8%	31.8%	3.0%
'BP' Percentage	5.2%	2.3%	0.6%	0.7%
'SBP' Percentage	3.1%	0.6%	0.4%	0.4%

Table 5.18: 2ⁿᵈ Pass Code Percentage of all 'R' Tweets

In each dataset, the social environment ("S") category is most common. Disasters are social events; in each disaster studied in this dissertation, the disaster occurred because a natural hazard impacted a large number of people. However, the disparity between the "S" category in the Haiti dataset as compared to the OK09, RR09 and RR10 datasets is notable. The social environment ("S") category in the Haiti dataset comprises almost 84% of all tweets relevant to situational awareness; the next highest percentage of "S" tweets is in the RR10 dataset, at almost 55%. The high proportion of social environment tweets in the Haiti dataset is an indication of what information tweets in the Haiti dataset communicated; very little was broadcast about the physical and built environments as compared to the human situation.

5.3 Third Pass Coding: Coding for Information Type

For the third pass of annotation, tweets are sorted into their second pass categories so that annotators view all social environment ("S") tweets and then all built environment ("B") tweets, then all physical environment ("P") tweets and so on; there are a total of seven combinations of second pass categories: B, S, P, SBP, BP, SB and SP. The list of third pass categories is below. Each third pass category appears under the second pass category to which it is assigned. After each category

title, a definition of the category is provided, along with an tweet from each dataset

that includes a segment which indicates the third pass category. Not all third pass

categories include examples from all datasets; if an information type is not

represented in a particular dataset, it is noted.

5.3.1 Social Environment

The following information type categories all include a description of human

actions regarding the mass emergency event. These are the "S" tweets; each

information type is numbered, and includes a definition and example tweets.

1. Advice - Information Space

Advice - Information Space labels those tweets that recommend or urge others to

follow and/or search for specified Twitter users; use, and/or search for specified

hashtags; go to specified links; or pay attention to particular news and/or media

outlets. This category includes tweets that provide information about the digital, or

online environment, as well as information on television and radio stations to watch

or listen to, and telephone numbers to call (e.g. emergency hotlines). Additionally,

these tweets may contain information that corrects hoaxes or false rumors being

spread via mainstream media, social media or microblogging services.

Examples:

OK09	`@fitnewbie I am not.  search #OKfires #MWCfires #OKCfires lots of info`
RR09	`@therealjosh is also tweeting nd flood info`
RR10	`Nice set of links for #fmflood info from @dofargomoorhead`
Haiti	`Last night's hoax on Twitter about American/JetBlue flying doctors/nurses to Haiti for free was just that a hoax. – AA spokesman Tim Smith`

Table 5.19: Tweet Examples, Advice – Information Space

2. Animal Management

Animal Management labels those tweets that include information about pet/animal

evacuations and care due to mass emergency.

Note: No *Animal Management* tweets were coded in the RR10 dataset.

Examples:

OK09	RT @cityofokc: OKC Animal Control is on the scene at 149th and Hiwasee with a response trailer. #okfires
RR09	there's an emergency animal shelter set up at the Fargo fairgrounds as well, RT if you can
Haiti	#News #World Group tackles animal welfare in Haiti after quake http://bit.ly/4FYQT2 (Associated Press)

Table 5.20: Tweet Examples, Animal Management

3. Caution

Caution labels those tweets that include information about warnings, watches,

declarations and other communications meant to inform readers about official and

unofficial notifications about the mass emergency event. These tweets may include

information about warnings issued by government agencies. Tweets may also include

information that is advising people/area residents to take care, to watch for potential

heath issues or other negative effects of the emergency, but is not necessarily an

official mandate or order.

Examples:

OK09	Gary England just said if you live SE of these fires don't climb in your bed and think you're safe tonight.
RR09	Clay County Sheriff Bill Bergquist, "Decide what you want to save." #fargoflood
RR10	Moorhead leaders are warning people not to remove dikes right away after the Red River crests this weekend. http://bit.ly/9aCuKi
Haiti	Haiti warned to brace for another big quake: Reuters http://bit.ly/9RzIW3

Table 5.21: Tweet Examples, Caution

4. Evacuation

Evacuation labels those tweets that indicate an evacuation has been officially ordered by emergency management, that voluntary evacuations are taking place or that mention or imply evacuation in some way.

Examples:

OK09	@cgorshing are you taking your family to the evacuation center at the Midwest City Community Center at 100 North Midwest City Blvd?
RR09	Retweeting @inforum: Breaking: Oak Grove Lutheran School now evacuating. http://www.inforum.com/event/article/id/235700/ #fargoflood #f ...
RR10	Absaraka, ND, on alert for possible evacuation: The Absaraka Dam north of that North Dakota town has people who li... http://bit.ly/a1Lasr
Haiti	Medical evacuations resume in Haiti http://tinyurl.com/ygqhyrp

Table 5.22: Tweet Examples, Evacuation

5. Fatality

Fatality labels tweets that provide information about people who have died or been killed, are suspected of having died, or which alludes to fear of death/dying.

Note: No *Fatality* tweets were coded in the OK09, RR09 or RR10 datasets.

Examples:

Haiti	Hundreds feared dead in #Haiti quake — The Guardian: Telegraph.co.#uk Hundreds feared dead in… http://goo.gl/fb/QOXg
Haiti	REU: Haitian Government Says 40,000 Buried, Another 100,000 Thought Dead In Earthquake #Haiti

Table 5.23: Tweet Examples, Fatality

6. Feeding/Hydration

Feeding/Hydration labels tweets that provide information about the provision of or need for food and/or water by populations affected by a mass emergency.

184

Note: Only the OK09 and Haiti datasets were coded with *Feeding/Hydration*. In addition, this is an information type that was identified through an analysis of government procedures and sociology of disaster literature, which took place after the linguistic analysis of Twitter data.

Examples:

OK09	REMINDER: Bethel Christian in Choctaw has food and supplies for victims of the #OKFires starting at 9am this morning. DM me for more info
Haiti	CNN reporting new food aid plan in Haiti – now handing out rice only to women

Table 5.24: Tweet Examples, Feeding/Hydration

7. General Population Information

General Population Information labels tweets that provide information about the population of a certain city, region or country. These tweets may include information about the number of people, ages, population densities, etc.

Note: Only the Haiti dataset was coded with *General Population Information*.

Examples:

Haiti	CanadaRights.com The help Haiti needs: Baltimore Sun (blog)Haiti has long been the poorest nation in the Americas... http://bit.ly/92G3st
Haiti	Wait, #Haiti had 380k orphans BEFORE the quake?

Table 5.25: Tweet Examples, General Population Information

8. Immediate Recovery

Immediate Recovery labels tweets that communicate information about a immediate post-impact efforts to return to a normal state, which may be at the individual, community, state, national or international level.

Examples:

OK09	Areas back open to residents w/id, mwc officials meeting to decide where to go from here, hot spots in oak cliff east, area closed #okfire
RR09	Fargo begins sandbag leevee removal week: The City of Fargo is encouraging homeowners to begin sandbag removal i... http://bit.ly/aTlyFr
RR10	Fargo levees coming down starting Friday: Clay levees that protected City of Fargo property during this spring's f... http://bit.ly/areD9P
Haiti	Boston Globe reports on PIH's relief/rebuilding efforts at a hospital in #Haiti & at our Boston office http://bit.ly/6aQnq9

Table 5.26: Tweet Examples, Immediate Recovery

9. Injury

Injury labels tweets that communicate information about people who have suffered harm due to a hazard agent or in preparation for the impact of a hazard agent. These tweets also include information about lack of injuries, or low levels of injuries.

Note: No *Injury* tweets were coded in the RR10 dataset.

Examples:

OK09	Gov. confirms at least 34 injured. Some hot spots in south central OK. "Amazing no lives were lost." Property loss could have been
RR09	suffering from back pain caused by bending over sand bags yesterday--needed Enrico's help to get out of bed. After icing, will try moving
Haiti	Almost half Haiti's injured may be children: study (AFP)http://bit.ly/9A9cZT

Table 5.27: Tweet Examples, Injury

10. Medical Attention

Medical Attention is used to label tweets that provide information about the health and medical needs of those affected by the mass emergency. These tweets may include information about providing care; medical personnel traveling to and/or helping at the site of the disaster; special attention needed for those with medical conditions; and needs specific to the health and wellness.

186

Note: Only the RR09 and Haiti datasets were coded with *Medical Attention*. In addition, this is an information type that was identified through an analysis of government procedures and sociology of disaster literature, which took place after the linguistic analysis of Twitter data.

Examples:

RR09	`@therealjosh It helps to know that the elderly and patients have all been evacuated. I think will win this fight! :) Thinking of you`
Haiti	`Stanford Hospital Medical Team Travels To Haiti To Assist In Relief Effort: A seven-person team of Stanford Hospit... http://bit.ly/6x8r8H`

Table 5.28: Tweet Examples, Medical Attention

11. Missing

Missing labels tweets that mention or imply that people are missing, unaccounted for or that their locations are otherwise unknown.

Note: No *Missing* tweets were coded in the OK09, RR09 or RR10 datasets.

Examples:

Haiti	`Filipinos unaccounted for - Manila Bulleting Online: PORT-AU-PRINCE, Haiti (AP) — The strongest earthquake in more... http://bit.ly/8iM072`
Haiti	`Continuing the search for missing students and staff from Lynn University who were on an academic mission in Haiti`

Table 5.29: Tweet Examples, Missing

12. Offer of Help

Offer of Help labels tweets that communicate Twitterers' ability or willingness to help those affected by a mass emergency. Individuals or groups of people may be offering help.

Examples:

OK09	RT @rhukill: PLEASE RETWEET. If anyone affected by the #okfires needs baby clothes please let @jeditanner know!
RR09	Anyone needing a place to stay right now in Fargo can DM me and I will help you out! I have an extra bed room.
RR10	If you need help during the flood, all of us on AM 970 WDAY are here for you! You call, we'll respond! 701-293-9000, 800-279-9329. #flood10
Haiti	American Airlines is taking doctors and nurses to Haiti for free!! Call 212-697-9767....pass it on

Table 5.30: Tweet Examples, Offer of Help

13. Preparation

Preparation labels tweets that communicate information about activities and actions meant to guard against or ready people/property for an impending hazard event, or mention that preparations in general are taking place.

Note: No *Preparation* tweets were coded in the Haiti dataset.

Examples:

OK09	@jwendz It can start anywhere - a simple spark can burn down a whole neighborhood. I watered the yard and roof and hope for the best #OKFire
RR09	We're building the dikes up to 44' now! They're at 43'.
RR10	ND Nat'l Guard plans to fill several 1-ton sandbags today, in case they're needed during the flood fight.... #fmflood #flood10

Table 5.31: Tweet Examples, Preparation

14. Report of Crime

Report of Crime labels tweets that include information about any illegal activity that has taken place, or which advise or warn against partaking in criminal activity.

Note: No *Report of Crime* tweets were coded in the OK09 dataset.

Examples:

RR09	Red River at Fargo: 37.37. 3 and a half feet below record crest. Garage Fires & Burglaries reported in flood areas Details on CH 4
RR10	Sheriff Bergquist: 30+ roads washed out in Clay County, MN -- don't drive around barricades... you could face charges....
Haiti	#haiti There are rumors of sexual violence in the camps. Lack of adequate shelter is a major problem. /by @CARE

Table 5.32: Tweet Examples, Report of Crime

15. Request for Help

Request for Help labels tweets in which Twitter users ask for help directly or

indirectly, or relay messages about others' request for help or aid of some sort. This

may include requests for others to create technological tools (e.g. map mashups), or to

provide any number of services. Additionally, *Request for Help* tweets may indirectly

ask for help by stating a need.

Examples:

OK09	Challenge for you mashup people in OKC: How can we see near-realtime map of OKC area fires?
RR09	Please send help to Fargo ND and Morehead MN. We need all the help we can get.
RR10	Flood fight volunteer turnout is light in Moorhead. Volunteers encouraged to report to Nemzek Hall at MSUM. http://bit.ly/9yo5kz
Haiti	@cnnbrk @RedCross People are dying at the Port-au-Prince Municipal Nursing Home. WATER AND FOOD IS NEEDED!!! #HelpHaiti #Haiti

Table 5.33: Tweet Examples, Request for Help

16. Request for Information

Request for Information labels tweets in which Twitter users ask for updates about the

mass emergency situation. These may include requests regarding the wellbeing and

safety of others, requests for the status of the hazard, and requests for information

about the infrastructure or specific property.

Examples:

OK09	@maggiehoey how bad is it up that way?
RR09	Any Fargo twitters know if University Ave and all the other roads are open again? #fargo #redriver #flood09
RR10	@gregandmarlo that's good! looking forward to listening tomorrow! any flood concerns up in Winnipeg?
Haiti	anyone have information on Borgne region of Haiti?

Table 5.34: Tweet Examples, Request for Information

17. Rescue

Rescue labels tweets that include information about people being rescued from perilous situations.

Note: No *Rescue* tweets were coded in the OK09, RR09 or RR10 datasets.

Examples:

Haiti	A Moment of Joy in Haiti -A young boy named Kiki was lifted from the rubble he had been trapped under for 8 days. http://ow.ly/i/n3n
Haiti	Check this video out -- Raw Video: Rescuers Save Trapped Woman in Haiti http://youtu.be/jBxj1t1FnRw These rescues are still going on.

Table 5.35: Tweet Examples, Rescue

18. Response – Community

Response - Community labels tweets that convey information about responses to a mass emergency by local organizations such as faith-based, non-profit and similar organizations or groups of community members.

Examples:

OK09	Bethel Christian will have 400 boxes of food, drinks, and hygiene kits to give to fire victims after 9am tomor. DM me for more info. Plz RT.
RR09	My wife's workplace just closed up until next week to allow their employees to fight the flood full time. Way to go! #flood09
RR10	More students from #Fargo will be helping to button up dikes today.... #fmflood #flood10
Haiti	BREAKING NEWS: Churches Responding As Haiti Earth Quake Kills Bishop And Thousands More - http://bit.ly/4qWC4r

Table 5.36: Tweet Examples, Response – Community

19. Response – Formal

Response – Formal labels tweets that convey information about response or lack of response to a mass emergency by local or state emergency responders such as fire departments, non-government organizations, the military and similar groups. This category also includes recommendations and advice from official community, state and national leaders.

Examples:

OK09	Gotebo FD is out on Fire at Ozark Trail and Hwy 54. Light rain. Fire contained.
RR09	Province sending ambulances to Fargo - http://tinyurl.com/dz8gf8 #manitobaflood #redriver #flood09
RR10	Levee breach in Harwood: National Guard shores up levee breach in Harwood. http://bit.ly/cuzhKI
Haiti	USS Carl Visnon (CVN 70) departed the waters near Port-au-Prince today. Carl Vinson was instrumental in providing a... http://bit.ly/aV1CiO

Table 5.37: Tweet Examples, Response – Formal

20. Response – Miscellaneous

Response – Miscellaneous labels tweets that mention information about response or lack of response to a disaster, but which cannot be labeled with *Response – Community*, *Response – Formal* or *Response – Personal*. These tweets do not contain

any information that allows the reader to assess who is responding, but which

nonetheless is communicating about response.

Note: No *Response – Miscellaneous* tweets were coded in the OK09, RR09 or RR10

datasets.

Examples:

| Haiti | 140 bottles of breast milk was delivered to the babies of Haiti however they may not get them. |
| Haiti | Relief teams helping Haiti's traumatized quake victims : http://tinyurl.com/y8m7349 |

Table 5.38: Tweet Examples, Response – Miscellaneous

21. Response – Personal

Response – Personal labels tweets that convey information about response or lack of

response to a mass emergency by persons acting alone or in small groups (e.g.

families).

Examples:

OK09	Live in SE OKC, going home to hook up all my water hoses, pray for @slmoore family, fire is close to their place #okfire
RR09	Enrico and the boys are moving books from the basement to the main level. We thought we were all good with our drain plug, but ....
RR10	People I know are packing up their houses and moving out to avoid the flood. Last year they were helicoptered out. Trying to avoid that.
Haiti	So u found out my dads gunna b movin dead bodies in haiti n helpin to build a hospital.. Eeek! :(

Table 5.39: Tweet Examples, Response – Personal

22. Sheltering

Sheltering labels tweets that communicate information about the location of

sheltering options for displaced people, as well as information about where safe areas

are located.

Note: No *Sheltering* tweets were coded in the RR10 or Haiti datasets.

Examples:

OK09	RT @redcrossokc: A shelter has opened in Grady County at Alex School located at 209 S. 2nd Street near Hwy 19 and Main St. #okfire
RR09	Flood Shelter Open at Alerus Center in Grand Forks, ND @ 4:00 pm Wed - RedCross (218)773-9565 #redriver #fargoflood #ndfloods #flood09

Table 5.40: Tweet Examples, Sheltering

23. Status – Community/Population

Status – Community/Population labels tweets that provide information about a

community or population as a whole and how they are managing with mass

emergency.

Examples:

OK09	Fires under control, residents allowed back into homes, we're safe
RR09	residents north of winnipeg cling to rooftops waiting for #flood rescue-http://tiny.cc/xmyYL
RR10	ChastityWalberg: Mayor Walaker: "We are in good shape today." #fmflood #flood10
Haiti	There were kids in the St Gerard skool when it collapsed. I haven't seen any sort of logistical backup in this part of town since the quake

Table 5.41: Tweet Examples, Status – Community/Population

24. Status – Personal

Status – Personal labels tweets that provide information about individuals as that

status relates to the emergency situation. Some are written by Twitter users who are

personally affected by a hazard ("I'm okay" or "I've been evacuated"). Additionally,

these tweets may be written by users asking about the well-being of others ("are you

okay?") and/or relaying information about the status of others ("John Doe is okay").

Examples:

OK09	@NataleeGodfrey I'm at a neighbors parents right now going to take kids to shawns parents I think I won't sleep till I know were clear
RR09	I'm crashing early to be ready for more flood fighting. I am sleeping in the basement so I can hear any H20 invasion. #flood09 #fargoflood
RR10	When asked how she's doing, homeowner at a loss for words. She just renovated home after last yr's flood, now river is 3 ft away #fmflood
Haiti	Wow this little girl is stuck under cement. This is sad. #haiti on CNN

Table 5.42: Tweet Examples, Status – Personal

5.3.2 Built Environment

The following information type categories all include a description of ways in which buildings, infrastructural components such as roads and bridges, property and other structures are affected by the mass emergency. These are the "B" tweets, and each information type is numbered, and includes a definition and example tweets.

25. Damage

Damage labels those tweets that provide information or allude to information about a structure, facility or property that has suffered from a hazard. This can be for public or private structures and property of any variety.

Examples:

OK09	Update: About 100 homes have now been destroyed or severely damaged in Midwest City, fire spokesman Jerry Lojka said. #OKFire #OKFires
RR09	Oak Grove School in Fargo has been lost to the flood. #fargoflood
RR10	In '09, #Fargo restored the backyards of 237 homes damaged by levees, claims averaged $1,700 – this yr. not as much damage #fmflood #flood10
Haiti	Haiti's cathedral was in ruins after the earthquake: BBC NewsThis past week, like everybody else on the globe, http://bit.ly/dtfC6u

Table 5.43: Tweet Examples, Damage

26. Status – Infrastructure

Status – Infrastructure labels tweets that contain information about the conditions,

openings/closures and other status updates of roads/highways, bridges, electricity,

water, and other public utilities, public transportation, airports and other basic

organizational aspects that allow cities, states and countries to function.

Examples:

OK09	#okfire I-35 reopened in Payne County, HWY 51 still closed
RR09	Hwy 2 in GF is still open #flood09
RR10	Odor prompts change in Mhd. water supply mix: The 2010 flood is prompting a change in the way Moorhead gets its wa... http://bit.ly/99x8AH
Haiti	Flights to Haiti suspended after quake http://goo.gl/fb/H4hc

Table 5.44: Tweet Examples, Status – Infrastructure

27. Status – Personal Property

Status - Personal Property labels tweets that convey information and/or ask questions

about the condition of personal property such as homes.

Examples:

OK09	@librarising79 And I am blessed that my home (which I love so much) is intact. Most of the wildfires are out or under control now.
RR09	Oakport development North of Moorhead, MN has lost 500 homes. #fargoflood
RR10	@thebroadbroad Thanks darling! My house is safe but my parents live by the river so we have been sandbagging it's looking better
Haiti	The earthquake collapsed many houses in Port-au-Prince, Haiti. I wish to go there, so I think to help something to do at this time

Table 5.45: Tweet Examples, Status – Personal Property

28. Status – Public Property

Status – Public Property labels tweets that communicate information about conditions

of public property such as a school, hospital, government facility and/or other

community building or publically accessible property.

Examples:

OK09	@cnn fire department in carter county oklahoma burned. documents, and a firetruck
RR09	Retweeting @inforum: Dike fails at Fargo's Oak Grove Lutheran School, two buildings lost http://is.gd/puWd
RR10	RT @redriveratfargo: Nothing to worry about, but there is now a waterfall filling up the el zagel golf course.
Haiti	My cousin was at school & jumped out the window when the earthquake hit & the building was shook. Building is gone life spared #godisgood

Table 5.46: Tweet Examples, Status – Public Property

5.3.3 Physical Environment

The following information type categories all include a description and/or comparative data regarding the hazard agent, e.g. location of a fire or level of a river. These categories also include information about physical aspects of the affected area, e.g. weather, geographical details, characteristics of city/county borders and similar information.

29. General Area Information

General Area Information labels those tweets that provide geographic or logistical information about areas under threat or experiencing disaster.

Examples:

OK09	Geo-note for our non-Oklahoma friends: the fires around Velma-Ratliff City-Fox-Loco-Meridian are in oil producing regions. #OKFires
RR09	@skangus GF built permanent levees after 97 flood of the Red. Difference was they declared entire neighborhoods greenspace to do it.
RR10	@femaregion8 ND and MN and border the Red River which is flooding now.
Haiti	Haiti earthquake occurred in complex, active seismic region http://bit.ly/6VIRyN

Table 5.47: Tweet Examples, General Area Information

30. General Hazard Information

General Hazard Information labels those tweets that provide information about the

hazard agent at hand, but are not specific to a mass emergency situation.

Examples:

OK09	@Leosulli Firestorm is a fire. Wind is Hi n fuels fire. It spreads n wind carries it. Fire 'jumps' from house to house. Cedars, grass catch.
RR09	@skangus Red River is like a pencil scratch in kitchen table. Plus it flows north. So downstream tends to backup.
RR10	Red River in #Fargo #Moorhead now at 30.47 feet -- "Major" flood stage is 30 feet -- #flood10
Haiti	@GottaLaff The earthquake happened because Haiti is on top of a major fault. Like LA and SFO are built on top of a major fault.

Table 5.48: Tweet Examples, General Hazard Information

31. Historical

Historical labels tweets that provide information comparing present hazard conditions

to past hazard conditions of the same or different types.

Examples:

OK09	"This is probably the worst disaster in Midwest City's history, even worse than the May 3rd tornadoes" PD Chief Brandon Clabes.
RR09	River at Hickson has exceeded 1997 levels #fargoflood #ndfloods #flood09 #redriver
RR10	#fmflood. Three 100-year floods in 15 years. So, after this, we're flood free for 300 years?
Haiti	Video: Catastrophe in Haiti. Strongest Earthquake in 200 Years http://ow.ly/16ksO1

Table 5.49: Tweet Examples, Historical

32. Prediction

Prediction labels tweets that provide information about forecasts of what will happen

with regards to the hazard, or what the hazard will do, e.g. where a fire will spread or

when a river will crest.

Note: No *Prediction* tweets were coded in the Haiti dataset.

Examples:

OK09	Well this fire here in oklahoma it looks like its going to get bigger than the 12 miles it is now
RR09	@velvetflamingo I haven't seen anything about the dike, specifically, but crest forecasts seem to be lowering.
RR10	Expected crest in Grand Forks is 49 feet. Permanent levee system is good to 54 feet. #flood10

Table 5.50: Tweet Examples, Prediction

33. Status – Hazard

Status – Hazard labels tweets that provide information about the location and/or state of a hazard such as the intensity of a fire and where it is burning, the height of a river or other body of water in relation to flood levels, and the strength of an earthquake and information about aftershocks.

Examples:

OK09	RT @d_stew: Fires have started rekindling in SE OKC near Choctaw. East winds are creating tough conditions for firefighters again.
RR09	ice jams worsen in north dakota
RR10	RT @fargofloodstage: The Red River at Fargo ND is at 26.05 ft which is 8.05 ft above flood stage #flood10 #fargoflood
Haiti	Thirteen aftershocks strike Haiti. Seven of them were above a magnitude 5. Here are images from Twitter: http://bit.ly/5Ue17R

Table 5.51: Tweet Examples, Status – Hazard

34. Weather

Weather labels tweets that communicate information about weather conditions, and specifically, how those conditions that may have effects on the progression or outcome of a mass emergency event.

Examples:

OK09	Wind gusts in Cordell reported 65mph+ #OKFires
RR09	Thursday rainfall on the way, but potentially not as heavy as early thought. Still more ice to melt north of Valley City & Jamestown.
RR10	RT @FargoMoorhead: Looking to really warm up over the next few days in #fargo Moorhead. Good and bad all at the same time. #fmflood
Haiti	@leashless re cold weather, that's not the case in #haiti with 17-31ºC today in capital area; maybe for kids and frail, only

Table 5.52: Tweet Examples, Weather

5.3.4 Third Pass Coding Kappa Calculations

The calculation of kappa values for the third pass categories involves taking each group of tweets that have been organized according to second pass category (e.g. all "SB" tweets) and calculating the agreement between the two annotators. These results are provided in Tables 5.53-5.56. Previous research I performed on Twitter data indicates that the most third pass categories that were ever assigned to one tweet is five, and additional annotation by the annotators and me revealed this to be the case. So, the maximum number of options for agreement or disagreement in each subset of tweets is the number of tweets in that subset, multiplied by five. The exception to this is for those tweets that only include information about the built environment, or the "B" tweets. There are only four third pass categories to choose from when annotating "B" tweets, so in this case, the total number of options to agree or disagree is four multiplied by the number of tweets in the "B" subset. The third pass kappa values for each dataset are below. By this point in the annotation process, the external annotators were experts, and I acted as the adjudicator.

Annotators	OK09 3rd Pass Kappa Values						
	BP	*B*	*SBP*	*SB*	*SP*	*S*	*P*
Preciado and Vollmann	91.9%	91.2%	81.8%	87.6%	86.8%	91.1%	93.0%

Table 5.53: OK09 3rd Pass Kappa Values

Annotators	RR09 3rd Pass Kappa Values						
	BP	*B*	*SBP*	*SB*	*SP*	*S*	*P*
Preciado and Vollmann	83.7%	91.9%	65.4%	81.6%	78.3%	86.9%	90.3%

Table 5.54: RR09 3rd Pass Kappa Values

Annotators	RR10 3rd Pass Kappa Values						
	BP	*B*	*SBP*	*SB*	*SP*	*S*	*P*
Preciado and Vollmann	96.8%	99.2%	79.2%	82.2%	84.8%	88.2%	98.3%

Table 5.55: RR10 3rd Pass Kappa Values

Annotators	Haiti 3rd Pass Kappa Values						
	BP	*B*	*SBP*	*SB*	*SP*	*S*	*P*
Preciado and Vollmann	93.7%	83.3%	79.3%	81.5%	87.4%	89.1%	92.3%

Table 5.56: Haiti 3rd Pass Kappa Values

Kappa calculations for all categories in all datasets are between good and fair, save the "SBP" category in the RR09 dataset, which was 65.4%. Further analysis of these tweets shows that one annotator was assigning more information type categories than the other, which is what led to the lower kappa value. I discussed this with the annotators, we practiced coding example tweets together, and all subsequent annotation resulted in higher agreement.

5.3.5 Third Pass Coding Results

The following sections provide a description of the third pass of coding for each dataset. In addition to the coding results, a discussion of third pass category development is included.

5.3.5.1 2009 Oklahoma Fires Third Pass Coding Results

There are 535 tweets which contain information relevant to situational

awareness in the OK09 dataset. Table 5.57 shows the total number of *instances* of

each information type, and the percentage of all tweets that contain information

relevant to situational awareness comprised of each information type. The total

number of instances of information in the OK09 dataset is 835, which is an average of

1.6 instances per tweet. Figure 5.9 displays the information in Table 5.57 in a graph.

OK09 3rd Pass Category	Number of Instances	Percentage of all 'R' Tweets
Status – Hazard	158	18.9%
Damage	115	13.8%
Weather	80	9.6%
Status – Personal	77	9.2%
Evacuation	73	8.7%
Advice – Information Space	59	7.1%
Response – Formal	58	6.9%
Status – Infrastructure	31	3.7%
Caution	28	3.4%
Response – Community	27	3.2%
Sheltering	24	2.9%
Request for Information	18	2.2%
Status – Personal Property	14	1.7%
Feeding/Hydration[32]	8	1.0%
Status – Public Property	7	0.8%
Immediate Recovery	7	0.8%
Offer of Help	7	0.8%
Response – Personal	6	0.7%
Injury	6	0.7%
Status – Comm/Population	5	0.6%
Request for Help	5	0.6%
Prediction	4	0.5%
General Hazard Information	4	0.5%
General Area Information	4	0.5%
Animal Management	4	0.5%
Historical	3	0.4%
Preparation	3	0.4%

Table 5.57: OK09 3rd Pass Tweet Instance Totals

---

[32] The Feeding/Hydration category was not part of the coding schema used to annotate the OK09 data, though it is part of the final offering regarding information communicated via Twitter in mass emergency situations. I added this category after the multiple-annotator coding process, so these tweets were not coded by multiple annotators, only by me.

Figure 5.9: OK09 3rd Pass Instances per Category

5.3.5.1.1 OK09 Third Pass Category Details

The third pass category options for the OK09 dataset were determined based on a multi-year study of the OK09 and RR09 events. These mass emergency events were the first to be closely examined, and the empirical analysis I performed in previous research uncovered the majority of the categories listed in Table 5.57, and displayed in Figure 5.9. Once I began to examine tweets for this dissertation, I further honed the categories which resulted in the final category schema for the OK09 dataset. For a category to be created, the initial criteria were that it must convey information that relates to situational awareness, and there must be a minimum of five tweets in one of the datasets that contain the information category.

The third pass coding schema evolved over time. One category that was originally a part of the schema—listed under social environment—was *Advice – General*. This category was meant to label those tweets that offered guidance on how

to manage situations related to a mass emergency, but which did not offer information covered in one of the other social environment categories, such as *Immediate Recovery* or *Preparation*. It became clear that those tweets labeled with *Advice – General* also fell into at least one other Social Environment category, which provided a more detailed explanation of the tweet content. Therefore, this category was removed from the annotation guidelines.

Another social environment category that was eventually removed from the third pass coding schema is *Volunteering*. There is mention of volunteers and volunteering in the OK09 and RR09 datasets, which is why this category was originally created. However, after analyzing both datasets after they were labeled with third pass codes, I found that the *Volunteering* category co-occurred with *Request for Help*, and/or *Offer of Help*. Since both the *Request for Help* and *Offer of Help* are applicable to more tweets—as not all tweets in these categories include information about volunteering—I removed *Volunteering* from the coding schema.

5.3.5.2 2009 Red River Flood Third Pass Coding Results

1483 tweets that contain information relevant to situational awareness were coded with third pass information categories in the RR09 dataset. Table 5.58 shows the breakdown of the number of instances of each third pass category, their percentage of all RR09 tweets that contain situational awareness information, and Figure 5.10 provides the visual depiction. There are 2550 total instances of information type in the 1483 "R" tweets, which is an average of 1.7 instances per tweet.

RR09 3rd Pass Category	Number of Instances	Percentage of all 'R' Tweets
Status - Hazard	564	22.1%
Preparation	274	10.7%
Advice – Info Space	205	8.0%
Response - Formal	164	6.4%
Historical	148	5.8%
Status – Infrastructure	141	5.5%
Status – Personal	139	5.5%
Prediction	134	5.3%
Evacuation	124	4.9%
Weather	115	4.5%
Status – Comm/Population	101	4.0%
Request – Help	82	3.2%
Response – Personal	77	3.0%
Response – Community	66	2.6%
Caution	45	1.8%
Request – Information	39	1.5%
Damage	31	1.2%
Status – Personal Property	29	1.1%
General Area Information	15	0.6%
Status – Public Property	10	0.4%
Offer of Help	8	0.3%
General Hazard Information	7	0.3%
Immediate Recovery	7	0.3%
Report of Crime	7	0.3%
Injury	6	0.2%
Animal Management	5	0.2%
Medical Attention[33]	5	0.2%
Sheltering	2	0.1%

Table 5.58: RR09 3rd Pass Tweet Instance Totals

[33] The Medical Attention category was not part of the coding schema used to annotate the RR09 data, though it is part of the final offering regarding information communicated via Twitter in mass emergency situations. I added this category after the multiple-annotator coding process, so these tweets were not coded by multiple annotators, only by me.
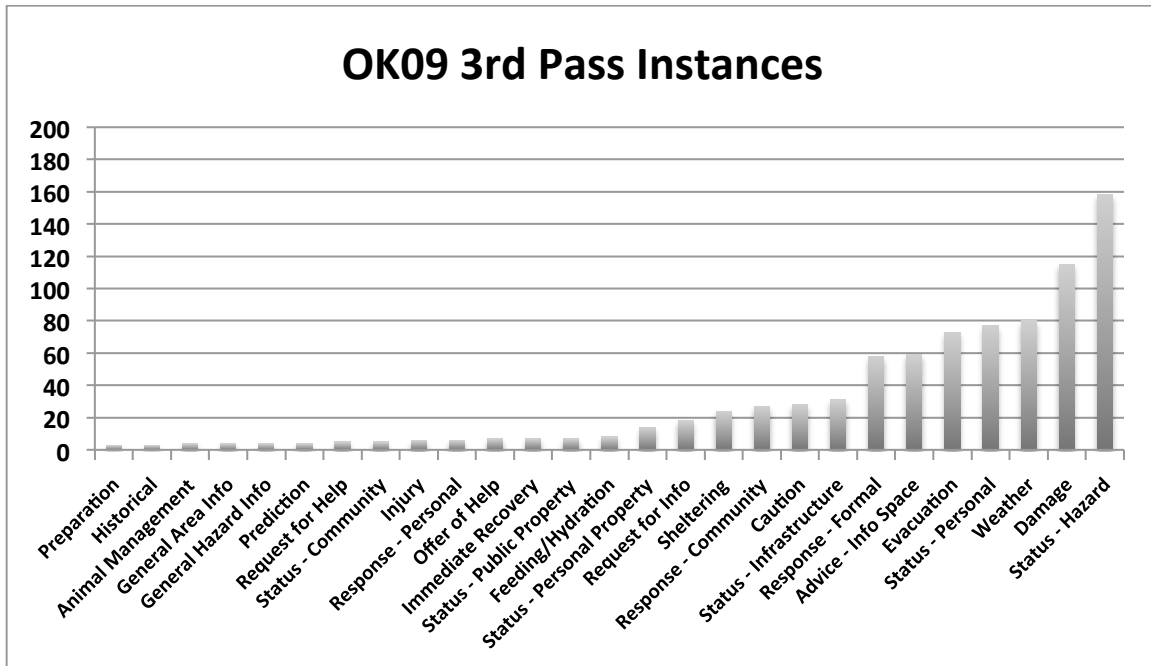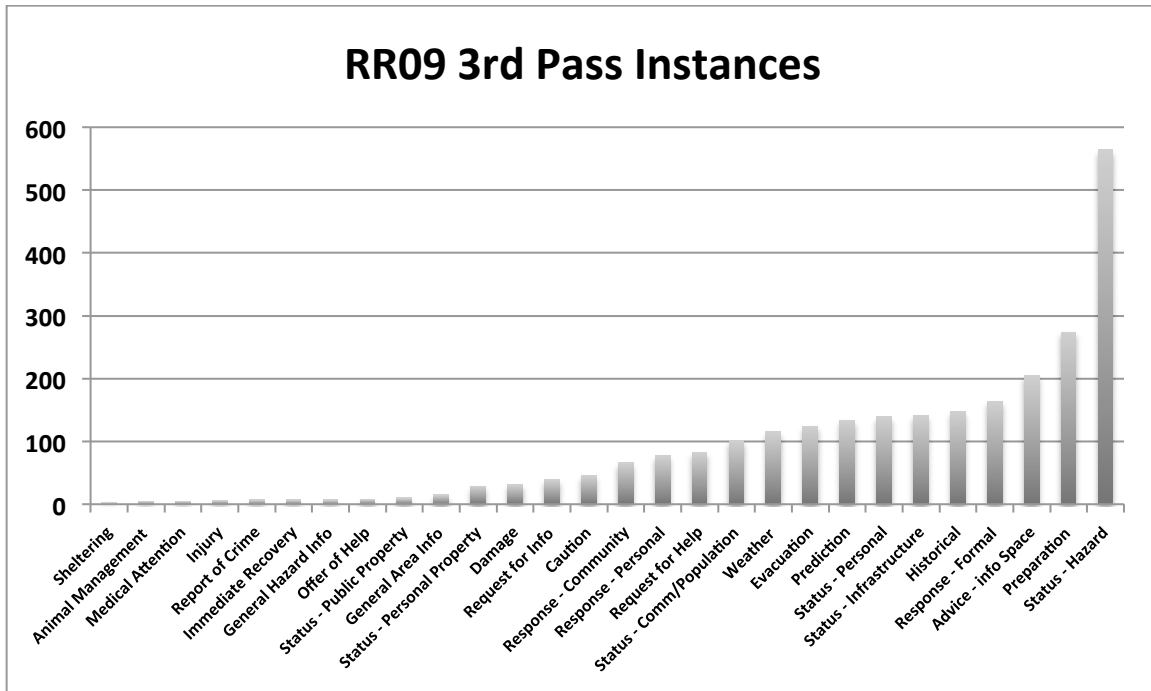
Figure 5.10: RR09 3rd Pass Instances per Category

5.3.5.2.1 RR09 Third Pass Category Details

The RR09 third pass coding schema is the same as that for OK09, save one category. A third social environment category that arose after annotators had finished coding the OK09 dataset, and had begun coding the RR09 dataset, is *Report of Crime*. Though I developed the coding categories for OK09 and RR09 in tandem, using input from each event, once the annotators got involved with the coding process, and we began to regularly discuss tweet content, it became clear that a *Report of Crime* category had emerged from the RR09 dataset.

An additional factor to note for the RR09 dataset is regarding the *Injury* category. There are six tweets coded with *Injury* in this dataset. However, three of them actually mention a fatality, as seen in this example:

```
Two deaths reported related to flooding along Red River in ND.
(health dept says they're cardiac-related, due to flood
prevention exertion.
```

The decision to include tweets that mention death in the *Injury* category is because

there were only three tweets that mentioned death in the RR09 dataset. At this time

during the coding process the minimum number of tweets required for a new category

to be named was five. It was not until the analysis of the Haiti dataset began that I

added the *Fatality* category, due to the large number of tweets that mention death in

the Haiti dataset. However, I made the decision to keep the original coding scheme

and code assignments in the RR09 dataset.

5.3.5.3 2010 Red River Flood Third Pass Coding Results

The results of the third pass of coding that was performed on the 1091 "R"

tweets in the RR10 dataset are displayed below. Table 5.59 provides the breakdown

of number of instances for each third pass category, the percentage each information

type represents of all "R" tweets, and Figure 5.11 provides a graphical representation

of that same information. The total number of third pass category instances in the

RR10 dataset is 1507, which is an average of 1.4 instances per tweet.

RR10 3rd Pass Category	Number of Instances	Percentage of all 'R' Tweets
Status – Hazard	353	23.4%
Advice – Information Space	224	14.9%
Preparation	219	14.5%
Response – Formal	150	10.0%
Response – Community	134	8.9%
Status – Infrastructure	94	6.2%
Status – Community	65	4.3%
Weather	52	3.5%
Request for Help	48	3.2%
Status – Personal	34	2.3%
Response – Personal	31	2.1%
Caution	31	2.1%
Prediction	11	0.7%
Damage	10	0.7%
Request for Information	9	0.6%
Immediate Recovery	9	0.6%
Historical	8	0.5%
Status – Personal Property	5	0.3%
Evacuation	5	0.3%
General Hazard Information	4	0.3%
Report of Crime	4	0.3%
Offer of Help	3	0.2%
General Area Information	2	0.1%
Status – Public Property	2	0.1%
Animal Management	0	0.0%
Injury	0	0.0%
Sheltering	0	0.0%

**Table 5.59: RR10 3rd Pass Tweet Instance Totals**
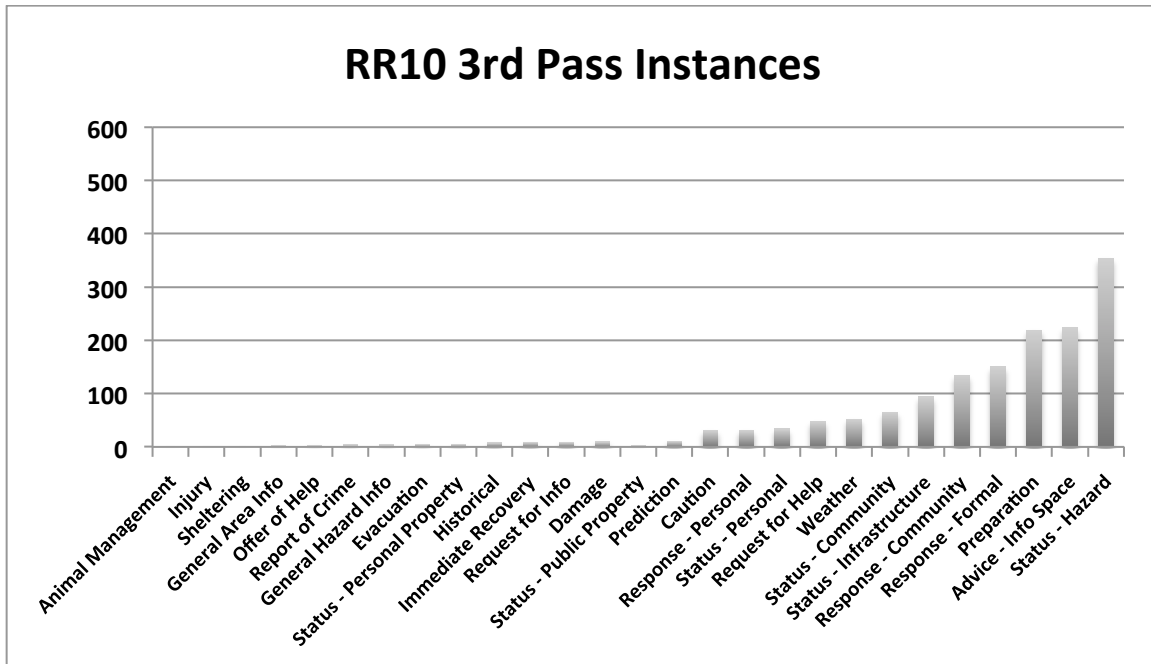
Figure 5.11: RR10 3rd Pass Instances per Category

5.3.5.3.1 RR10 Third Pass Category Details

The RR10 third pass coding categories are the same as those used for the

RR09 dataset. When performing the third pass of coding on all datasets, annotators

were instructed to identify any tweets that contained information relevant to

situational awareness but which did not fit into any of the previously defined

categories from RR09; no new information types were revealed. This is not surprising

as the 2010 Red River flood was similar to the 2009 Red River flood. Though the

2009 flood was more severe, caused more damage and involved a record-breaking

river crest level, Twitter users in the 2010 flood were communicating the same types

of information as the previous year, save for the lack of information about *Animal

Management, Injury* and *Sheltering*—none of these information types is mentioned in

the RR10 dataset.

209

5.3.5.4 2010 Haiti Earthquake Third Pass Coding Results

The third pass of coding on the Haiti dataset resulted in 1414 instances of information for the 943 R tweets, which is an average of 1.5 instances per tweet. Results of the third pass are shown in Table 5.60 and Figure 5.12.

Haiti 3rd Pass Category	Number of Instances	Percentage of All 'R' Tweets
Advice – Information Space	229	16.2%
Response – Formal	144	10.2%
Fatality	139	9.8%
Status – Personal	108	7.6%
Status – Community/Population	108	7.6%
Response - Miscellaneous	84	5.9%
Damage	82	5.8%
Request for Help	61	4.3%
Status – Hazard	55	3.9%
Rescue	49	3.5%
Report of Crime	41	2.9%
Feeding/Hydration[34]	38	2.7%
Medical Attention[6]	35	2.5%
Response – Community	32	2.3%
Response – Personal	31	2.2%
Request for Information	28	2.0%
Injury	21	1.5%
Status – Public Property	19	1.3%
Missing	19	1.3%
General Population Information	16	1.1%
Status – Infrastructure	15	1.1%
Caution	13	0.9%
Immediate Recovery	10	0.7%
Offer of Help	9	0.6%
Historical	8	0.6%
General Area Information	6	0.4%
Evacuation	5	0.4%
Animal Management	3	0.2%
Weather	2	0.1%
General Hazard Information	2	0.1%
Status – Personal Property	2	0.1%
Prediction	0	0.0%
Sheltering	0	0.0%
Preparation	0	0.0%

Table 5.60: Haiti 3rd Pass Tweet Instance Totals

---

[34] The Feeding/Hydration and Medical Attention categories were not part of the coding schema used to annotate the Haiti data, though it is part of the final offering regarding information communicated via Twitter in mass emergency situations. I added these categories after the multiple-annotator coding process, so these tweets were not coded by multiple annotators, only by me.
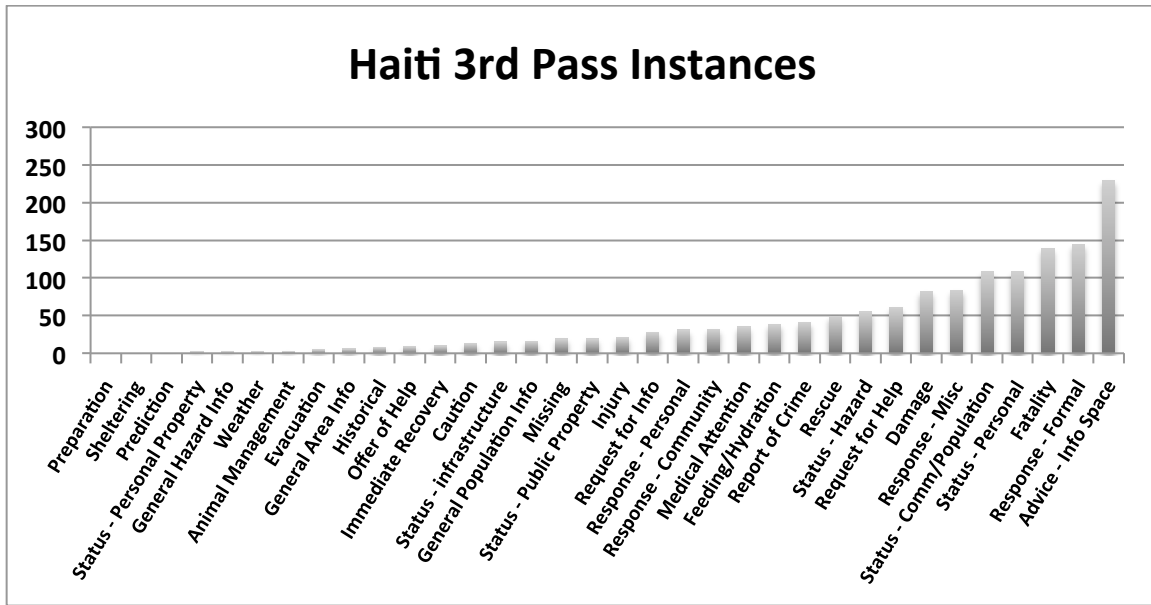
Figure 5.12: Haiti 3ʳᵈ Pass Instances per Category

5.3.5.4.1 Haiti Third Pass Category Details

As annotators worked their way through the third pass of coding for the Haiti dataset, several new information types were identified—all are included in the Social Environment ("S") second pass category. The first information type to be added to the third pass coding schema during the Haiti annotation was *Rescue*, because tweets include information about people being saved from perilous situations in the aftermath of the earthquake. In the OK09, RR09 and RR10 datasets, there was no mention in the datasets of people placed in potentially deadly circumstances that required rescue.

Another category added during the Haiti annotation is *Fatality*. Both Vollmann and Preciado had independently suggested a new category entitled *Casualty*. However, the coding schema already included an *Injury* category, and a strict definition of "casualty" specifies that it describes both the injured and the dead due to an accident, mass emergency or war. Therefore, I added the category *Fatality*

211

and maintained the existing category of *Injury*. The *Fatality* category applies to tweets that include information about people who have died, and the *Injury* category applies to tweets that include information about people who have suffered bodily harm, but who are alive.[35]

The *Response – Miscellaneous* category also arose out of the Haiti third pass annotation. This category was identified because in the Haiti dataset, there are tweets that include information about responding to the aftermath of the earthquake, but which do not indicate if it is a personal, community or formal response effort. Many response efforts took place and were communicated about on Twitter, but these efforts were not always attributed to a particular person or group of people.

Another information type category that emerged during the Haiti annotation effort is *General Population Information*. Tweets in the Haiti dataset include information about the Haitian population, such as the number of orphans and the average age of a Haitian citizen. Twitter users communicated this type of information, likely to inform others about the state of the Haitian people previous to the devastating earthquake. Many in Haiti were in a precarious situation before the earthquake struck, making post-earthquake circumstances potentially dire.

The final new category added to the third-pass coding schema as a result of the Haiti annotation is *Missing*. In the aftermath of the earthquake, many peoples' locations were unknown, and Twitter users broadcast information about specific individuals and groups of people who were missing. Updates on the search efforts for missing people were also provided.

---

[35] The exception to this is noted earlier in this chapter

Another change that took place in the third pass coding schema was a modification to the *Status – Community/Population* category. In the Haiti dataset, many tweets include information about the current state of "Haiti." "Haiti" is used in this sense to imply the Haitian people as a whole. Previous annotation of the other three datasets required that annotators label tweets that mention the status of affected populations with *Status – Community*, as the other three events are focalized, and tweets tend to mention people in particular neighborhoods, towns and cities. Since tweets in the Haiti dataset often focus on large numbers of Haitian people, the *Status – Community* category was amended to become the *Status – Community/Population* category. It is still the case that some tweets in the Haiti dataset refer to specific areas or regions of Haiti, hence the decision to augment the *Status – Community* category, as opposed to creating a new *Status – Population* category.

The final amendment to the third pass coding schema based on linguistic analysis of Twitter data was to further specify the definitions for the *Response – Community* and *Response – Formal* categories. In performing the third pass coding on the Haiti dataset, annotators noted uncertainty when they came across tweets that mentioned response by the Red Cross, or by the Salvation Army. In the context of the Haiti earthquake, responses by these organizations are considered formal. The Red Cross, the Salvation Army and other aid organizations have local as well as national and international chapters that respond to various events as needed. In the case Haiti, it was an international effort on behalf of these organizations, requiring more personnel and resources than a response to a focalized event does. Annotators were trained to differentiate between tweets that mention response by these organizations

213

based on if they were coding tweets from a focalized or a diffused event. In focalized events, tweets that mention response by the Red Cross and similar organizations are labeled with *Response – Community*. In diffused events, tweets that mention response by the Red Cross and similar organizations are labeled with *Response – Formal*.

5.3.6 Third Pass Coding Outcomes

An examination of Twitter communications that take place during mass emergency reveals that the behavior of Twitter users can provide an indication of each event as it occurs on the ground. Additionally, the types of information communicated by users provides insight into how they harness the capabilities of Twitter to provide and guide others to information that may lead to situational awareness.

5.3.6.1 Information Types in the 2009 and 2010 Red River Floods

The three most commonly communicated types of information in both the 2009 and 2010 Red River floods are *Status – Hazard, Advice – Information Space,* and *Preparation,* in that order. Table 5.61 provides a breakdown of the percentage each of these information types represents of each dataset. For the RR09 dataset, over 40% of all information tweeted during the flood is comprised of these three information types, and for the RR10 dataset, that number jumps to over 50%.

	RR09	RR10
Status – Hazard	22.1%	23.4%
Preparation	10.7%	14.9%
Advice – Info Space	8.0%	14.5%

Table 5.61: RR09 and RR10: Most Frequent Information Types

One conclusion I can draw from the figures presented in Table 5.61 is that *Status – Hazard* is a type of information frequently communicated by Twitter users

214

during both of the Red River floods. These tweets include information about river level and areas that were flooding. Residents of the Red River Valley monitor the river level frequently during flood season; they are aware of what river levels imply with regards to when they need to begin flood preparations, when they need to consider evacuation, and when they can safely return to their homes post-flood. The percentage of the dataset represented by the *Status – Hazard* category is very similar for both the RR09 and RR10 datasets.

The second most commonly tweeted information type for the two Red River events is *Preparation*. Every spring, the residents of the Red River Valley are aware of flooding danger, and are frequently placed on flood watch or flood warning. In both 2009 and 2010, flooding was predicted for the region, which allowed for area residents to prepare. Many tweets mention sandbagging, dike/levee building, moving items to upper floors of homes and other buildings, and other preparatory measures. Though the RR09 flood was more damaging and included a record-breaking crest, it is interesting to note that the proportion of *Preparation* tweets is 4.2% higher in the RR10 dataset.

Though the difference in *Preparation* tweets between the RR09 and RR10 datasets is small, it may be an indication of how area residents were confronting the 2010 flood after experiencing the previous years' disaster. The residents of the Red River region were placed on flood alert in late February 2010. They were informed that flooding would be similar to the previous year, which was a devastating event; the 2010 flood alert likely incited fear and worry among area residents. Also in late February, National Guard troops were ordered to begin levee construction, and

community members also began flood preparation efforts. Though there was forewarning before the crest of the 2009 flood, the heightened state of vigilance among Red River Valley communities in 2010 is a potential reason for the higher percentage of *Preparation* tweets in the RR10 dataset. The sociology of disaster literature also indicates that "previous community activity provides a residue of learning that is applied in subsequent situations" (Dynes, 1970). This phenomenon may point to the higher percentage of *Preparation* tweets, in 2010 as area residents learned from and remembered the previous year. An additional factor that explains the proportion difference in *Preparation* tweets in the RR10 dataset is due to sampling decisions. The RR09 dataset is comprised of only local individuals, whereas the RR10 dataset is comprised of users who tweeted situational awareness information, who are not necessarily local individuals. It is possible that in expanding the Twitter user population in the 2010 dataset, more tweets were analyzed that include *Preparation* information.

The *Advice – Information Space* category also shows a proportion difference between the 2009 and 2010 Red River floods. In 2009, 8.0% of the dataset was comprised of *Advice – Information Space* tweets, and in 2010, that number almost doubled to 14.5%. This difference may be attributed to how people used Twitter one year after the record-breaking flood of 2009. Twitter use grows every year, with more tweets being sent and more users joining the service.[36] It is possible that in 2010, more people were aware of Twitter and the role it could play in helping to attain situational awareness; it is also possible the higher proportion of *Advice – Information*

---

[36] http://blog.twitter.com/2011/03/numbers.html

Space tweets in the 2010 dataset is due to sampling decisions, as this dataset includes not only local individual users, but organizations, media outlets and non-locals. However, the doubling of the *Advice – Information Space* category within one year, during a similar flooding event in the same region may serve as an indication of how Twitter is viewed by affected populations and concerned others during mass emergency events as a useful tool for spreading important information.

Information types that are not frequently communicated during each Red River flood differ. In the 2009 Red River dataset, only one type of information is mentioned in fewer than five tweets, which is the *Sheltering* category. Sheltering options were available during the 2009 flood, and a few Twitter users communicated about them. I associate the small number of tweets that include sheltering information to the fact that the flood was predicted, and allowed area residents time to make arrangements regarding a safe place to stay for the duration of the flood.

In the 2010 Red River dataset, three types of information are not represented: *Sheltering*, *Animal Management*, and *Injury*. I attribute the lack of *Sheltering* tweets in the RR10 dataset to the same reasons as those mentioned above for the RR09 dataset. The absence of *Animal Management* tweets may imply that area residents were prepared to take care of their animals and communicating information about animal care was not a priority, it may imply that animal shelters were not available, it may imply that not much information was available with regards to animal management, or a combination of these factors. The lack of *Injury* tweets is not necessarily an indication that no injuries occurred, but it is likely that injury levels

were low enough not to cause concern among those who tweeted about the 2010 flood.

The RR10 dataset also includes information types with fewer than five mentions: *General Area Information*, *General Hazard Information*, *Offer of Help* and *Report of Crime*. Twitter users do not mention much information about the Red River region, nor about floods in general in 2010. The 2009 Red River flood dataset has more information about each of these categories, which I attribute to the serious nature of the 2009 flood, and to the fact that it had been twelve years since the region had experienced flooding of this extend. In 2009, more Twitter users were broadcast information about how floods transpire, and how the geography of the Red River Valley make it a prime flood location. In 2010, it is possible that due to the previous year's flood, Twitter users did not feel it necessary to mention as much general information, but to focus more on information specific to the 2010 situation.

There are also few tweets that mention offers of help to area residents in the 2010 Red River flood dataset. I attribute the low number of *Offer of Help* tweets to the nature of the 2010 flood, and the region in which it took place. Residents of the Red River Valley are frequently placed on flood alert and had experienced a devastating flood one year previously; people in the area are adept at flood preparation tactics and know how to respond when threat of flood is afoot. While some Twitter users broadcast information that offers help regarding the flood situation, my belief is that most users conclude that their fellow community members are aware of the situation and know how to handle any flood-related event that may occur. Another infrequent information type in the RR10 dataset is *Report of Crime*.

218

There were some tweets that mention criminal activity, such as caution against swimming or boating, because those are illegal activities during a flood threat. However, crime was not a common topic among Twitter users in the RR10 dataset, and did not appear to be a concern as compared to river levels and flood preparation.

5.3.6.2 Information Types in the 2009 Oklahoma Fire

Like the Red River datasets, the most frequently tweeted information type in the OK09 dataset is *Status – Hazard*, which comprises 18.9% of the dataset. Similar to residents of the Red River Valley, those living in central and southern Oklahoma during the spring of 2009 were concerned with the state of the hazard they were experiencing. Most tweets in this dataset convey information about the location and intensity of the fire. However, the other two most commonly tweeted information types in the OK09 dataset differ from the flood datasets, which can be attributed to the nature of the event. *Damage* and *Weather* are the next most frequent types of information Twitter users communicate.

The Oklahoma fires damaged many homes and buildings, and area residents communicated about the damage as it was occurring. In addition, weather reports were of great importance to area residents. Knowing wind direction and intensity is crucial during a wildfire, as this information serves as an indication of where fires will likely spread. Many Twitter users communicated about wind, as well as about rain predictions. During the fires, people were hoping for rain to help extinguish the flames.

Those information types that had fewer than five mentions in the OK09 dataset are: *Preparation, Prediction, Historical, Animal Management, General Area*

Information, and *General Hazard Information*. The small number of *Preparation* and *Prediction* tweets in this dataset is not surprising. Some Twitter users mention watering yards or wetting roofs in case the fires were to travel toward their homes, and some mention where they think the fires would spread. However, because the Oklahoma fires were not predicted, and there was minimal action area residents could take to prepare in the event the fires spread toward their property, there was little mention of preparation and prediction. In addition, few Twitter users compare the Oklahoma fire to previous events. Some mention that the fires were the worst disaster some cities had experienced, but for the most part, comparisons were not made between the current fires and earlier disasters.

Tweets that mention how to deal with animals during the fires are also at a minimum in the OK09 dataset. Some Twitter users do communicate that area Animal Control is responding to the fires, and mention of displaced pets also occurs. However, for the most part, the Twitter users in the OK09 dataset are concerned with the state of the fires, the potential for fire spread, and the state of property damage. Additionally, few Twitter users communicate general information about the region experiencing fire. Some mention that areas under threat are in oil producing regions, or that the dry, windy conditions typical of this part of Oklahoma make it easy for fires to start and spread, but for the most part, communications about the characteristics of the region are not something on which Twitter users focus. General information about fires is also not a large topic of communication, though some users provide information about how fires spread, and how dangerous they can be.

I attribute the low number of tweets that mention animal welfare to the fact that many in the region under threat were able to evacuate, and pet or livestock owners were likely to have brought their animals with them, or to have made other safe arrangements. I associate the low numbers of general area information and general hazard information tweets to the fact that the Oklahoma fires, while a damaging and dangerous event, lasted approximately two days. Again, since Twitter users mostly communicated about the state of the fire and where it was headed, providing information about how fires behave in general, or how the regional terrain was an ideal place for fire to spread were not commonly tweeted topics as they was not as immediate as fire status, weather or damage reports.

5.3.6.3 Information Types in the Haiti Earthquake

The most frequently tweeted information types in the Haiti dataset are *Advice – Information Space, Response – Formal*, and *Fatality*. The position of the *Advice – Information Space* category as the most frequently tweeted information type in the Haiti dataset (at 16.2%) is an indication that many Twitter users in the examined dataset were concerned with directing others to information sources during the emergency period of the earthquake, which may indicate that most Twitter users were not located in Haiti.

Sarcevic et al., (2012) performed research on the use of Twitter during the Haiti earthquake to understand how it was used for medical coordination. The researchers were interested in identifying Twitter users who were tweeting from "on the ground" in Haiti, and after a series of sampling decisions, they identified 110 Twitter users who were actively tweeting on the ground in the aftermath of the

221

earthquake, which is .12% of the much larger ~3 million tweet sample with which they started. Though Sarcevic et al. focus on a specific activity that took place via Twitter, their finding that a small proportion of users were located in Haiti speaks to the likelihood that of all Twitter users in the sample used in this dissertation, it is reasonable to suggest that very few were tweeting from the ground.

The presumption that few Twitter users in the dataset used in this dissertation were located in Haiti, but were tweeting from remote locations, speaks to the proportion of *Advice – Information Space* tweets. Much of the information broadcast on Twitter in the aftermath of the earthquake provides information such as links to lists of missing persons, pointers to television or radio stations broadcasting pertinent information, and phone numbers people can call for help. The frequency of *Advice – Information Space* tweets in the Haiti dataset serves as an indication of how users are harnessing the capabilities of Twitter to respond to mass emergencies when they were not able to help in a tangible way.

The second most frequently tweeted category in the Haiti dataset is *Response – Formal*, which represents 10.2% of the dataset. The devastating effects of the earthquake prompted an international response by many government, military, international aid and international volunteer organizations, and many Twitter users tweeted information about these response efforts or the need for response efforts by nation states, non-governmental organizations and similar groups.

The *Fatality* category is the third most common type of information represented in the Haiti dataset, comprising 9.8%. The Haiti earthquake caused the death of hundreds of thousands of people, and Twitter users communicated

information about the deceased. The high proportion of tweets that include information about those killed by the earthquake serves as an indication of the catastrophic nature of the Haiti event.

Three information type categories are not included in the Haiti dataset: *Preparation, Sheltering* and *Prediction*. It is not surprising to see that Twitter users did not communicate information about preparatory measures or prediction, as earthquakes are not predictable; therefore preparation just in advance of impact is not possible.

The earthquake left one million Haitians homeless (BBC, 2010), and many aid organizations, government agencies and other concerned groups responded to the earthquake to provide immediate help to those affected, which included shelter (CNN, 2010). However, the most frequently tweeted types of information in this dataset are concerned with information spread, response efforts and those who were killed in the earthquake. While sheltering is an important and necessary component to disaster response, it is not a topic users communicated about in the Haiti dataset. The reason for the lack of *Sheltering* tweets may be because Twitter users were more concerned with the life-or-death situation in which the Haitian population found itself, and sheltering was seen as a non-immediate concern. It is also possible that there simply was not a lot of sheltering information available for users to communicate, or that they were unaware of information about shelters.

The additional information types in the Haiti dataset that have very little representation are *Animal Management, Weather, General Hazard Information*, and *Status – Personal Property*. There are few tweets concerned with animal welfare.

Though there are many aid organizations are concerned with the state of the animal population after disasters occur, in the Haiti dataset, Twitter users are primarily focused on the state of the human population. Twitter also users did not often communicate about weather. Though worry about rain was a concern approximately one month after the earthquake struck (Rubin, 2010), in the immediate aftermath, little concern for weather conditions was expressed on Twitter. General information about earthquakes was also not a subject many Twitter users communicated about. I attribute this to the fact that the earthquake caused so much damage and death that those who tweeted about it were more concerned with its effects, not with communicating why earthquakes occur or what causes them. The status of personal property was also not frequently mentioned in the Haiti dataset. Though many lost their homes to the earthquake, the status of personal property was not a source of worry for those who tweeted about the event, as the status of the human population was constantly in question.

5.4 Analysis of Qualitative Coding Results Across All Events

The results of the qualitative coding process provide the opportunity to analyze Twitter behavior that takes place during the different mass emergency situations. There are commonalities and differences in the types of information Twitter users communicated during each mass emergency event examined in this dissertation. Across all four datasets, the top three types of information Twitter users communicated comprise between 36.7-52.8% of the entire dataset. This is an indication that though Twitter users communicate about a variety of information, a large portion of their attention is focused on only a few types of information, which

224

differ across each emergency event. The maximum number of information types communicated during an event is twenty-nine, which was during the Haiti earthquake. Table 5.62 provides an overview of the top three information types for each dataset.

	RR09		RR10		OK09		Haiti	
1st Most Frequent Information Type	Status – Hazard	22.1%	Status – Hazard	23.4%	Status – Hazard	18.9%	Advice – Info Space	16.2%
2nd Most Frequent Information Type	Preparation	10.7%	Preparation	14.9%	Damage	13.8%	Response – Formal	10.7%
3rd Most Frequent Information Type	Advice – Info Space	8.0%	Advice – Info Space	14.5%	Weather	9.6%	Fatality	9.8%
Total Percentage of 'R' Tweets Comprised of Top 3 most Frequent Information Types		40.8%		52.8%		42.3%		36.7%

Table 5.62: Top Three Information Types per Mass Emergency Event

A comparison of results from the third pass coding to results of the second pass coding uncovers an interesting finding. In all four datasets, the most highly represented second pass coding category is "S," or tweets that contain only *Social Environment* information. However, in the OK09 dataset, the top three third pass information types are in the *Physical Environment* or *Built Environment* second pass categories. Almost 43% of the OK09 dataset is comprised of information types that fall under the physical and built environment categories, but the majority of the dataset as a whole focuses on the social environment. This finding indicates that though the Twitter users in the OK09 dataset are focused on the state of the fire and how it is impinging on the built environment, societal factors represent a large portion

of their tweet communications; these social environment categories are simply more spread out across many different specific types of information.

The spring of 2009 was a time when two of the datasets were collected—the 2009 Red River flood and the 2009 Oklahoma fires. While both datasets show that *Status – Hazard* was the most frequently tweeted information type, they differ in the second and third most frequently information types, which may serve as an indication of the nature of each event. *Preparation* and *Advice – Information Space* are the second and third most frequently tweeted information types in the RR09 dataset, while *Damage* and *Weather* are the second and third most frequently tweeted information types in the OK09 dataset. The second and third most frequent information types in a fire as compared to a flood indicates that during the fire, Twitter users were concerned with structural damage and weather conditions, both of which were concerning factors during the 2009 Oklahoma fires, as 270 buildings were burned and high winds created worry while promises of rain created calm among area residents. During the 2009 Red River flood, preparatory measures were taken by area residents, local and state government agencies, and these activities are reflected in Twitter communications in the RR09 dataset. The 2009 flood also prompted many to spread information related to the flood, as indicated by the *Advice – Information Space* tweets.

The *Advice – Information Space* third pass category is a top-three category in all but the OK09 dataset. I attribute the lower proportion of *Advice – Information Space* in the OK09 dataset due to the nature of the emergency event. The OK09 event involved the shortest emergency period of all four mass emergencies. People had

226

little, if any, warning about the fires. Though Twitter users in the OK09 dataset provided links to live streaming of the fires and websites that posted wind direction and speed, directing others to these information sources was not as widespread as providing information about the fire status, property damage and weather updates.

However, the emphasis placed on the *Advice – Information Space* category in three of the four datasets examined suggests that Twitter users are employing Twitter as a means to provide situational awareness in an implicit way. Tweets in this category do not always explicitly provide information about how to prepare for a flood, where to find shelter or what areas experienced damage. They do, however, direct others to additional sources that provide such information.

The most frequently tweeted information types for each mass emergency event serve as an indication of each emergency event as it unfolded on the ground, and demonstrate the differences between focalized and diffused disaster situations. In the focalized events of the 2009 Oklahoma fires, and the 2009 and 2010 Red River floods, Twitter users communicate mostly about the status of the hazard agent, e.g. level of the river and size of the fire. In the Haiti event, *Status – Hazard* is an infrequently tweeted information type. Both Red River floods and the Oklahoma fires involved ongoing hazards, i.e. the hazard agent impacted the surrounding environments for an extended period of time. The 2010 Haiti earthquake lasted for thirty seconds. Due to the short duration and high magnitude of the Haiti earthquake, I can speculate that it is unlikely that during similar earthquake events that *Status – Hazard* would be a frequently tweeted information type.

The second and third most frequently communicated information types in Haiti reflect the situation that unfolded on the ground after the earthquake struck. The catastrophic nature of the Haiti earthquake prompted responses from the international community; the days and weeks following impact were spent searching for survivors, providing medical attention, and assessing the death and destruction. Many of these activities were carried out by formal responders, e.g. members of the International Red Cross/Red Crescent Societies, Doctors Without Borders and military members from various countries. Many Twitter users communicated about these efforts by formal response agencies; they provided updates on when responders would arrive, the numbers of responders from particular organizations or agencies, and what tasks responders were performing. This type of information was tweeted about frequently, which I attribute to Twitter users' concern about response efforts and how the hundreds of thousands of affected Haitians were being assisted and supported by the international community.

Over 300,000 people died as a result of the Haiti earthquake (Pan American, 2011). This immense number of fatalities was shocking and upsetting to many, and those who tweeted about the Haiti earthquake frequently mentioned information about those who passed away as a result of the earthquake. Tweets mention the number of deaths known at the time the tweet was broadcast; they also provide information about specific individuals or groups of people known to have died. The amount of tweets that include information about death due to the earthquake is an indication that it was a topic on which Twitter users were focused. The inordinate number of lives lost likely spawned the large amount of Twitter communications that

mention death. Though disaster events regularly occur throughout the world, catastrophic events like Haiti are less common. The high number of fatalities that were the result of the Haiti earthquake was a deeply troubling occurrence, which likely contributed to the large number of tweets that mention fatality.

Looking at how information types are distributed in each dataset provides insight into what Twitter users communicate about, as well as the role Twitter plays in these time- and safety-critical situations. People use Twitter to broadcast advice on where to go to find information, to provide statistics about a hazard event, to talk about the effects of the hazard on the human population, to offer information about preparatory activities and to give weather reports. All of this information may contribute to situational awareness during mass emergency and provide affected populations with information they may otherwise be without.

5.5 Qualitative Coding Outcomes

The three-phase qualitative analysis performed on the 2009 and 2010 Red River floods, the 2009 Oklahoma fires, and the Haiti earthquake datasets reveals what information is communicated via Twitter that may contribute to situational awareness. This analysis points to events—such as property damage and human injury—that take place during each mass emergency. Understanding what information types are most frequently communicated is useful, as it provides an indication of the information on which Twitter users focus. Additionally, having an overall picture of the information available via Twitter is helpful because it allows for researchers to develop an understanding of what activities occur during mass emergencies, regardless of whether those activities are tweeted about often or not.

229

The types of information presented in chapters 4 and 5 can help affected populations make informed decisions in mass emergency situations. However, the process of identifying such information among millions of tweets for the use of automatic methods that use natural language processing (NLP) techniques. The next step of this research involves leveraging VerbNet (Kipper-Schuler, 2006), a lexical resource that provides semantic and syntactic information about classes of English verbs. The approach toward using natural language processing that I focus on in this dissertation is to identify the verbs that tweet authors use to convey situational awareness information. This analysis and its implications are presented in chapter 6.

Chapter 6

Mapping Verbs to Situational Awareness Communications

In chapter 4, I explained the inductive analysis and discourse analytic process used to understand the information Twitter users communicate in mass emergencies. In chapter 5, I presented the results of the annotation process on those tweets that contain information relevant to situational awareness (SA). The findings from chapters 4 and 5 serve as the foundation for this chapter, which focuses on contributions to natural language processing (NLP) techniques that can be developed based on behavioral analyses and understandings of tweet content. The goal of this chapter is to explore options that are useful for identifying tweets that contain the situational awareness information detailed in previous chapters.

6.1 The Value of Linguistic Features in NLP Applications for Situational Awareness

Previous research done by my colleagues and me has focused on natural language processing techniques to identify tweets that contain situational awareness information (Verma et al., 2011). In this research, tweets were manually coded based on whether they contained information that could contribute to situational awareness, if they were written in a subjective or objective style, a formal or informal register, and with a personal or impersonal tone. This hand-annotated data showed that a large proportion of those tweets which contain situational awareness information are written in an objective, impersonal, formal manner (ibid.)

A classifier to identify situational awareness information was then developed using the hand-annotated data. The classifier used basic lexical features such as unigrams, bigrams and part-of-speech. In addition, classifiers to predict subjectivity, register and tone of the tweet were developed. The predicted tags from these classifiers were then used as features for a classifier that could automatically identify tweets that contain information relevant to situational awareness. The classification results of this research showed that the style, register and tone of tweets are features that improve the classification of tweets that contain situational awareness information. These findings demonstrate that training a classifier to identify linguistic features is a useful approach toward the classification of tweets that contain situational awareness information.

6.2 Focus on Verbs

Human language is an evolving, nuanced, contextualized communicative method. Different languages have different grammars that guide how successful communication takes place. In English—which is the only language considered in this dissertation—the *verb* is the central, organizing element of the sentence (Manning & Schütz, 1999). To illustrate the centrality of the verb in English, consider the following sentences, each of which includes only three words, two of which remain the same throughout the list:

1. Anne grows cabbage.
2. Anne cooks cabbage.
3. Anne eats cabbage.

In each of these examples, there is a description of an event taking place that involves *Anne* and *cabbage*. English speakers understand the event described in each sentence:

232

a growing event, a cooking event and an eating event. The sentences rely upon the verb to convey each event meaning. *Anne* and *cabbage* are arguments of the verb, or roles which play a specific part in the sentence. In the examples above, Anne is the agent in the event, meaning that Anne is the *doer*, and cabbage is the *product* or *patient* of each event, meaning that cabbage is the element that is in a particular state or undergoing a change of state in each sentence. However, it is the meaning of the *verb* that dictates how the reader understands what each sentence communicates.

6.2.1 VerbNet

For a machine to classify text according to the message that text conveys, the machine must correctly identify the event that is taking place and the event participants (Palmer et al., 2010). In English, the verb is the element of a sentence that usually communicates the main idea or event (Kipper-Schuler, 2006). The assumption of the research presented here is that verb meaning dictates verb behavior (Levin, 1993). VerbNet (Kipper-Schuler, 2006) is a robust lexical resource based on the verb classes originally created by Levin (1993) that groups together verbs with similar meanings and places them into classes. VerbNet includes 5,879 verbs distributed across 270 classes.[37] The goal of VerbNet is to provide a broad-coverage linguistic resource that provides classes of verbs which have similar meanings, and which exhibit similar semantic and syntactic behavior.

Knowing that verb classes have been shown to be a useful resource for a variety of natural language processing tasks (Kipper-Schuler, 2006), the idea behind using VerbNet verb classes in this research is to exhibit that verbs in the same class

---

[37] http://verbs.colorado.edu/verb-index/index.php

often communicate similar information regarding a mass emergency situation. If a particular VerbNet class is highly correlated with a given information category, then an NLP classifier using VerbNet class membership as a feature would have a higher chance of categorizing the information type correctly than a classifier without VerbNet information.

VerbNet classes include class name, members of the class, one or more syntactic frames, which provide thematic roles and their preferred argument position with regards to the verb, and semantic predicates, which describe the relation between the event and the event participants (Kipper-Schuler, 2006; Palmer et al., 2009). An example of a VerbNet class is shown in Table 6.1.[38]

Class:	contiguous_location-47.8
Members:	border, contain, edge, skirt, straddle, etc.
Frame:	NP V NP
Roles:	Theme1 [+concrete], Theme2 [+concrete]
Example:	"Italy borders France."
Syntax:	Theme1 V Theme2
Semantics: Exist	Contact (During (E), Theme1, Theme2) Exist (During (E), Theme1, (During (E), Theme2)
Subclass:	contiguous_location-47.8-1
Members:	abut, meet, touch, etc.
Frame:	NP V
Example:	"Italy and France touch."
Syntax:	Theme1 (and) Theme2 V
Semantics: Exist	Contact (During (E), Theme1, Theme2) Exist (During (E), Theme1, (During (E), Theme2)

Table 6.1: VerbNet Class contiguous_location-47.8

As the example VerbNet class in Table 1 shows, VerbNet is hierarchically organized, meaning that a class can have a parent class, with up to three child classes nested below it. The child subclass inherits all information from the parent class, and adds

---

[38] NP = noun phrase, V = verb, Theme = Entity being located, E = event

additional information to it. In Table 1, the subclass contiguous_location-47.8-1

inherits all of the roles and frames listed in the parent class, and includes a new

frame: "NP V." However, the frame "NP V" in the child class does not apply to verbs

in the parent class, i.e. one cannot say "Italy and France border."

In addition, verbs often belong to more than one class, because the same verb

may describe different events, i.e. one verb can have multiple senses. For example,

the verb *edge* belongs to the *contiguous_location-47.8* class, and the *fill-9.8* class.

The sense of *edge* in the *contiguous_location-47.8* class means "share a boundary," as

in "The grass edges the sidewalk." The sense of *edge* in the *fill-9.8* class means

"provide with a border," as in "The tailor edged the dress with lace."

There is a potentially wide range of uses that VerbNet can provide for natural

language processing purposes. There have been efforts to create a classifier that

automatically classifies verbs into their VerbNet classes (Brown et al, 2011), because

once a classifier has successfully identified the correct VerbNet class, it has the

ability to access the semantic and syntactic information presented for the given class.

The semantic and syntactic information is helpful for identifying event participants

and the relation those participants have with the verb. In addition, the outputs of

VerbNet classification done either automatically or manually have been used for tasks

including semantic role labeling and word sense disambiguation (Brown et al., 2011).

6.3 Verbs in Situational Awareness Tweets

The goal of identifying verbs used in tweets that convey information relevant

to situational awareness is to provide a resource that demonstrates which VerbNet

classes indicate information relevant to situational awareness. The VerbNet class

information can serve as a linguistic feature that provides a classifier with information to identify tweets that contain situational awareness information. VerbNet classes are useful because the classes provide a list of verbs that may not be present in any of the Twitter data I examined, but which may be used to describe similar information in unseen data. In other words, if a particular VerbNet class is relevant to situational awareness, and a classifier identifies a verb in that class that is used in a previously unseen tweet, then that tweet is more likely to be identified as containing situational awareness information.

In chapter 5, I explained three passes of qualitative coding. The second pass of coding—in which annotators identify tweets that contain information about the social, built and physical environments—underpins the VerbNet analysis described in following sections.

I focus on the second pass category because as I began to analyze the VerbNet class members used in situational awareness tweets, it became clear that regularities in verb use for specific types of information (i.e. third pass categories) were present, but for the most part, using VerbNet classes to indicate specific information was not useful because VerbNet classes are too general to indicate specific information types. However, the result that *did* emerge from the identification of VerbNet classes is that they may be used to identify tweets that fall into the higher-level categories of social, built and physical environments. So, I focused on identifying VerbNet classes that were represented in tweets that contain information about the social, built and physical environments.

6.3.1 Identifying Verbs

To discover if VerbNet classes serve as a useful feature in contributing to the identification of tweets that contain situational awareness information, I manually identified the verbs used in tweets that contain information relevant to situational awareness in all four datasets examined in this dissertation. I created a spreadsheet entitled "VerbNet Analysis" that contained three pages—one each for tweets that contain information about the social environment, the built environment and the physical environment. The next step was to identify those tweets that could be used for this phase of analysis. Tweets were selected based on the following criteria: they had to include a verb, and the verb had to be meaning-bearing. Tweets that contain no verb, e.g. ``Haiti death toll 150000 in capital alone: minister - ABC Online: The GuardianHaiti death toll 150000 in capital al... http://bit.ly/7QaC5`` are not eligible for this phase of the research. Tweets that contain light or copula verbs are also not eligible, e.g. ``I am not in the line of the fire yet.`` Though both of these tweets contain information relevant to situational awareness, the first does not include a verb, and the second uses a copula verb "am" which is not a meaning-bearing verb. "Am" links the subject "I" and the prepositional phrase "in the line of fire yet," but the verb "am" does not provide any semantic meaning on its own.

The process of identifying tweets with meaning-bearing verbs that indicate information about the social, built or physical environment involved analyzing every tweet that contains information relevant to situational awareness in each dataset. If a tweet contains a meaning-bearing verb, e.g. ``RT @redcrossokc The American Red``

Cross are sheltering 30 fire evacuees at the Midwest City Community

Center. #okfire" where the verb "shelter" indicates that a shelter is open and

available at a particular location, then the tweet is copied into the "social

environment" VerbNet Analysis spreadsheet page. This process continues for all

tweets that contain verbs which indicate information relevant to situational

awareness.

The goal of this analysis was to identify verbs used in tweets that contain

situational awareness information. As the analysis progressed, the same verbs would

often appear repeatedly. Once I identified a tweet with a particular verb, I copied that

tweet to the VerbNet Analysis spreadsheet, and I did not copy any additional tweets

that contain the same verb, because the goal of this phase of the analysis was to

identify what verbs were used, not the frequency of those verbs.

Once I analyzed tweets that contain situational awareness information in all

four datasets, and compiled a spreadsheet that listed a tweet using each verb I

identified, I then manually identified the VerbNet class for each verb.[39] This process

involved going to the VerbNet website, searching for each verb, and then identifying

the VerbNet class associated with the verb. Each tweet was labeled with the verb used

to communicate information about the social, built or physical environment, and with

the appropriate VerbNet class.

The purpose of this phase of the research is to look for VerbNet classes that

have high representation in tweets that communicate situational awareness

information about the social, built or physical environments, because these classes are

---

[39] Some verbs do not have VerbNet class assignments, and these are disregarded for purposes of this
analysis.

likely to contain additional verbs not included in my data, but which communicate

similar information and which may be used in unseen data. Therefore, my goal was to

locate verbs used to convey information about the social, built or physical

environment, and then identify those VerbNet classes that are represented with more

than two verbs across all datasets. To illustrate, I turn to the examples below.

1. `140 bottles of breast milk was` **`delivered`** `to the babies of Haiti however they may not get them.`

2. `Province` **`sending`** `ambulances to Fargo - http://tinyurl.com/dz8gf8 #manitobaflood #redriver #flood09`

3. `UPS is` **`shipping`** `for Free TMO to HAITI- it has to be UNDER 50lbs.` **`Ship`** `whtevr u can! if possible!`

In these examples, I focus on the information about the social environment

communicated in the above tweets—in this case, the fact that an event was performed

by an animate agent in which a concrete theme was sent from one place to another.

The segment of the tweet that indicates social environment information in the first

example is: "`140 bottles of breast milk was delivered to the babies of Haiti`". The verb used to convey social environment information is "deliver." The

segment of the second example that implies social environment information is:

`Province sending ambulances to Fargo`. The verb used in this example to

indicate information about the social environment is "send." The segments in the

third example that implies social environment information are: `UPS is shipping`

`for Free TMO to HAITI`; `Ship whtevr u can`. The verb used in this example is

"ship."

The same VerbNet class includes the verbs *deliver, send* and *ship*—the class

send-11.1. Other members of this class include *convey, post, transfer* and *transport*.

239

All of the verbs in the *send-11.1* class have similar meanings, and they all exhibit similar semantic and syntactic behavior. The verb analysis of tweets that contain information relevant to situational awareness reveals that if a verb is in the *send-11.1* class, it is used in a tweet that contains information about the social environment. Tweets from my datasets that include information about the built and physical environments do not include verbs in the *send-11.1* class. In the following section, I outline those VerbNet classes that indicate information that falls under the social, built and physical environment categories.

6.4 VerbNet Classes Indicating Social, Built and Physical Environment Information

Table 6.2 outlines the VerbNet classes whose members have high representation among tweets, meaning more than two class members were used to communicate information about the social, built or physical environment. A total of thirty-nine verbs were identified in nine classes. The VerbNet classes listed in Table 6.2 are exclusive to social, built or physical environment tweets.

	Social Environment	Built Environment	Physical Environment
VerbNet Class	beg-58.2	destroy-44	calibratable_cos-45.6-1
	confront-98		weather-57
	escape-51.1		
	establish-55.5-1		
	fulfilling-13.4.1		
	send-11.1		

Table 6.2: VerbNet Classes Indicative of Social, Built and Physical Environments

6.4.1 VerbNet Classes Represented in Mass Emergency Twitter Data

This section provides an overview of each VerbNet class, and how the classes are relevant to the social, built and physical environments. Below, I provide the name of the VerbNet class, the verbs used in tweets that convey information about the social, built or physical environment, as well as example tweets with the verbs highlighted in bold. It is important to note that many example tweets include more than one verb, but it is the highlighted verb that I have identified as the verb which indicates the social, built or physical environment and which is the member of a VerbNet class that is unique to the social, built or physical environment. Semantic roles (Palmer et al., 2010) used in the explanations below include:

Agent: initiator or doer of an event
Patient: entity in a state or undergoing a change of state
Theme: entity that is moved or changes location
Experiencer: entity that perceives an action, but has no control

6.4.1.1 Social Environment VerbNet Classes

Each of the VerbNet classes below is represented throughout the four datasets examined in this dissertation by tweets that contain more than two of its verb members. The verbs used in these tweets all describe information relevant to situational awareness as situational awareness is found in the social environment.

1. VerbNet class: beg-58.2

    Verbs in tweets: ask, beg, request

    Example 1 (Haiti): `Bodies pile up as Haiti` **`begs`** `for aid`
                       `http://bit.ly/9331nU`

    Example 2 (RR09): `RT @bistrib: Fargo` **`asks`** `for help: AP: A storm`
                      `system moving into North Dakota today increased`
                      `the flood threat.. http://tinyurl.com/d3zntg`

241

Verbs in this class require an animate agent who seeks a concrete or abstract entity. In example 1, the people of Haiti (animate agent) "beg for aid," i.e. they seek help from others. In example 2, "Fargo"—which implies the people of Fargo—"asks for help." Tweets that include verbs in the beg-58.2 class communicate information about a requester making an appeal for some procedure to take place, or for some entity to be provided. There are only five verbs in the beg-58.2 class, three of which are represented in the Twitter data; the additional verbs are: *importune* and *supplicate*.

2. VerbNet class: confront-98

   Verbs in tweets: face, tackle, target

   Example 1 (Haiti): `#News #World Group` **`tackles`** `animal welfare in Haiti after quake http://bit.ly/4FYQT2 (Associated Press)`

   Example 2 (RR09): `Cass County Sheriff Paul Laney says people caught cutting into roads to divert water from their yards will` **`face`** `charges.`

Verbs in the confront-98 class involve an animate agent and a theme. In example 1, "News World Group" is the agent in a tackling event that involves meeting the challenge of animal welfare in the aftermath of a disaster. In example 2, "people" will "face charges," meaning they will encounter allegations that they have engaged in illegal activity. Tweets that include verbs in this class communicate information about confronting a problem or issue. Other verbs in this class which may be used to convey similar information about an animate agent addressing difficulties include *approach, confront* and *undertake*.

3. VerbNet class: escape-51.1

   Verbs in tweets: arrive, depart, escape, leave, return

   Example 1 (OK09): `Residents allowed` **`to return`** `to homes.`
   `Firefighters   adddressing hotspots:`
   `http://tinyurl.com/cf5p4s #OKfire`

   Example 2 (Haiti): `@abigvictory  In Jeremie, Haiti people fr PAP r`
   `arriving` `by the thousands w/the clothes on their`
   `backs.help www.haitianhealthfondation.org`

   This class includes verbs that describe a theme ("residents" or "people") moving *to* some location. Tweets that include using verbs in this class may also include information about moving *from* some location. Tweets that use verbs in the escape-51.1 class include information about the transfer or movement of entities in relation to the disaster situation. Additional verbs in this class that may convey similar information are: *enter, flee* and *vacate*.

4. VerbNet class: establish-55.5

   Verbs in tweets: arrange, open, prepare

   Example 1 (OK09): `RT @redcrossokc: An evacuation center` **`has opened`**
   `at the Midwest City Community Center located at`
   `100 North Midwest City Blvd.  #okfire`

   Example 2 (Haiti): `Businessman` **`arranges`** `burial rites for about 2,500`
   `quake victims - http://bit.ly/9T9AZC #cnn`

   The verbs in this class describe the launching or creation of a process or entity ("an evacuation center" or "burial rites") by an animate agent. In example 1, though no explicit agent is mentioned, the evacuation center was opened by someone or some organization, i.e. there is an implicit animate agent. In example 2, the animate agent is "Businessman." Tweets that use verbs in the establish-55.5

class communicate information about the start of a process or entity that relates to the mass emergency situation. Additional verbs in this class that may be used to convey similar information include: *establish*, *launch* and *organize*.

5. VerbNet class: fulfilling-13.4.1

   Verbs in tweets: issue, provide, supply

   Example 1 (Haiti): `Help- Please `**`Provide`**` Address for people to send`
   `                  stuff to~RT @Reuters: Water, drugs are big need`
   `now                for Haiti http://bit.ly/76ZeRv`

   Example 2 (OK09): `Velma, OK:  14+ homes &  biz have been damaged /`
   `                  destroyed, All streets  seem to be open. School`
   `                  canceled. water boil order `**`issued.`**` #OKFi`

   Verbs in this class describe an agent that presents a theme ("address" and "water boil") to a recipient. Tweets that use verbs in the fulfilling-13.4.1 class include information about fulfilling a need or distributing an entity. Additional verbs in this class that could be used to communicate comparable information are: *present* and *furnish*.

6. VerbNet Class: send-11.1-1

   Verbs in tweets: deliver, post, re-post, send, ship, transport

   Example 1 (RR09): `Province `**`sending`**` ambulances to Fargo –`
   `                  http://tinyurl.com/dz8gf8 #manitobaflood`
   `#redriver                #flood09`

   Example 2 (RR10): `Sandbags `**`delivered`**` ahead of expected Fargo flood`
   `                  –  Yahoo! News http://bit.ly/dgcY98`

   Verbs in this class are used to convey information about the transfer of a concrete theme, such as people or supplies ("ambulances" and "sandbags"), from one location to another. Tweets that include send-11.1-1 verbs provide the reader with

information about the arrival of some form of entity. Additional verbs in this class that may be used in tweets with similar information are: *convey* and *transfer*.

6.4.1.2 Built Environment VerbNet Class

The VerbNet class listed in this section is represented by tweets across all four datasets that contain more than two of its members. The verbs used in these tweets all describe situational awareness information as it is communicated with respect to the built environment, e.g. how facilities and infrastructure are affected by the mass emergency situation.

1. VerbNet class: destroy-44

Verbs used in tweets: damage, destroy, devastate, ruin

Example 1 (RR09): `LA Times: Flood waters '`**`devestate`**`' Fargo scheool http://bit.ly/3S27eu`

Example 2 (OK09): `Oklahoma Winds whip wildfires: Officials say 100 houses` **`destroyed`** `http://tiny.cc/QUX8r`

Verbs in this class express an agent or a cause in the form of a natural hazard agent that lead to a negative change of state to a concrete patient which experiences an undesirable effect from the action performed by the agent or cause. In example 1, "Flood waters" are the agent that affect "Fargo school," in example 2, the implied agent are the "wildfires," which affect "houses." The tweets that include verbs in this class communicate damage to buildings, facilities and other forms of property. Additional verbs from the *destroy-44* class that may be used in tweets to describe similar circumstances of undesirable effects on the built environment include *demolish, level, ravage, raze, ruin, shatter* and *wreck*.

245

6.4.1.3 Physical Environment VerbNet Classes

The VerbNet classes in this section contain verbs used in tweets that communicate information about the physical environment, e.g. about the geographical layout of a region, weather conditions or features of a natural hazard agent.

1. VerbNet class: calibratable_cos-45.6-1

    Verbs used in tweets: climb, drop, increase, lower, move, rise, shift

    Example 1 (RR10): `6:15 AM: Red River Stage at Fargo, ND is 30.53 feet. The river should` **`drop`** `below 30 feet (major stage) in the next 24 hours. #flood10`

    Example 2 (RR09): `RT @JanetStewartCBC : St Andrews firefighters say the Red is` **`rising`** `in spots by 10 inches every hour. 30 houses already flooded. #flood09`

    Verbs in the calibratable_cos-45.6-1[40] class communicate information about a patient ("the river" and "the Red") undergoing a change of state that is measurable on a scale ("below 30 feet" and "by 10 inches"). Tweets that include verbs from this class relay information about an entity changing position in some measurable way. Other verbs in this class which might communicate similar information are: *climb, decline, decrease, grow,* and *increase*.

2. VerbNet class: weather-57

    Verbs used in tweets: blow, gust, rain, roar, snow

    Example 1 (OK09): `20 miles Southwest Lawton area, winds steady at 30+ MPH` **`gusting`** `to 50+ MPH.`

    Example 2 (RR09): `Its` **`snowing`** `in the Fargo Moorhead aria. We might need a little more help.`

---

[40] cos = change of state

246

Verbs in the weather-57 class describe meteorological conditions or conditions of hazard agents as they move through an area. Additional verbs in this class that may communicate similar information include *hail, pour, sleet* and *storm*.

6.5 Evaluating VerbNet Classes

The nine VerbNet classes described in section 6.4 include 195 verbs; 111 verbs in the six social environment VerbNet classes, 25 verbs in the one built environment VerbNet class, and 59 verbs in the two physical environment VerbNet classes. These classes were identified based on an analysis of the approximately four thousand tweets that contain situational awareness information across all four datasets. To test whether VerbNet classes are a useful feature for detecting tweets that contain situational awareness information on tweets that were not previously analyzed, I performed the analyses described below.

6.5.1 VerbNet Classes Tested for Social, Built and Physical Environment Tweet Presence

Once I identified the nine VerbNet classes described in section 6.4, I then calculated the proportion of tweets across all datasets that include information about the social environment, the built environment and the physical environment that contain one of the verbs in the identified classes. Table 6.3 provides an outline of the VerbNet classes whose members have high representation in tweets that contain situational awareness information, and how many verbs are included in all VerbNet classes that are unique to tweets that convey information about the social, built and physical environments.

		Social Environment	Built Environment	Physical Environment
VerbNet Class		beg-58.2	destroy-44	calibratable_cos-45.6-1
		confront-98		weather-57
		establish-55.5-1		
		fulfilling-13.4.1		
		escape-51.1		
		send-11.1		
Total Classes		6	1	2
Total Verbs		111	25	59

Table 6.3: Social, Built and Physical Environment VerbNet Classes, Class Totals and Verb Totals

To see how members of these VerbNet classes were represented across all datasets, I compiled three files of tweets from all four datasets: (1) one file of all tweets that contain information about the social environment, (2) one file of all tweets that contain information about the built environment, and (3) one file of all tweets that contain information about the physical environment. I then compiled three lists of verbs[41]: (1) a list of the 111 verbs included in the social environment VerbNet classes, (2) a list of the 25 verbs in the built environment VerbNet class and (3) a list of the 59 verbs in the physical environment VerbNet classes.

My colleague Amanda Hughes then performed the following three tasks: (1) match tweets that contain information about the social environment to the 111 verbs in social environment VerbNet classes; (2) match tweets that contain information about the built environment to the 25 verbs in the built environment VerbNet class; (3) match tweets that contain information about the physical environment to the 59 verbs in the physical environment VerbNet classes. This process is outlined in Figure 6.1.

---

[41] Verb lists included the conjugated forms of each verb. For example, for the verb cross, the string match accounted for cross, crosses, crossing and crossed.
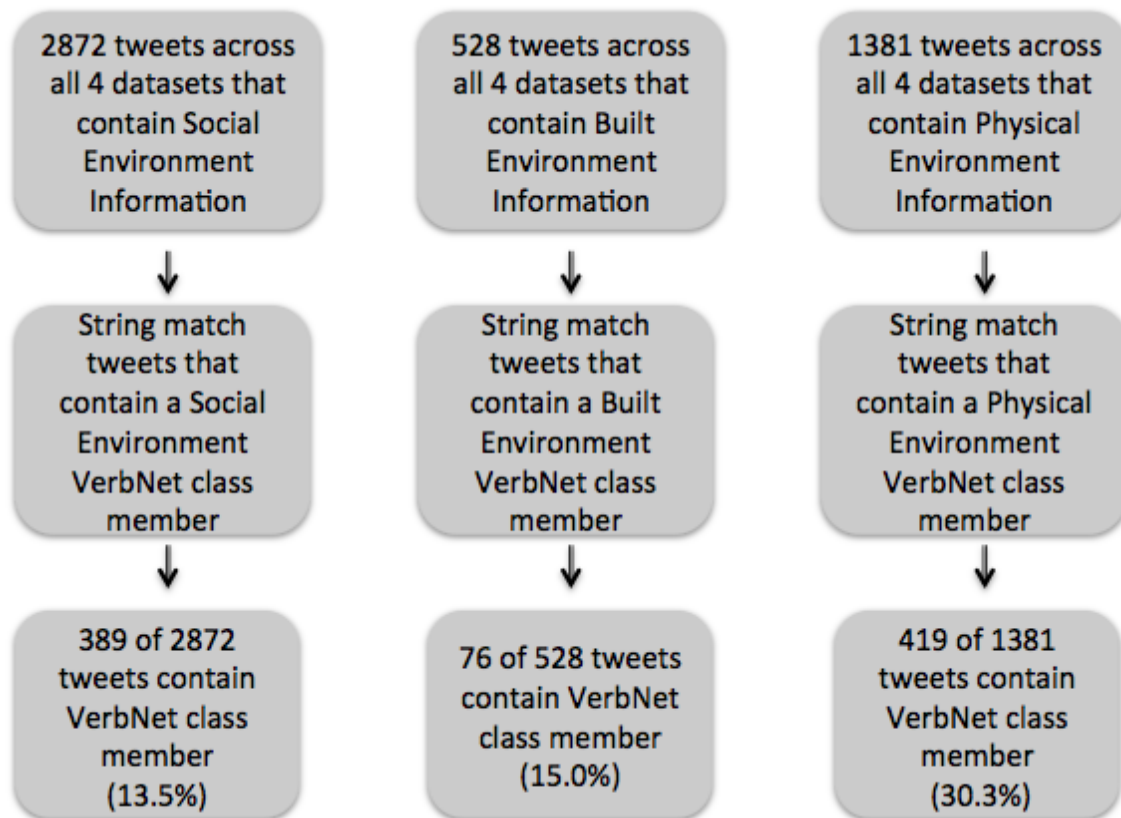
2872 tweets across all 4 datasets that contain Social Environment Information	528 tweets across all 4 datasets that contain Built Environment Information	1381 tweets across all 4 datasets that contain Physical Environment Information
↓	↓	↓
String match tweets that contain a Social Environment VerbNet class member	String match tweets that contain a Built Environment VerbNet class member	String match tweets that contain a Physical Environment VerbNet class member
↓	↓	↓
389 of 2872 tweets contain VerbNet class member (13.5%)	76 of 528 tweets contain VerbNet class member (15.0%)	419 of 1381 tweets contain VerbNet class member (30.3%)

Figure 6.1: Process of string matching tweets that contain a VerbNet class member

Since Hughes performed a string match in the process shown in Figure 6.1, I manually analyzed each tweet to ensure that verbs were used to communicate an event in the social, built or physical environment. In reading and checking tweets, I noted that some tweets were a match because they contained one of the verb members used as a nominalization. Nominalization occurs when a noun is formed from another part of speech, often verbs. Nominals can work substantively as the head of a phrase; they often denote an event described in a tweet. In the case of the data I examine here, some verbs are used as zero-derived nominals, meaning the nominal takes the same

249

form as a verb. For example, "damage" can be either a verb, as in "the earthquake damaged the buildings" or it can be a nominal, as in "the damage to the buildings was bad." In both cases, the event being communicated is one of damage. Therefore, if a verb is acting as a nominal, and that nominal is responsible for conveying the event described in the tweet, I counted the tweet as containing one of the VerbNet class members. The results of the process shown in Figure 6.1 are shown in Table 6.4.

Information Communicated in Tweet	Total number of tweets that contain a VerbNet class verb/Total number of tweets	Proportion of Tweets that Contain a VerbNet Class Verb
Social Environment	389/2872	13.5%
Built Environment	76/528	15.0%
Physical Environment	419/1381	30.3%

Table 6.4: Social, Built and Physical Environment Tweets - VerbNet Class Analysis

The results in Table 6.4 show the proportion of tweets that communicate information about the social, built or physical environment which contain one of the verbs from one of the VerbNet classes I identified as indicating information about the social, built or physical environments. These results provide a starting point for understanding how to move forward regarding the use of VerbNet classes as a feature for the classification of tweets that communicate situational awareness information. A majority of tweets do not contain one of the verbs in the identified VerbNet classes, which indicates that additional features are necessary to classify tweets according to the social, built or physical environment.

However, the results in Table 6.4 also show that tweets that include information about the physical environment have a higher presence than the other two categories; almost one third of tweets that contain information about the physical environment include a VerbNet class member. This higher proportion of tweets that

250

contain a VerbNet class member in the physical environment category is because

78.4% of the physical environment tweets are from the 2009 Red River flood dataset

and the 2010 Red River flood dataset.

In both of the Red River datasets, Twitter users often communicate about the

status of the hazard, which includes the level of the river. One of the physical

environment VerbNet classes I identified is *calibratable_cos-45.6*, which contains

verbs that describe events of value change according to some scale. In both Red River

datasets, many tweets communicate information about the river "rising," "falling,"

"dropping" and "increasing," which are all verbs in the *calibratable_cos-45*.6 class.

Due to the circumstances around the floods in both 2009 and 2010, and the high

proportion of tweets that mention the river level, this result suggests that for future

floods of the same nature (i.e. predicated floods in areas equipped with river level

monitoring equipment), the *calibratable_cos-45*.6 VerbNet class is a useful feature

for identifying tweets that communicate information about the physical environment.

6.5.2 VerbNet Classes Tested on Unseen Data

The next test I performed to see if VerbNet classes are a useful feature for

tweet categorization was to use the same classes, but to test them on previously

unseen data.

The Haiti dataset used for this dissertation is a sample of 4,005 tweets randomly

selected from a larger dataset of 230,409 tweets. First pass coding results (reported in

chapter 5) reveal that 23.5% of tweets from this sample of 4,005 contain information

relevant to situational awareness. An illustration of the Haiti dataset sampling and

annotation process is shown in Figure 6.2. This figure shows that from a dataset of

approximately 4 million tweets, my colleague Sudha Verma removed tweets that

contained one of the stop words listed in section 3.2.3.2.1, and all duplicate tweets.

This left approximately 1.7 million tweets. Verma then took a random sample of

300,000 tweets, which was a manageable size to process through the situational

awareness classifier colleagues and I developed based on linguistic features (Verma

et al., 2011). This resulted in a dataset of 230,409 tweets. I then took a random

sample of 4,005 tweets to use for manual annotation, which is the Haiti dataset that

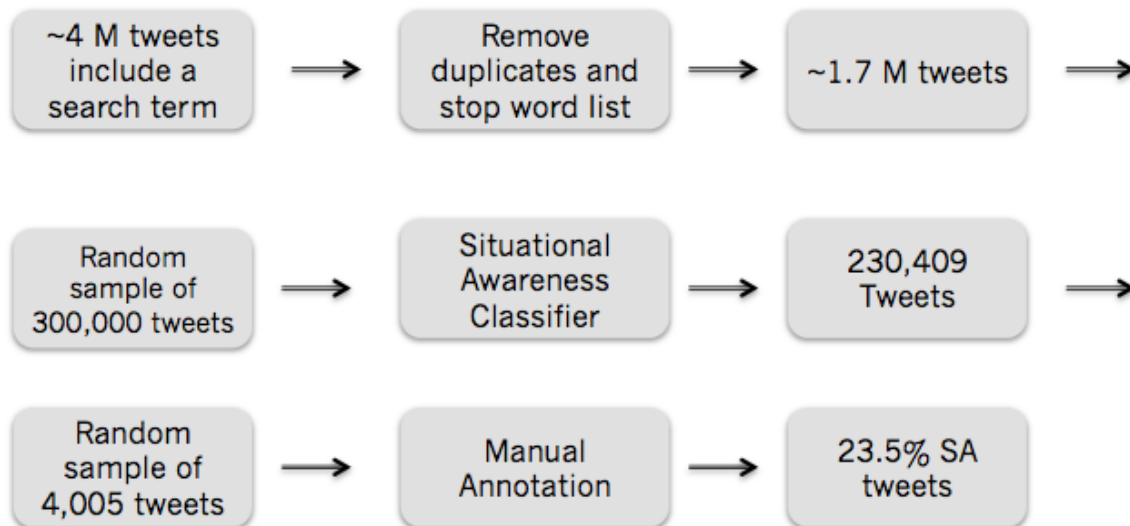went through the first, second and third passes of coding.

Figure 6.2: Haiti Data Sampling and First Pass Coding Process

Hundreds of thousands of tweets sent during the Haiti earthquake were not

analyzed in chapters 4 and 5, which presents the opportunity to test the validity of

VerbNet classes on unseen tweets. For this phase of testing, I sought to determine if

the VerbNet classes were useful for identifying tweets that contain situational

awareness information of any sort. I asked my colleague Ali Alzabarah to match

tweets in the remaining 226,404 tweets not included in my original sample of 4,005

Haiti tweets that contained one of the 195 verbs in the nine VerbNet classes I

identified above.[42] This returned a set of 29,289 tweets. I took a random sample of

4,000 tweets from this dataset of 29,289 tweets, and annotators Preciado and

Vollmann annotated them with first pass codes, i.e. annotators coded tweets as off-

topic ("O"), on-topic and relevant to situational awareness ("R") or on-topic and not

relevant to situational awareness ("N"). After this annotation, I adjudicated the data

and the results show that of this 4,000 tweet dataset, 1,305 tweets, or 32.6% of this

new Haiti dataset, include situational awareness information.[43] To illustrate the results

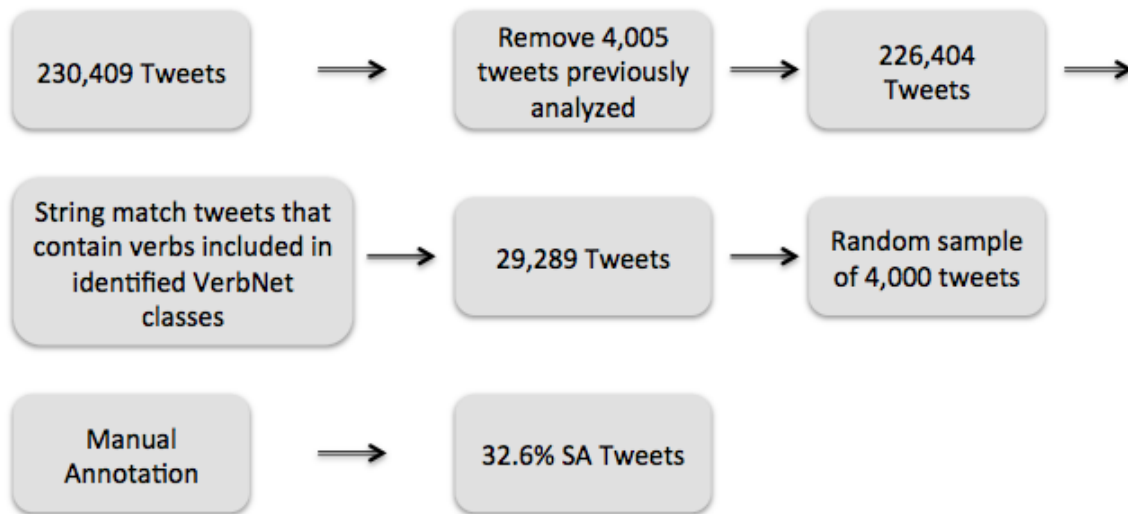of this process, I turn to Figure 6.3:

Figure 6.3: Analysis of Unseen Haiti Data

By starting with previously unanalyzed dataset of 226,404 tweets, then

performing a string match to locate those tweets that contain one of the 195 verbs

---

[42] Alzabarah filtered out the 29,289 tweets that contained one of the verbs in the nine identified
VerbNet classes; no part-of-speech tagging was performed. This led to the selection of tweets that
contained the same strings as those of the verbs listed in the verb list, but which were not in fact verbs.
E.g. tweets that include the phrase "Red Cross" are a match because they include the word "cross,"
which is also a verb listed in my VerbNet list. Tweets such as this were discarded.
[43] I approach the use of nominals in the same way as described in section 6.5.1.

included in the identified VerbNet classes, and then annotating a random sample of comparable size to the original Haiti dataset examined in chapters 4 and 5, this analysis shows that the proportion of tweets that contain situational awareness information rose by 9.1%, which indicates that using VerbNet classes is a useful approach toward identifying tweets that communicate situational awareness information.

6.5.2.1 Identification of New Verbs

An additional aspect of this phase of analysis involved identifying each verb used in those tweets that contain situational awareness information. My original analysis revealed that thirty-nine verbs out of the 195 verbs in the nine identified VerbNet classes were used in tweets that contain situational awareness. The subsequent analysis I did on previously unseen tweets shows an additional twenty-nine verbs of those 195 were used in tweets that convey situational awareness. By using the VerbNet classes as a starting point, previously unseen tweets that contain situational awareness information were identified because they contain one of the verb members, which is a promising result and shows that the information contained in VerbNet classes is useful for identifying situational awareness tweets. An outline of the VerbNet classes for the social, built and physical environments are shown in Tables 6.5-6.7. Each table provides the VerbNet class name. Below the class name, the verbs identified in my original analysis of tweets from all four datasets are listed as "previously identified verbs." The following row provides a list of "newly identified verbs." These are verbs not previously identified, but which are members of

the VerbNet classes, and which were present in tweets in the 4,000 tweet set of

previously unseen Haiti data.

VerbNet Classes and Verbs Used in Tweets that Contain Situational Awareness Information	Social Environment
	1. VerbNet Class: beg-58.2
	Previously Identified Verbs: ask, beg, request
	Newly Identified Verbs: no new verbs
	2. VerbNet Class: confront-98
	Previously Identified Verbs: face, tackle, target
	Newly Identified Verbs: combat, confront, handle
	3. VerbNet Class: escape-51.1
	Previously Identified Verbs: arrive, depart, escape, leave, return
	Newly Identified Verbs: cross, descend, emerge, enter, fall, flee
	4. VerbNet Class: establish-55.5
	Previously Identified Verbs: arrange, open, prepare
	Newly Identified Verbs: establish, launch, mount, organize
	5. VerbNet Class: fulfilling-13.4.1
	Previously Identified Verbs: issue, provide, supply
	Newly Identified Verbs: present
	6. VerbNet Class: send-11.1
	Previously Identified Verbs: deliver, post, re-post, send, ship, transport
	Newly Identified Verbs: convey, dispatch, forward, hand, pass, smuggle, transfer

Table 6.5: Social Environment VerbNet Classes, previously identified verbs, newly identified verbs

VerbNet Classes and Verbs Used in Tweets that Contain Situational Awareness Information	Built Environment
	1. VerbNet Class: destroy-44
	Previously Identified Verbs: damage, destroy, devastate, ruin
	Newly Identified Verbs: demolish, level, ravage, shatter, wreck

Table 6.6: Built Environment VerbNet Classes, previously identified verbs, newly identified verbs

VerbNet Classes and Verbs Used in Tweets that Contain Situational Awareness Information	Physical Environment
	1. VerbNet Class: calibratable_cos-45.6-1
	Previously Identified Verbs: climb, drop, increase, lower, move, rise, shift
	Newly Identified Verbs: grow, surge, soar
	2. VerbNet Class: weather-57
	Previously Identified Verbs: blow, gust, rain, roar, snow
	Newly Identified Verbs: no new verbs

Table 6.7: Physical Environment VerbNet Classes, previously identified verbs, newly identified verbs

6.5.2.2 Verbs in Tweets That Do Not Contain Information Relevant to Situational Awareness

An additional analysis I performed was to identify verbs used in tweets that do not contain information that contributes to situational awareness. The results of this analysis show that many tweets contain the same verbs as the 195 verbs represented in the nine VerbNet classes listed in Table 6.3. However, these tweets are used in different senses, and are therefore members of different VerbNet classes—classes that were not identified as having high representation across the four datasets I examine, and which are not unique to tweets that contain information about the social, built or physical environment. Examples of such tweets are listed here:

1. `@missmiranda oh. i'm sorry. i'm bouncing all over the place tonight but i'm` **`devastated`** `by Haiti`

   VerbNet Class: amuse-31.1

2. `@misscrumb haiti has always needed help. with the love being` **`showered`** `on her now, she can rise to greater glory. we believe this right.`

   VerbNet Class: spray-9.7-1

3. `in light of the haiti disaster, you can support the american red cross by` **`transferring`** `starpoints. http://bit.ly/8pedla`

   VerbNet Class: contribute-13.2-2

In the first example, the tweet author writes that s/he is "devastated" by Haiti. The sense of *devastate* in this tweet—to convey an emotional state—is the sense described by the VerbNet class amuse-31.1, which is not one of the previously identified classes that indicate information relevant to situational awareness. The verb *devastate* is also a member of the destroy-44 VerbNet class, and as *devastate* is used in that class, it indicates the destruction of concrete items, which is an indication of information relevant to situational awareness.

In the second example, the tweet author mentions that there is "love being showered" on Haiti. The use of the verb *shower* in this tweet indicates the profuse heaping of an abstract noun, which places it into the spray-9.7-1 VerbNet Class; this is not a class I identified in Table 6.3. *Shower* is also a verb in the weather-57 class, which contains verbs that convey information about weather conditions and precipitation. When *shower* is used to communicate information about the weather, it indicates situational awareness information.

In the third example, the Twitter user indicates that one can support the American Red Cross by "transferring starpoints." The use of *transfer* indicates that an abstract theme is being moved from one location or context to another, which is the sense of *transfer* in the contribute-13.2-2 VerbNet Class. I did not identify the contribute-13.2-2 VerbNet Class as one of the nine classes that contain verb members used in tweets that contain situational awareness information. *Transfer* is also a member of the send-11.1 class, which is listed in Table 6.3. As *transfer* is used in this sense, it indicates the movement of a concrete theme from one location to another,

which is the sense used in tweets that communicate information that contributes to situational awareness.

The analysis I outline in this section points to the advantage of using VerbNet class names as a feature for identifying situational awareness information communicated in tweets. If classifiers can successfully assign the VerbNet class name to a tweet, and the identified class is one whose verb members are used in communication information relevant to situational awareness, the classifier has a better chance of predicting tweets that contain actionable, tactical information that people can use to make informed decisions in situations of mass emergency.

6.6 Resources for Natural Language Processing

VerbNet is a well-developed resource that provides researchers with a lens through which to analyze human language. VerbNet has been used in natural language processing applications for a variety of purposes; the objective of this chapter is to offer a method for harnessing the capacities of VerbNet by linking VerbNet classes to empirically-based understandings of language use in Twitter communications that contribute to situational awareness.

The results of the analyses presented in this chapter show that using VerbNet classes as a source for identifying situational awareness tweets is promising. The idea behind the analyses presented in this chapter is to lay the foundation for natural language processing applications that classify text according to categories of information that describe mass emergency events. The empirical analysis of Twitter communications broadcast during mass emergency illustrates how VerbNet classes provide a useful feature for tweet classification.

However, VerbNet classes are not the only potentially useful feature for predicting whether tweets contain situational awareness information. The analyses I present above show that using VerbNet classes as a feature is encouraging, but other features are needed to identify tweets that contain situational awareness information, as not all tweets that contain situational awareness information use one of the verb members in the nine identified VerbNet classes. In addition, more research in this area will involve using the semantic and syntactic information contained in each VerbNet class to identify event participants, which can lead to more fine-grained categorization of tweets.

Additional features that my colleagues and I consider are named entities and semantic roles. Ongoing annotation efforts in Project EPIC focus on annotating tweets with named entities such as *person, organization, facility, location,* and *artifact* (Corvey et al., 2010; 2012). These labels are used to annotate spans within tweets that refer to real-world entities. For example, in this segment taken from a tweet sent during the 2009 Red River flood: "`Province sending ambulances to Fargo,`" "Province" is annotated with *organization*, "ambulances" is annotated with *artifact*, and "Fargo" is annotated with *location*. Semantic roles, which are defined here according to PropBank guidelines (Palmer et al, 2010), are also potentially useful features. For the same tweet segment listed above, "Province" is the *agent*, or doer of the event, "ambulances" is the *theme*, or entity moving from one location to another, and "Fargo" is the *destination*, or final location for the entity being moved.

Named entities and semantic roles abstract over specific mentions of event participants within tweets to provide classifiers with higher-level linguistic features.

259

Many tweets that communicate situational awareness information do not contain one of the verbs in the identified VerbNet classes, and the information provided with named entities and semantic roles can serve as features that classifiers can use to identify situational awareness information in the absence of such a verb. In addition, for tweets correctly identified as containing information relevant to situational awareness, named entities and semantic roles can provide classifiers with additional information to classify these tweets into the social, built and physical environment categories, and into specific information type categories.

During mass emergency events, millions of tweets may be broadcast that refer to the event. Previous chapters in this dissertation describe the information relevant to situational awareness that Twitter users communicate. Locating such tweets among the millions that are broadcast is a task that requires computational methods. The development of these computational methods rests on knowledge of how Twitter users communicate situational awareness information. This chapter has provided insight into how an existing lexical resource can inform natural language processing applications that can identify the tactical, actionable information that exists among the mass of Twitter communications broadcast during mass emergency events.

Chapter 7

Conclusion

This chapter provides an overview of the findings and contributions of this dissertation. I use discourse analysis to examine Twitter communications broadcast during four mass emergency events; perform a three-pass qualitative coding process that reveals what information types contribute to situational awareness are broadcast by Twitter users during each mass emergency; and identify verbs and VerbNet classes used in tweets that communicate information relevant to situational awareness. These findings lead to the conclusions and the proposed future research explained below.

7.1 Findings and Contributions

The findings that emerged from this research are described in this section.

(1) ***An outline and definitions of the information types communicated via Twitter in mass emergency situations***.[44] Based on the discourse analysis of Twitter communications broadcast during four mass emergency events, I identified thirty-two specific types of information that contribute to situational awareness. Subsequent analysis of the sociology of disaster literature, government documents and additional research on the use of Twitter in mass emergency uncovered three additional types of information. The information types are listed in Tables 7.1-7.3:

---

[44] No Twitter communications were identified that include information about Environmental Impact. I include it in my final outline of information types communicated via Twitter because based on government publications and Environmental Protection Agency literature, concerns about the impact on the environment are a consideration during mass emergency situations.

Social Environment	
Advice – Information Space	Preparation
Animal Management	Report of Crime
Caution	Request for Help
Evacuation	Request for Information
Fatality	Rescue
Feeding/Hydration	Response – Community
General Population Information	Response – Formal
Immediate Recovery	Response – Miscellaneous
Injury	Response – Personal
Medical Attention	Sheltering
Missing	Status – Community/Population
Offer of Help	Status – Personal

Table 7.1: Social Environment Information Types

Built Environment	
Damage	Status – Personal Property
Status – Infrastructure	Status – Public Property

Table 7.2: Built Environment Information Types

Physical Environment	
Environmental Impact	Prediction
General Area Information	Status – Hazard
General Hazard Information	Weather
Historical	

Table 7.3: Physical Environment Information Types

(2) ***A discourse analytic description of tweets that communicate information relevant to situational awareness.*** The process of identifying the information types broadcast via Twitter in mass emergency involved an analysis of the language Twitter users employ in their communications about the event. This analysis involved close attention to context, background knowledge and linguistic phenomena that Twitter users rely on when communicating the information types presented above.

(3) ***A resource for the qualitative coding of Twitter communications broadcast during mass emergency.*** The identification of information types and the coding process described in chapters 4 and 5 serve as an example for similar

research going forward. The cyclical process of identifying tweets that contain situational awareness information, identifying what that information specifically communicates, and then applying those information types to tweets involved a three-pass coding process based in empirical, inductive analysis of tweets. The results of each of these passes of coding serve to categorize tweets at different levels of granularity.

(4) ***A description of what information types are communicated during four mass emergency events.*** The results of the three-pass qualitative coding process show what information Twitter users in each dataset attend to, and how that information differs or is similar across each emergency event. The findings from the qualitative coding process suggest that the information Twitter users communicate is indicative of each mass emergency event.

(5) ***A resource for the development of natural language processing techniques that classify tweets based on the content of the tweet message.*** The three-pass coding process of tweets described in chapter 5; the identification of verbs used in tweets that contain situational awareness information; and the subsequent identification of VerbNet classes that have high representation in those tweets serves as a foundation for how VerbNet can be used as a feature for natural language processing applications to identify tweets that communicate tactical, actionable information.

7.2 Significance of Findings

The results of this research demonstrate that people use Twitter to communicate information that contributes to situational awareness in times of mass emergency. The detailed analysis of tweet content uncovers the ways in which Twitter users rely on linguistic phenomena, context, background and commonsense knowledge to convey many types of information that can help affected populations make more informed decisions when time is limited and safety is in question. This research also provides an overview of the information Twitter users communicate in different mass emergency situations. Additionally, this dissertation provides an approach for the development of natural language processing tools and resources to aid in the effort of automatically identifying tweets that contain situational awareness information.

7.3 Future Research

Future directions that begin with the research presented in this dissertation include the expansion of the types of mass emergency events analyzed. Three types of mass emergency events are analyzed in this dissertation—flood, wildfire and earthquake. The analysis of Twitter communications broadcast during additional types of events will uncover what Twitter users communicate during hurricane, tornado and additional mass emergencies caused by natural hazards. Results of these future analyses will serve to verify, expand and/or alter the existing offerings of this research.

In addition, this research introduces directions for the development natural language processing techniques. Next steps will involve creating classifiers that use VerbNet classes as a feature, along with additional features such as named entities and semantic roles. My colleagues and I can also explore possibilities for the development of natural language processing techniques that are specific to particular types of disaster events, e.g. classifiers that focus on focalized or diffused events; flood situations or earthquake situations, and similar specific uses. Finding the best approach toward the automatic identification of situational awareness information communicated in tweets is a task that will involve further training and testing of classifiers.

Bibliography

Aguirre, B.E. & Tierney, K.J. (2001). Testing Shibutani's Prediction of Information Seeking Behavior in Rumor. Preliminary Paper #307. University of Delaware, DRC.

Bach, K. (1998). Speech acts. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved from: http://www.rep.routledge.com/article

Barnes, S.J. & Böhringer, M. (2009). Continuance Usage Intention in Microblogging Services: The Case of Twitter. In *Proc. ECIS 2009*, 8-10.

Bayerl, P.S., Lüngen, H., Gut, U. & Paul, K.I. (2003). Methodology for Reliable Schema Development and Evaluation of Manual Annotations. In *Proc. K-CAP '03*.

BBC News. (2010, February 10). Haiti will not die, President Rene Preval insists. BBC News. Retrieved from http://news.bbc.co.uk/2/hi/americas/8511997.stm

Bikel, D.M, Schwartz, R. & Weischedel, R.M. (1999). An Algorithm that Learns What's in a Name. *Machine Learning (34)*, 211-231.

Brown, S.W., Dmitriy D. & Palmer, M. (2011). VerbNet class assignment as a WSD task. In *Proc. IWCS 2011*, 85-94.

Cable News Network (CNN). (2010, January 13). Aid Begins Flowing to Haiti. *CNN World*. Retrieved from: http://articles.cnn.com/2010-01-13/world/haiti.aid.response_1_port-au-prince-airport-haiti-world-bank?_s=PM:WORLD

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20(1)*, 37–46.

Conflict, Humanitarian and Security Department (CHASE) - Department for International Development. (2010, January 23). Haiti Situation report: Haiti Earthquake 2010, 1800 hours 23 January 2010. *reliefweb*. Retrieved from: http://reliefweb.int/node/342689.

Corvey, W.J., Vieweg, S., Rood, T. & Palmer, M. (2010). Twitter in Mass Emergency: What NLP Techniques Can Contribute. In *Proc. NAACL HLT Workshop on Computational Linguistics in a World of Social Media*, 23-24.

Corvey, W.J., Verma, S., Vieweg, S., Palmer, M. & Martin, J.H. (2012). Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events. In *Proc. Language Resources and Evaluation (LREC)*.

Crystal, D. (1997). *A Dictionary of Linguistics and Phonetics*. Malden, MA: Blackwell Publishing.

Department of Homeland Security. (2011). National Preparedness Goal. *FEMA.gov*. Retrieved from: http://www.fema.gov/library/viewRecord.do?id=3773

del-Teso-Craviotto, M. (2006). Language and sexuality in Spanish and English dating chats. *Journal of Sociolinguistics, 10(4)*, 460-480.

Drabek, T. E. (1996). The Social Dimensions of Disaster. *FEMA Emergency Management Higher Education Project College Course Instructor Guide*. http://training.fema.gov/EMIWeb/edu/completeCourses.htm

Dynes, R.R. & Quarantelli, E.L. (1968). What Looting in Civil Disturbances Really Means. *University of Delaware Research Center*, Article #15.

Dynes, R. (1970). *Organized Behavior in Disaster*. Heath.

Dynes, R. (1994). Community Emergency Planning: False Assumptions and Inappropriate Analogies. *International Journal of Mass Emergencies and Disaster, 12(2)*, 141-158.

Dynes, R. & Tierney, K.J. (1994). *Disasters, Collective Behavior, and Social Organization*. Newark, NJ: University of Delaware Press.

Endsley, M.R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems *Human Factors, 37(1)*, 32-64.

Endsley, M.R., Bolté, B., & Jones, D.G. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design*. New York, NY: Taylor & Francis.

Federal Emergency Management Agency (FEMA). (2009, March 15). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat031509.shtm

Federal Emergency Management Agency (FEMA). (2009, March 17). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat031709.shtm

Federal Emergency Management Agency (FEMA). (2009, March 18). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat031809.shtm

Federal Emergency Management Agency (FEMA). (2009, March 23). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat032309.shtm

Federal Emergency Management Agency (FEMA). (2009, March 24). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat032409.shtm

Federal Emergency Management Agency (FEMA). (2009, March 28). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat032809.shtm

Federal Emergency Management Agency (FEMA). (2009, March 29). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat032909.shtm

Federal Emergency Management Agency (FEMA). (2009, March 30). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat033009.shtm

Federal Emergency Management Agency (FEMA). (2009, April 5). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat040509.shtm

Federal Emergency Management Agency (FEMA). (2009, April 9). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat040909.shtm

Federal Emergency Management Agency (FEMA). (2009, April 11). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat041109.shtm

Federal Emergency Management Agency (FEMA). (2010, February 5). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2010/nat020510.shtm

Federal Emergency Management Agency (FEMA). (2010, March 8). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2010/nat030810.shtm

Federal Emergency Management Agency (FEMA). (2010, March 15). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2010/nat031510.shtm

Federal Emergency Management Agency (FEMA). (2010, March 19). Situation report. *FEMA.gov*. Retrieved from: http://www.fema.gov/emergency/reports/2009/nat031910.shtm

Federal Emergency Management Agency (FEMA). (2010, April 10). Disaster Recovery Report – 1907. *North Dakota – Flooding, FEMA-1907-DR*. Retrieved from*:* www.fema.gov/pdf/news/pda/1907.pdf

Federal Emergency Management Agency (FEMA). (2010, April 19). Disaster Recovery Report – 1900. *Minnesota Flooding – FEMA-1900-DR*. Retrieved from: www.fema.gov/pdf/news/pda/1900.pdf

Fritz, C.E. & Marks, E.S. (1954). The NORC Studies of Human Behavior in Disaster. *Journal of Social Issues* 10(3), 26-41.

Gabbard, R., Marcus, M. & Kulick, S. (2006). Fully Parsing the Penn Treebank. In *Proc. of the Human Language Technology Conference of the North American Chapter of the ACL*, 184–191.

Gumperz, J.J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.

Hagar, C. (2006) Using research to aid the design of a crisis information management course. Presented at *Association of Library & Information Science Educators SIG*: Multicultural, Ethnic & Humanistic Concerns (MEH), Information Seeking and Service Delivery for Communities in Disaster/Crisis. San Antonio, TX.

Hagar, C., & Haythornthwaite, C. (2005). Crisis, farming and community. *Journal of Community Informatics, 1*, 41-52.

Harrald, J. & Jefferson, T. (2007). Shared Situational Awareness in Emergency Management Mitigation and Response. In *Proc. HICCS 2007*.

Herring, S.C. (2001). Computer-mediated Discourse. In: D. Schiffrin, D. Tannen, and H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (612-634). Oxford: Blackwell Publishers.

Honeycutt, C. & Herring, S. (2009). Beyond Microblogging: Conversation and Collaboration via Twitter. In *Proc. HICSS 2009,* 1-10.

Horn, L. (1999). Implicature. In *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: MIT Press.

Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006). OntoNotes: The 90% Solution. In *Proc. of the Human Language Technology Conference of the North American Chapter of the ACL,* 57-60.

Huberman, B.A., Romero, D.M. & Wu, F. (2009). Social Networks that Matter: Twitter under the Microscope. *First Monday, 14(1),* January 5, 2009.

Hudson, G. (2000). *Essential Introductory Linguistics.* Malden, MA: Blackwell Publishing.

Hwang, J.D., Bhatia, A., Bonial, C., Mansouri, A., Vaidya, A., Xue, N., & Palmer, M. (2010). PropBank Annotation of Multilingual Light Verb Constructions. In *Proc. Fourth Linguistic Annotation Workshop*, 82–90.

Java, A., Song, X., Finin, R. & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage Communities. In *Proc. WEBKDD and SNA-KDD 2007 workshops*.

Johnstone, B. (2008). *Discourse Analysis.* Malden, MA: Blackwell Publishing.

Kiesler, S., Siegel, J. & McGuire, T.W. (1984). Social Psychological Aspects of Computer-Mediated Communication. *American Psychologist, 39(10)*, 1123-1134.

Kipper-Schuler, K. (2006). *VerbNet: a broad-coverage, comprehensive verb lexicon.* PhD Thesis, University of Pennsylvania.

Kreps, G. A. & Bosworth, S.L. (1994). *Organizing, Role Enactment, and Disaster: A Structural Theory.* Cranbury, NJ: Associated University Presses.

Manning, C.D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press.

Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge, UK: Cambridge University Press.

Matheus, C.J., Kokar, M.M. & Baclawski, K. (2003). A Core Ontology for Situation Awareness. In *Proc. International Society of Information Fusion* (ISIF). 545-552.

McNutt, M. (2009, April 23). Federal help sought for Oklahoma wildfire victims. *NewsOK.* Retrieved from: http://newsok.com/federal-help-sought-for-oklahoma-wildfire-victims/article/3363783

Mendoza, M., Poblete, B. & Castillo, C. (2010). Twitter Under Crisis: Can we trust what we RT? KDD Social Media Analytics Workshop, In *Proc. SOMA 2010*.

Mileti, D.S. (1999). *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Washington DC: John Henry Press.

Mileti, D.S., & Peek, L. (2000). The social psychology of public response to warnings of a nuclear power plant accident. *Journal of Hazardous Materials, 75(2-3)*, 181-194.

Naaman, M., Boase, J. & Lai, C. (2010). Is it Really About Me? Message Content in Social Awareness Streams. In *Proc. CSCW 2010*, 189-192.

National Oceanic and Atmospheric Administration. (2009A). *The Wildfire Outbreak of April 9th, 2009*. Retrieved from: http://www.srh.noaa.gov/oun/?n=events-20090409.

National Oceanic and Atmospheric Administration. (2009B). *NOAA Says Prepare for Major Spring Flooding on Red River*. Retrieved from: http://www.noaanews.noaa.gov/stories2009/20090205_redriver.html.

National Oceanic and Atmospheric Administration. (2009C). *Flood Damages Suffered in the United Statues During Water Year 2009*. Retrieved from: http://www.weather.gov/hic/flood_stats/Summaries/.

National Oceanic and Atmospheric Administration. (2010A). *Stage Set for Possible Major Spring Floods on Red River, Devils Lake*. Retrieved from: http://www.nws.noaa.gov/pa/fstories/2010/fs23feb2010.php

National Oceanic and Atmospheric Administration. (2010B). *Imminent Flood Threat in Midwest, South and East Also at Risk*. Retrieved from: http://www.noaanews.noaa.gov/stories2010/20100316_springoutlook.html

Neal, D.M. & Phillips, B.D. (1995). Effective Emergency Management: Reconsidering the Bureaucratic Approach. *Disasters, 19(4)*, 326-337.

New ice jam raising water levels, and fears, north of Winnipeg. (2009, March 26). Retrieved from http://www.cbc.ca/news/canada/manitoba/story/2009/03/26/mb-ice-jam-lockport.html

New York Times. (2011, July 22). Haiti. *The New York Times*. Retrieved from: http://topics.nytimes.com/top/news/international/countriesandterritories/haiti/index.html.

NewsOK.com (Producer). (2009, April 10). Fire prompts voluntary evacuations [video webcast]. Retrieved from: http://newsok.com/fire-prompts-voluntary-evacuations/multimedia/video/19218123001

North Dakota State University. (2011). *Flood Stages vs. Geographic Locations in the Fargo-Moorhead Region*. Retrieved from: http://www.ndsu.edu/fargo_geology/floodgeography.htm

Oklahoma Wildfire Damage Estimates Increased to $30M. (2009, April 20). Retrieved from: http://www.insurancejournal.com/news/southcentral/2009/04/20/99765.htm

Orlikowski, W.J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science, 3(3),* 398-427.

Palen, L., Vieweg, S., Sutton, J., Liu, S.B. & Hughes, A.L. (2007). Crisis Informatics: Studying Crisis in a Networked World. In *Proc. Third International Conference on E-Social Science*.

Palen, L. & Vieweg, S. (2008). The Emergence of Online Widescale Interaction in Unexpected Events: Assistance, Alliance and Retreat. In *Proc. CSCW 2008,* 117-126.

Palen, L., Vieweg, S., Liu, S.B. & Hughes, A.L. (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007 Virginia Tech Event. *Social Science Computer Review, Special Issue on e-Social Science*.

Palen, L., Anderson, K.M., Mark, G., Martin, J., Sicker, D., Palmer, M. & Grunwald, D. (2010). A Vision for Technology-Mediated Support for Public Participation & Assistance in Mass Emergencies and Disasters. In Proc. ACM-BCS Visions of Computer Science 2010.

Palen, L., Vieweg, S. & Anderson, K.M. (2011). Supporting the "Everyday Analyst" in Safety- and Time-Critical Situations. *The Information Society Journal, 27(1),* 52-62.

Palmer, M., Gildea, D. & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics, 31 (1),* 71-106.

Palmer, M., Hwang, J.D., Brown, S.W., Kipper-Schuler, K., & Lanfranchi, A. (2009). Leveraging Lexical Resources for the Detection of Event Relations. In *Proc. AAAI Spring Symposium on Learning by Reading*, 81-87.

Palmer, M., Gildea, D. & Xue, N. (2010). Semantic Role Labeling. In Hirst, G. (ed.), *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool.

Pan American Health Organization/World Health Organization. (2011). Earthquake in Haiti – One Year Later. *PAHO/WHO Report on the Health Situation*.

Panyametheekul, S. & Herring, S.C. (2003). Gender and Turn Allocation in a Thai Chat Room. *Journal of Computer Mediated Communication 9(1)*.

Phillips, B. (2002). Qualitative Methods and Disaster Research. In Stallings, R.A. (ed), *Methods of Disaster Research* (194-211). Xlibris.

Phillips, T. (2010). Haiti earthquake creating a generation of amputees, doctors warn. *The Guardian*. Retrieved from: http://www.guardian.co.uk/world/2010/jan/21/haiti-doctors-warn-amputee-crisis

Presidential Policy Directive/PPD-8. (2012a). Working Draft—National Protection Framework Review Package. Retrieved from: http://www.fema.gov/ppd8

Presidential Policy Directive/PPD-8. (2012b). Working Draft—National Response Framework Review Package. Retrieved from: http://www.fema.gov/ppd8

Qu, Y., Wu, P.F. & Wang, X. (2009). Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake. In *Proc. HICSS 2009*.

Quarantelli, E.L. & Hultaker, O. (1986). Introductory Statement. *International Journal of Mass Emergencies and Disasters 4(2)*.

Quarantelli, E.L. (2002). The Disaster Research Center (DRC) Field Studies of Organized Behavior in the Crisis Time Period of Disasters. In Stallings, R.A. (ed), *Methods of Disaster Research* (94-126). Xlibris.

Roberts, M. (2011, March 15). Haiti cholera 'far worse than expected', experts fear. *BBC News*. Retrieved from: http://www.bbc.co.uk/news/health-12744929

Romero, S. & Lacey, M. (2010, January 12). Fierce Quake Devastates Haitian Capital. *The New York Times*. Retrieved from: http://www.nytimes.com/2010/01/13/world/americas/13haiti.html.

Rubin, J. (2010, February 10). In Haiti, quake is sure to be followed by torrential rain. *The Los Angeles Times*. Retrieved from: http://articles.latimes.com/2010/feb/10/world/la-fg-haiti-shelter11-2010feb11

Saeed, J.I. (2009). *Semantics*. Malden, MA: Wiley-Blackwell.

Searle, J.R. (1975). A Taxonomy of Illocutionary Acts. In K. Gunderson (Ed.) *Language, Mind and Knowledge*. University of Minnesota Press.

Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., & Anderson, K. (2012). "Beacons of Hope" in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake. In *Proc. CSCW 2012*.

Sarter, N.B. & Woods, D.D. (1991). Situation Awareness: A Critical But Ill-Defined Phenomenon. *International Journal of Aviation Psychology 1(1),* 45-57.

Schneider, S.M. & Foot, K.A. (2004). Crisis Communication and New Media: The Web After September 11. In: P.N. Howard & S. Jones (Eds.), *Society Online: The Internet in Context* (137-153). Thousand Oaks, CA: Sage.

Schwert, D.P. (2003). A Geologist's Perspective on the Red River of the North: History, Geography, and Planning/Management Issues. In *Proc. 1st International Water Conference, Red River Basin Institute, Moorhead, MN.*

Shibutani, T. (1966). *Improvised News: a sociological study of rumor.* New York, NY: Bobbs-Merrill Company.

Shklovski, I., Palen, L. & Sutton, J. (2008). Finding Community Through Information and Communication Technology in Disaster Events. In *Proc. CSCW 2008,* 127-136.

Simon, H.A. (1996). *The Sciences of the Artificial.* Cambridge, MA: MIT Press.

Sonnenwald, D.H. & Pierce, L.G. (2000). Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing and Management, (36),* 461-479.

Sood, R., Stockdale, G. & Rogers, E.M. (1987). How the News Media Operate in Natural Disasters. *Journal of Communication*, 37(3), 27-41.

Spellman, J. (2010). *Heading off disaster, one tweet at a time.* September 22, 2010. http://articles.cnn.com/2010-09-22/tech/natural.disasters.social.media_1_social-media-twitter-first-responders?_s=PM:TECH.

Sproull, L. & Kiesler, S. (1993). *Connections.* Cambridge, MA: MIT Press.

Starbird, K., Palen, L., Hughes, A.L. & Vieweg, S. (2010). Chatter on The Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proc. CSCW 2010,* 241-250.

Starbird, K. & Palen, L. (2011). "Voluntweeters:" Self-Organizing by Digital Volunteers in Times of Crisis. In *Proc. CHI 2011,* 1071-1080.

Steinlechner, B. (2011, January 17). Logistics specialists help UNICEF deliver nutrition supplements to Haitian children. *United Nations Children's Fund - UNICEF.org.* Retrieved from: http://www.unicef.org/infobycountry/haiti_57472.html

Tierney, K.T., Lindell, M.K. & Perry, R.W. (2001). *Facing the Unexpected: Disaster Preparedness and Response in the United States*. Washington, D.C.: Joseph Henry Press.

Tierney, K.J. (2007). Recent Developments in U.S. Homeland Security Policies and Their Implications for the Management of Extreme Events. In: H. Rodriguez, E.L Quarantelli & R.R. Dynes (Eds.) *Handbook of Disaster Research* (405-412). SpringerLink.

Trask, R.L. (1999). *Key Concepts in Language and Linguistics*. London, UK: Routledge.

Turner, R.H. (1994). Rumor as Intensified Information Seeking: Earthquake Rumors in China and the United States. In: R.R. Dynes & K.J. Tierney (Eds.), *Disasters, Collective Behavior and Social Organization* (244-256). Newark, NJ: University of Delaware Press.

United States Agency for International Development (USAID). (2010). *USAID Post-Earthquake Response*. Retrieved from: http://www.usaid.gov/ht/helphaiti.html

Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Schram, A., & Anderson, K.M. (2011). Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. In *Proc. ICWSM 2011*, 385-392.

Vieweg, S., Palen, L., Liu, S., Hughes, A. & Sutton, J. (2008). Collective Intelligence in Disaster: Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shooting. In *Proc. ISCRAM 2008*, 44-54.

Vieweg, S., Hughes, A.L., Starbird, K. & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proc. CHI 2010*, 1079-1088.

SARAH VIEWEG

CURRICULUM VITAE

2300 Arapahoe Ave. #241
Boulder, CO 80302

sarah.vieweg@colorado.edu
www.sarahvieweg.com
Mobile: 720-352-7334

INTERESTS

My research interests are in Human-Computer Interaction (HCI), Computer-Supported Cooperative Work (CSCW), Social Media, Microblogging, and Natural Language Processing (NLP). I use ethnographic, discourse analytic and user-centered design methods to understand behavior in both online and offline settings, with a focus on developing intuitive technologies that help users locate and organize useful, personally meaningful information.

EDUCATION

Ph.D. in Technology, Media and Society

University of Colorado, Boulder, CO May 2012

Advisor: Leysia Palen, Ph.D.

M.A. in Linguistics

University of Colorado, Boulder, CO August 2007

Graduate Certificate in Human Language Technology

University of Colorado, Boulder, CO May 2007

B.A. in Economics and French

University of Illinois, Urbana-Champaign, IL May 1999

RESEARCH EXPERIENCE

Research Assistant
EPIC Lab (Empowering the Public with Information in Crisis), Fall 2007 – Spring 2012
I conduct qualitative and quantitative research to understand human behavior through technology use during times of mass emergency. Specifically, I analyze how people use Twitter to communicate information that leads to situational awareness in safety-critical situations. Results of my research inform the development of computational tools that enable people to better find, organize and understand timely, relevant information.

Project Supervisor and Verb Sense Annotator
Verb Sense Annotation Project, Fall 2005 – Summer 2007
I worked on a team to create lexical resources for an NSF- and DARPA-funded NLP research project on training machines to process and interpret human language. Additionally, I trained, mentored and managed research assistants.

PUBLICATIONS

REFEREED JOURNAL ARTICLES

J3. Chi, Ed H., Sean Munson, Gerhard Fischer, Sarah Vieweg & Cynthia Parr. 2010. **Advancing the Design of Technology-Mediated Social Participation Systems**, IEEE Computer 43(11): 29-35.

J2. Palen, Leysia, Sarah Vieweg & Kenneth M. Anderson. 2010. **Supporting "Everyday Analysts" in Safety- and Time-Critical Situations**. The Information Society, 27(1): 52-62.

J1. Palen, Leysia, Sarah Vieweg, Sophia B. Liu & Amanda L. Hughes. 2009. **Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007 Virginia Tech Event**. Social Science Computer Review, Special Issue on e-Social Science 27(5): 1-14.

LONG CONFERENCE PUBLICATIONS, REVIEWED IN FULL

C9. Verma, Sudha, Sarah Vieweg, William Corvey, Leysia Palen, James H. Martin, Martha Palmer, Aaron Schram and Kenneth M. Anderson. 2011. **NLP to the Rescue?: Extracting "Situational Awareness" Tweets During Mass Emergency**. Proceedings of the AAAI Conference on Weblogs and Social Media, ICWSM 2011.

C8. Vieweg, Sarah, Amanda L. Hughes, Kate Starbird & Leysia Palen. 2010. **Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness.** Proceedings of the ACM conference on Computer Human Interaction, CHI 2010.

C7. Starbird, Kate, Leysia Palen, Amanda L. Hughes & Sarah Vieweg. 2010. **Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information**. Proceedings of the ACM conference on Computer Supported Cooperative Work, CSCW 2010.

C6. Palen, Leysia & Sarah Vieweg. 2008. **The Emergence of Online Widescale Interaction in Unexpected Events: Assistance, Alliance and Retreat**. Proceedings of the ACM conference on Computer Supported Cooperative Work, CSCW 2008.

C5. Vieweg, Sarah, Leysia Palen, Sophia B. Liu, Amanda L. Hughes & Jeannette Sutton. 2008. **Collective Intelligence in Disaster: An Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shootings**. Proceedings of the Information Systems for Crisis Response and Management Conference, ISCRAM 2008. **Tied for best student paper award**

C4. Hughes, Amanda L., Leysia Palen, Jeannette Sutton, Sophia B. Liu & Sarah Vieweg. 2008. **Sight-Seeing in Disaster: An Examination of On-Line Social Convergence**. Proceedings of the Information Systems for Crisis Response and Management Conference, ISCRAM 2008. **Tied for best student paper award**

C3. Liu, Sophia B., Leysia Palen, Jeannette Sutton, Amanda L. Hughes & Sarah Vieweg. 2008. **In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster**. Proceedings of the Information Systems for Crisis Response and Management Conference, ISCRAM 2008. **Tied for best student paper award**

C2. Palen, Leysia, Sarah Vieweg, Jeannette Sutton, Sophia B. Liu & Amanda L. Hughes. 2007. **Crisis Informatics: Studying Crisis in a Networked World**. Proceedings of the Third International Conference on E-Social Science, eSS 2007.

C1. Duffield, Cecily Jill, Jena D. Hwang, Susan Windisch Brown, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis & Martha Palmer. 2007. **Criteria for Manual Clustering of Verb Senses.** Proceedings of the 29th Annual Conference of the Cognitive Science Society, CogSci 2007.

BOOK CHAPTERS AND INVITED PUBLICATIONS

B3. Starbird, Kate, Leysia Palen, Sophia B. Liu, Sarah Vieweg, Amanda Hughes, Aaron Schram, Kenneth M. Anderson, Mossaab Bagdouri, Joanne White, Casey McTaggart, and Chris Schenk. 2011. **Promoting Structured Data in Citizen Communications during Disaster Response: An Account of Strategies for Diffusion of the "Tweak the Tweet" Syntax.** To appear in Christine Hagar (Ed.) Crisis Information Management: Communication and Technologies, Chandos Publishing.

B2. Palen, Leysia, Kate Starbird, Sarah Vieweg, Amanda Hughes. 2010. **Twitter-based information distribution during the 2009 Red River Valley flood threat.** Bulletin of the American Society for Information Science and Technology, 36(5): 13–17.

B1. Liu, Sophia B., Leysia Palen, Jeannette Sutton, Amanda L. Hughes & Sarah Vieweg. 2009. **Citizen Photojournalism During Crisis Events**. In Stuart Allen and Einar Thorsen (Eds.), Citizen Journalism: Global

Perspectives. New York: Peter Lang. *reprint of a previous conference paper

WORKSHOPS AND POSTERS

W7. Vieweg, Sarah. Forthcoming. **Twitter Communications in Mass Emergency: Contributions to Situational Awareness.** CSCW 2012.

W6. Vieweg, Sarah. 2011. **Feminist Conversation Analysis: Opportunities for HCI**. CHI 2011 Workshop on Feminism and Interaction Design.

W5. Corvey, William J., Sarah Vieweg, Travis Rood & Martha Palmer. 2010. **Twitter in Mass Emergency: What NLP Techniques Can Contribute.** NAACL HLT Workshop on Computational Linguistics in a World of Social Media, 23-24.

W4. Vieweg, Sarah & Kate Starbird. **Microblogging in Mass Emergency**. CHI 2010 Workshop on Microblogging: What and How Can We Learn From It?

W3. Vieweg, Sarah. 2010. **The Ethics of Twitter Research**. CSCW 2010 Workshop on Revisiting Ethics in the Facebook Era: Challenges in Emerging CSCW Research.

W2. Vieweg, Sarah. 2008. **Social Networking Sites: Reinterpretation in Crisis Situations**. CSCW 2008 Workshop on Social Networking in Organizations.

W1. Duffield, Cecily Jill, Jena D. Hwang, Susan Windisch Brown, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis & Martha Palmer. 2007. **Criteria for the Manual Grouping of Verb Senses.** Proceedings of the Linguistic Annotation Workshop, Prague, Czech Republic.

ORGANIZED WORKSHOP

w1. Rotman, D., Vieweg, S., Yardi, S., Chi, E., Preece, J., Shneiderman, B., Pirolli, P. & Glaisyer, T. **From Slacktivism to Activism: Participatory Culture in the Age of Social Media.** Proceedings of the 29th International Conference on Human Factors in Computing Systems, CHI 2011.

DOCTORAL COLLOQUIUM

DC1. Vieweg, S. 2010. **Microblogged Contributions to the Emergency Arena: Discovery, Interpretation and Implications**. ACM conference on Computer Supported Cooperative Work, CSCW 2010.

INVITED TALKS AND PANELS

P11. "Twitter Communications in Mass Emergency: Contributions to Situational Awareness" Invited talk at Virginia Tech University, November 21, 2011.

P10. "The Computational World: Thinking about Twitter, Crisis and Communication" Presented to the Computational World class at the University of Colorado, March 11, 2011.

P9. "ICT and Critical Environments" Presented to the Digital and Social Systems class at the University of Colorado, September 18, 2009.

P8. "Speech acts and the Winograd/Flores and Suchman debate" Presented to the HCI Remixed class at the University of Colorado, December 9, 2008.

P7. "Collective Intelligence in Disaster: Community Response in an Online Arena" Presented to Community Informatics Class at Dominican University, November 19, 2008.

P6. Evaluator on Student Project Panel, User Interaction Design class at the University of Colorado, March 31, 2008.

P5. "Ethics, Theories and Thoughts on HCI" Presented to the User Interaction Design class at the University of Colorado, March 12, 2008.

P4. Evaluator on Student Project Panel, User Interaction Design class at the University of Colorado, February 6, 2008.

P3. "HCI Researchers at CU" Presented to the User Interface Design class at the University of Colorado, November 28, 2007.

P2. Evaluator on Student Project Panel, User Interface Design class at the University of Colorado, November 7, 2007.

P1. "Verb Sense Disambiguation" Presented with Cecily Jill Duffield and Jena Hwang to the Département de Linguistique et de Traduction at the Université de Montréal, November 15, 2006.

TEACHING POSITIONS

Teaching Assistant
User Interaction Design (CSCI 4838/6838) Spring 2009

Course Grader
Readings in HCI (CSCI 7000) Fall 2008

Course Grader
User Interaction Design (CSCI 4838/6838) Fall 2007, Spring 2008

PROFESSIONAL SERVICE

Reviewer
ACM Conference on Human Factors in Computing (CHI)
ACM Conference on Computer-Supported Cooperative Work (CSCW)
ACM TOCHI Journal
New Media and Society Journal
International Conference on Weblogs and Social Media (ICWSM)
International Conference on Intercultural Collaboration
Information Systems for Crisis Response & Management Conference (ISCRAM)
Culture, Language and Social Practice Conference (CLASP)

Conference Organizer
Culture, Language and Social Practice (CLASP) 2007

Committee Member
University of Colorado New Venture Challenge Business Plan Competition, 2009-2011
University of Colorado New Venture Challenge Executive Committee, 2009-2010
University of Colorado Emergency Preparedness Committee, 2007

HONORS/AWARDS

Human Computer Interaction Consortium (HCIC), Pacific Grove, CA June 14-18, 2011 (Full Funding)
Invited Student Participant

Human Computer Interaction Consortium (HCIC), Fraser, CO Feb. 4-7, 2009 (Full Funding)
Invited Student Participant

Technology Mediated Social Participation (TMSP) workshop Palo Alto, CA, Dec. 10-11, 2009 (Full Funding)
Initiative to organize agenda around research and education to promote technology-mediated interaction and participation in line with national and international priorities

Mike Meleshkin Award for "Best PhD Paper"
Information Systems for Crisis Response and Management (ISCRAM) Conference 2008