Symbol Grounding in Social Media Communications

Andrew J. Hampton
Wright State University

SYMBOL GROUNDING IN SOCIAL MEDIA COMMUNICATIONS

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

ANDREW J. HAMPTON
B.S., University of Central Florida, 2009
M.S., Wright State University, 2013

2018
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

April 10, 2018

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Andrew J. Hampton ENTITLED Symbol Grounding in Social Media Communications BE ACCEPTED IN PARTIAL FULFILLMENT OF REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Valerie L. Shalin, Ph.D.
Dissertation Director

Scott N. J. Watamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Graduate Program Director

Barry Milligan, Ph.D.
Interim Dean of the Graduate School

Committee on
Final Examination

Valerie L. Shalin, Ph.D.

John M. Flach, Ph.D.

Ion Juvina, Ph.D.

Amit P. Sheth, Ph.D.

ABSTRACT

Hampton, Andrew J. Ph.D., Department of Psychology, Human Factors & Industrial / Organizational Psychology Ph.D. program, Wright State University, 2018. Symbol Grounding in Social Media Communications.

Social media data promise to inform the disaster response community, but effective mining remains elusive. To assist in the analysis of community reports on disaster from social media, I draw on an integrated model of psycholinguistic theory to investigate the patterns by which language use changes as a function of environmental influence. Using social media corpora from several disasters and non-disasters, I examine variations in patterns of lexical choice between domain independent paired antonyms with respect to an Internet-specific base rate to determine generic sentinels of breach of canonicity. I examine social media content with respect to disaster proximity and examine relative proportions of actionable content in messages containing words that indicate breach. Results indicate a preliminary set of antonym pairs that vary consistently with respect to breach. Despite the absence of correlation with actionable content density, two related findings support the role of a psycholinguistic perspective on the mining of social media data. First, several diagnostic pairs reflect human function in an environment independent of sentiment. Second, the analysis of sentiment by spatial proximity suggests an increase in positive sentiment with proximity. Both findings motivate the continued study of how human behavior contributes to the production of social media messages, and hence the analysis of the messages they produce. I note several methodological contributions resulting from this work, including the expanded set of informative domain independent lexical items, consideration of base rates that both enables detection of departure from canonicity and reduces reliance on anonymous reporting, and a complement to sentiment analysis that is sensitive to environmental variability. Theoretical contributions include consolidation of disparate threads of language production research (including a focus on grounding). Finally, I identify several limitations in my own analysis, and more general concerns regarding the mining of social media data, to guide future work.

Keywords: psycholinguistics, lexical choice, breach, disaster response

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

I. INTRODUCTION

Language necessarily reflects aspects of the circumstance from which it arises. Some researchers emphasize a separability of environment from cognitive linguistic processes to facilitate the study of isolated linguistic processes (e.g., Saussure, 1979/2011), whereas others embrace the inextricability of environment and language (Peirce, 1894; Peirce, 1868/2015), and still others frame as a narrative tool, sensitive to breaches of expectations (Bruner, 2003). In this dissertation, I argue for the necessity of considering all three perspectives to identify generic aspects of language that coincide with patterns of environmental change in empirically identifiable ways. This could potentially support automated reasoning backward from linguistic choice to noteworthy departures in the world. Social media offer a rich source of data for this effort, with both environmentally situated and computationally accessible examples of language production. Disaster provides a strong natural manipulation of context. Though a substantial literature exists on the interpretation of social media during disaster, this literature rarely considers human language production processes as a factor in the interpretation.

This approach also introduces potential improvements to disaster response. The evaluation of need during disaster response and recovery presents a daunting challenge. Disaster disrupts the normal functioning of a society (Perry, 2007) creating what Quarantelli (2008) called the *problems of living* (p. 893) that need *solving*. Metrics such as emergency call volume and hospital admissions are coarse. Satellite imagery from a flood, for example, may lag changes in need. Moreover, disaster conditions do not neatly correspond to unidimensional physical sensors. Flooding requires more information than rainfall or storm surge metrics. Prior weather events, topography, demographics, and socio-cultural factors such as construction practices and infrastructure all impact whether a given event constitutes a human disaster.

Humans readily identify situations as critical or not, leveraging an intuitive sense of circumstances breaching expectations and bypassing the need to reason separately about these kinds of contributory factors. While human observers already play a part in the disaster response system (e.g., by calls to 9-1-1 or visits to emergency rooms) bandwidth concerns and computational inaccessibility of such metrics render these only marginally useful for directing aid in wide-scale disaster events. Sheth (2009) points to the potential of "citizen journalism" and multi-modal information broadcast from handheld computers, creating what he calls citizen sensor networks.

For example, when earthquakes struck Virginia in the summer of 2011, people noted the event as exceptional and immediately posted this fact on their Twitter accounts, or else sent text messages to the same effect (Hotz, 2011). Cellular data travel much faster than seismic waves, resulting in social networks receiving information about the earthquake as much as forty seconds before they felt the actual waves. In comparison, the existing U.S. Geological Survey's conventional warning system had a best alert time of around two minutes, and it required people to sign up to receive alerts. Social media data from those situated in the environment promise to reflect the experienced disruption and resulting functional stress[1]. A central thesis of this work is that an understanding of human language processes and motivations will inform the interpretation of social media as they relate to the disaster environment.

Twitter, the world's largest microblogging service, currently has 328 million active users (Twitter, 2017a) producing updates in 140 characters or fewer[2]. Eighty-three percent of these users log in on their mobile devices. These data are accessible from anywhere and remain (potentially) available for years. Further, Twitter's functionality can sometimes remain even when Internet and voice lines fail (Kawamura, & Ohsuga, 2013; Li & Rao, 2010). People carry their reporting devices with them at all times, and many of them can be relied upon to report salient features of their surroundings. This means that a sizable portion of the population at any given time and place constitutes a citizen sensor network producing data that reflect (and in some way measure) their circumstances.

However, difficulty remains in finding, interpreting, and scaling relevant, actionable signal in a virtual firehose of noise. Many researchers (e.g., Palen & Liu, 2007; Sheth, 2009; Starbird, 2011) analyze social media message content, sentiment, organization, and dispersal (among other topics) making progress regarding message filtering and analysis. Despite this obvious potential, we have limited guidelines for interpreting or leveraging the power of social media. Significant work remains in explicitly identifying the function that relates citizen sensor language content to the environment experienced. Madey, Szabo, and Barabási (2006) developed a prototype

[1]Note that "stress" does not necessarily denote a negative influence (Selye, 1978).
[2] The limit was recently doubled to 280 characters. This does not change the principles of my approach.

algorithm to determine the rough boundaries of disaster based on aggregate movement of phones and call metadata, but this frequency-over-time approach attempts no more than a binary outline of the affected area. Without content based interpretation, this algorithm also fails to consider the capacity of humans to act as intelligent and insightful observers within these boundaries. Sentiment analysis (e.g., Pang & Lee, 2008; Wang, Chen, Thirunarayan, & Sheth, 2012), on the other hand, leverages human insight and response, but in disaster contexts this may prove misleading because public response to disaster is not uniformly negative (Rodríguez, Trainor, & Quarantelli, 2006). Suedfeld (1997), in his analysis of the impacts of reactions to societal trauma and uncertainty, notes the importance of considering both *distress*, associated with breakdown, and *eustress* that may result after having risen to a challenge.

This dissertation considers the processing assumptions that allow people to serve as real-time sensors of their environment to guide the selection of informative metrics. I rely on psycholinguistic theory from both the process model and pragmatic viewpoints. I propose a model to integrate these related but traditionally separate viewpoints by combining language process models with context sensitive speech act theory (Searle, 1983) and Bruner's (2003) notion of narrative breach. The integrated model supports the study of novel measures of social media data to determine their relationship to uncertainty and stress derived from environmental factors. The resulting set of *sentinels of breach* potentially contributes information in disaster response scenarios, supplementing what is available through conventional approaches like sentiment analysis. My contribution is primarily methodological, with relatively modest theoretical contributions concerning the role of social media as narrative.

Below I examine the relationship between language use and the environment in several ways. Fundamentally, behavior, including language, should vary with respect the nature of the event in which it is situated. In exceptional circumstances, variations should be empirically observable when compared to normative conditions. Further, because geographic proximity to disaster provides a natural manipulation of context, it permits an examination of language choice relative to varied experience in the environment.

Background

As shown in Figure 1, social media messages may bear some relationship to environmental conditions. The fundamental issue here is known as "symbol grounding"—the process of generating and comprehending the mapping between a linguistic symbol, such as "severe" or "all", and the environment. Presumably social media messages respect principles of informativeness common in any communication (Grice, 1975). A particularly dangerous storm should promote words pertaining to above average size, and objectively bad experiences should promote the word "bad" over "good".

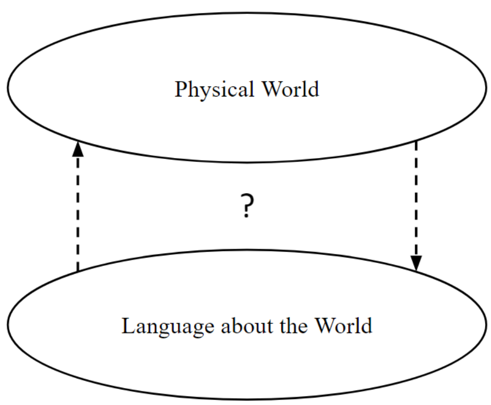

Figure 1. A naive model of the relationship between language and the world.

Computer and information scientists, such as Sheth, Palen, and Liu lead the effort to provide the technology that realizes the potential of social media to indicate areas of need and public response. While such work recognizes some properties of the human who provides this information (such as sentiment) insufficient research addresses the central thesis of this dissertation: the processes that generate the message play a role in the interpretation of that message. Moreover, this suggests alternative metrics in the analysis of social media. This thesis motivates the following modification to the initial figure, as shown in Figure 2.

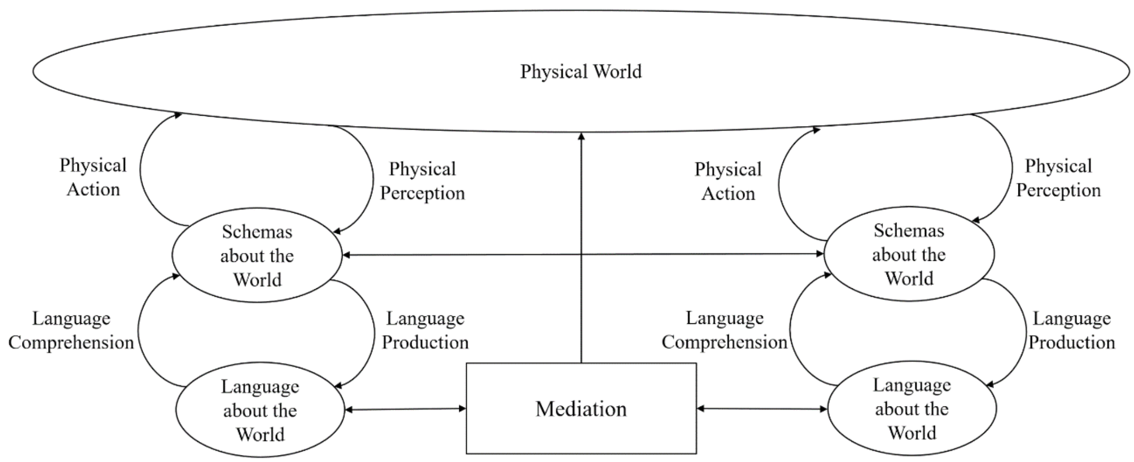

Figure 2. Grounded integration of perception, action, cognition, linguistics, and mediation.

Figure 2 organizes the following introduction of related literature. First, I describe the general philosophical perspective Figure 2 suggests. Then, I align the existing empirical literature with processes in the figure, considering perception and action as distinct from language comprehension and production and the psychological perspective, with an additional consideration for the influence of mediated communication. I then examine conventional computer science approaches to analyzing social media data. I conclude this introduction with implications of the model for expanding our notion of the language metrics that inform the condition of the world.

Philosophical Considerations

In this section, I examine language processes from a philosophical perspective, including its separable study apart from the environment, the distinction between of language and thought, canonicity and breach, as well as the implications of having multiple intelligent agents contributing to discourse. Such differences in the conceptualization of language processes and purposes have had profound, but somewhat hidden influences on the empirical work discussed in a later section.

Role of the environment. Figure 2 includes the physical world, mental representation (schemas), and physical representation. This depiction is a rotated copy of Roy's model of symbol grounding (see Figure 3).

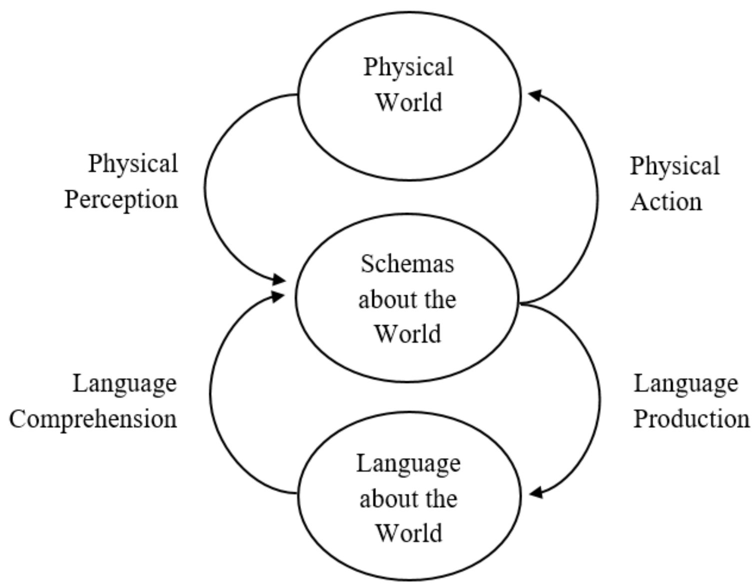

Figure 3. Roy's model of symbol grounding, distinguishing between the physical world, schemas internal to the agent about the world, and language about the world (rotated from the original horizontal depiction)

Acknowledging the impossibility of separating instruction from physical environment for complex machines, Roy (2005) created a model depicting a triadic interaction among language about the world and the world itself, with schemas serving as the integral mediation between them. Drawing from the semiotic triangle of Ogden and

Richards (1923), he specifies “physical world”, “schemas about the world”, and “language about the world”, creating a model whereby perception and action schemas mediate the relationship between words and the world. These schemas provide grounding, coherence, and a conceptualization of the means by which knowledge affects humans’ interaction with their environment.

Triadic models[3] such as Roy’s, inspired by the semiotic theories of Peirce (e.g. Peirce, 1894; Peirce, 1868/2015), appear essential to systems that require symbolic grounding in the environment (Noth, 2009). The requirement is specific to multi-agent collaborative activity. Fully autonomous robots may not require schemas to operate. However, when an outside agent provides goal-oriented commands to a machine (e.g., “pick up *all* the marbles”) the robot and the commanding agent require some shared understanding of “all” in the present context, likely all the marbles in present view, but not marbles hidden in a drawer.

Including the physical world in a language-based model of communication specifically contrasts with the more conventional view “bottom(ing) out in symbolic, human generated descriptions of the world” (Roy, 2005, p. 174). The conventional model, considered dyadic (Flach & Voorhorst, 2017), appears most prominently in a series of lectures by Ferdinand de Saussure compiled a century later (1979/2011). The dyadic perspective focuses on the relationship between meaning and symbols (typically

[3] An influential triadic model presented in Flach & Voorhorst (2017) offers an alternate view. The primary difference lies in their conception of language as a medium between agent and environment rather than as a product of or catalyst to that interaction. In models such as Roy’s, language is separable from cognition requiring language specific production and comprehension processes (Fodor, 1983). Moreover, not all cognitive behavior results from the processing of explicit symbolic representations.

language). Saussure's notion of "arbitrariness of the sign" refers to its (lack of) relationship to the environment and specifically excludes phenomena such as onomatopoeia. Rather, the sign takes on meaning with respect to the agent's knowledge. Outside of robotics, such a view has been enormously influential in computer science (albeit with considerable support from cognitive psychologists such as Landauer & Dumais, 1997) where word meaning is characterized by the lexical company it keeps. Philosophers of mind refer to this view as *intensional semantic*s (e.g., Heim & Kratzer, 1998).

Of course, this will never suffice when the primary focus becomes the relationship between symbols and the world. *Extensional semantics* concerns itself with that relationship. *Deictic* terminology—including pronouns and spatio-temporal referents such as "here", "today", or "now"—has no meaning apart from the context in which it appears. This important phenomenon suggests that *no symbol can have a completely fixed meaning apart from the environment* to which it refers. The context-sensitive relationship between a symbol and the environment allows a restaurant server to refer comprehensibly to a customer as a "ham sandwich" (Nunberg, 1979). The *ad hoc* reference works because in that environment that customer ordered a ham sandwich, and no other current customer did the same.

These examples are consistent with Peirce (1894), who conceives of meaning as arising from the interaction of three interconnected parts: the object (some entity), the sign (something which represents the object—often symbolically or linguistically), and the interpretant (an agent's conception of the object via the sign). Symbolic cognition, then, becomes dependent on the environment (the object, broadly construed) in which it occurs. While contemporary cognitive science admits both extensional and intensional notions of semantics, it typically argues for symbolic "bridges" that allow for the study of symbol manipulation apart from the environment to which it refers (Newell, 1980). These fundamentally dyadic models have had substantial influence on psycholinguistic research.

Language and thought. Roy's model employs schemas as an interface between language and the world it represents. In separating schemas and language, Roy takes a position on a longstanding debate dating to at least the Sapir-Whorf investigations (e.g., Whorf, 1956). The initial question was whether the categories marked with linguistic distinctions drive perception. The consensus answer is "no". Languages that lack refined color labels do not disable the detection of these differences in controlled stimuli, maintaining a distinction between language and experience. Most contemporary psycholinguists make the same distinction between language and thought (e.g., Gleitman & Papafragou, 2013; Pinker, 1995). A plethora of linguistic phenomena (e.g., irony, polysemy, metaphor) emphasizes the separability of linguistic representation and meaning. Numerous propositions have the same truth conditions (e.g., "This beef is 75% lean" vs. "This beef is 25% fat") but the distinctive styles have different functional meanings as evidenced by subsequent reasoning (Tversky & Kahneman, 1981).

Vygotsky (1969) provides useful insight into the relationship between language and thought. He proposes different evolutionary and developmental foundations for these, linking thought to motor behavior, and language to culture. According to Vygotsky, language is a means to absorb the culture, and by inference, the means for distributing cultural knowledge. Language, including the lexicon, preserves distinctions that are culturally relevant and reflected in schemas. Language is thus a culturally determined, symbolic artifact of human thought.

Although certainly biologically enabled, an individual utterance constructed by humans is the result of language-specific *intentional processes* conceptually separable from the intentionality embedded in schemas. That is, the way an observer understands the world stands distinct from the way in which she represents it with language, based on her intention in relaying the information to a particular conversational partner. For example, a speaker may choose to suppress established detail in order to convey relevance with minimal recipient effort (Wilson & Sperber, 2004). All representational artifacts (e.g., maps and train schedules), which form a central focus of human factors psychology, reflect intentionality in production and intentionality in use. Such artifactual representations are also conceptually separable from those signs whose appearance is exclusively grounded in physical law. Leaves do not choose whether, how, or when to rustle in the wind. They cannot time their announcements to avoid concurrent noise, and thereby attract attention. The appearance of density gradients with distance and movement do not choose to make themselves apparent or add contrast under poor viewing conditions. Human language, however, reflects intentional choice, shaping what messages are important enough to merit articulation (Gricean maxim of quantity), with lexical and syntactic choices that further the articulation of intent (Gricean maxim of manner).

Breach. Bruner's (2003) work on narrative concerning *breach* of normative conditions converges with the Gricean maxim of quantity: "...to be worth telling, a tale must be about how an implicit canonical script has been breached, violated, or deviated from in a manner to do violence to what Hayden White calls the ‘legitimacy’ of the canonical script” (Bruner, 2003, p. 11). Bruner further points out that knowledge, and by extension language, are never “point-of-viewless”. Breach is a socio-culturally determined phenomenon. Consider my lab mate’s recent experience (P. Garvarik, personal communication, 2017): a single 20-watt light source in an American hotel room is a breach; in a Rwandan hotel it is not. Our American visitor in Rwanda determines breach presumably due to unexpected constraints on activity imposed by the relative darkness. Indeed, the breach was judged sufficiently interesting to volunteer as a story told at a Christmas party. Narratives are culturally transmitted constructions, constrained by individual characteristics and relationships; they are built by a community and interpreted by an individual. The simultaneous construction and representation of human experience, then, derives from the individual’s exposure to and mastery of the culture and its narrative conventions, including what is worth saying and what linguistic markers make distinctions.

While the cultural constructions perspective provides a framework for interpreting narrative structure in a community, examination of the individual and her language behavior centers on the psychological constructs of uncertainty, degree of control over an outcome, and stress that accompany breach. While the darkness of the Rwandan hotel constituted breach, it was the ensuing stress that would impact my friend’s complaint. This breach–stress relationship allows a comprehensive consideration of situated language behavior, promising a generalizable construct to guide speech production, spanning different disasters, and potentially positive or negative experiences.

Multi-agent model. Figure 2 includes two agents and provides a purpose for language. The Pickering and Garrod model for language (2004; see Figure 4), upon which Figure 2 is substantially based, emphasizes the presence of multiple agents. Influenced by Levelt (1993), Pickering and Garrod's model suggests a fixed, overlapping procession through stages of production and comprehension (while allowing reciprocal influence between levels, unlike Levelt). Language, in their conception, aims to arrive at a shared understanding, with exchanges iteratively moving toward that end. According to their model, language usage strives toward *alignment* between interlocutors across several layers of analysis.

Support for the focus on alignment in production goes back at least to Garrod and Anderson (1987), who employed a maze game task to identify linguistic patterns of dyads, noting striking similarities in adjacent speaking turns. They described this phenomenon as input/output coordination, focusing on the pragmatic and semantic parity between members of a dyad (and assuming alignment at the lexical level), but allowing for the possibility of alignment on various cognitive levels that impact linguistic production, meshing with the principles proposed by both Dell (1986) and Levelt (1993)[4]. This opens the door for the specification of alignment at the level of situation models as the driver of situated linguistic interaction.

Figure 2 therefore portrays two conversational partners, *A* and *B* (left and right, respectively). The indistinct boundary between *A* and *B*'s environments reflects the possibility that they may share critical features to varying and uncertain degrees (for

---

[4] Dell (1986) detailed a spreading activation mechanism that allowed reciprocal influence among the various language production levels. Levelt (1993) proposed a fixed, overlapping, and unidirectional progression through these stages leading to articulation. Both recognized the importance of interacting levels of cognitive functioning in language production.

example, the same rainfall, the same transportation infrastructure, or the same utilities, but not necessarily the same for all of these). These discrepancies in experience motivate the use of language to achieve some higher-order task-specific goal, for example to articulate need not readily apparent to a partner. H. Clark is a prominent example of this multi-agent language research tradition (e.g., 1979; 1980; 1992; Clark & Chase, 1972; Clark & Carlson, 1982; Clark & Schaefer, 1989).

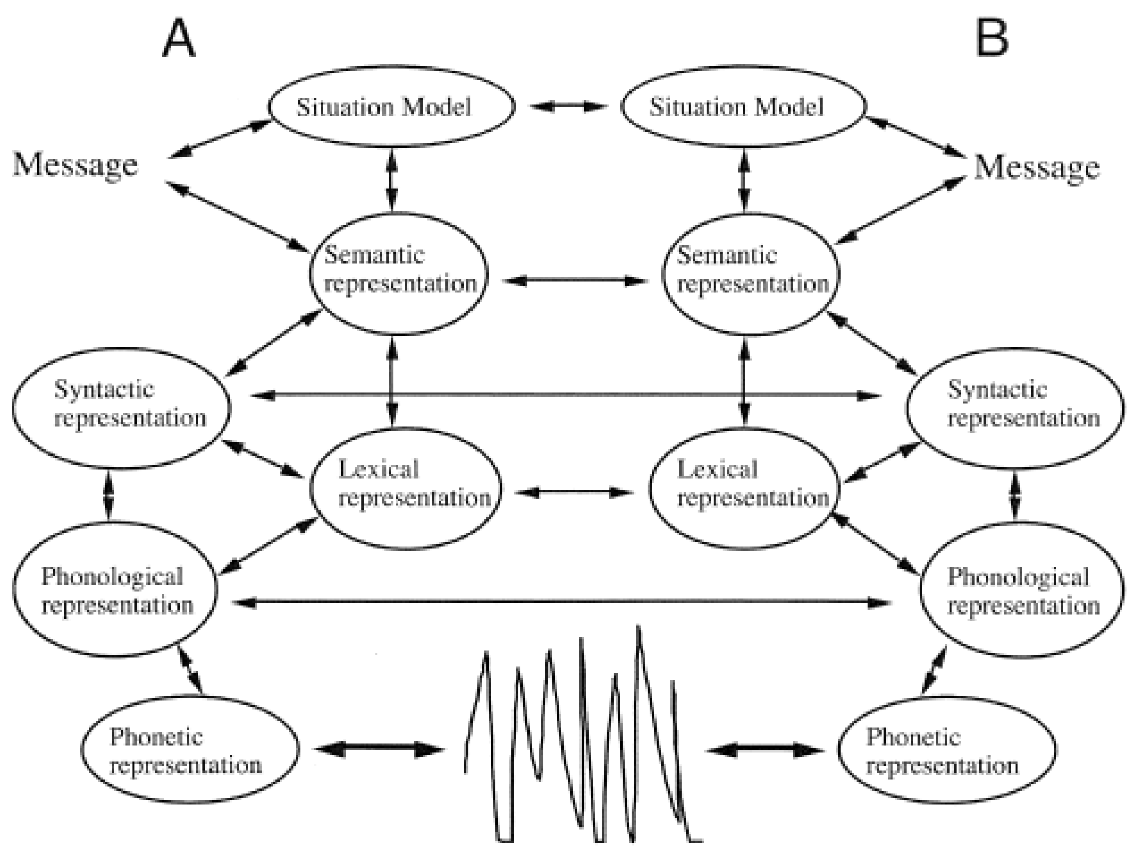

Figure 4. A psycholinguistic process model of face-to-face dialogue from Pickering & Garrod (2004).

There are several notable limitations of the original Pickering and Garrod model (Figure 4). First, it omits explicit concern for the environment, which is presumably subsumed by the situation model. This omission would seem to overlook a primary purpose of task-oriented communication: to identify real-world objects and events and develop a coordinated response.[5]

Second, and related, Pickering and Garrod have no explicit consideration for the intention of the speaker. Intentionality, classically encompassed by speech act theory (Searle, 1983), has no acknowledgement beyond “Message”—left undefined in their model. As conversational partners and their language constitute aspects of the environment, alignment in itself addresses the symbol grounding problem (i.e., the referents for lexical items) by creating a shared understanding of terms, but in Figure 4 participants strive toward grounding only for its own sake and not to accomplish some situated goal. While alignment is the observable process, I suggest that the implicit communication goal between agents is an alignment in situation models *grounded in the environment.* Interlocutors consistently strive toward this alignment, leveraging other levels of linguistic production to do so.

Pickering & Garrod’s model, like that of Dell (1986) does allow reciprocal influence among the stages. Partners iteratively align at each of the stages, including but certainly not limited to phonetic representation. Niederhoffer and Pennebaker (2002) studied both face to face and chat room communications and found that people speaking to one another tend to match many linguistic patterns consistent with alignment. For example, the number of words per turn and the types of words used converge on both turn-based (proximal), and conversational (overall) levels of analysis. These principles extend to broader contexts as well. Scholand, Tausczik, and Pennebaker (2010) found

---

[5] Recall the motivation of Roy’s model of language in robotics. The grounding problem arises from the collaboration between agents.

that level of formality (syntax), word syllable length (lexical choice), and relative use of categories (semantics) all tend to correlate within social networks, as measured by the Linguistic Inquiry Word Count system of language analysis (Pennebaker, Francis, & Booth, 2001).

Alignment manifests in the use of common ground (e.g., Clark & Brennan, 1991; Clark & Marshall, 1981), whereby conversational partners establish *ad hoc* referents for task terminology. The phenomenon of deixis, previously noted, illustrates the function of common ground, to enable the use of abbreviated description. Deixis occurs when conversational participants use context-dependent terminology, "indexed" to the current situation. Consider the proposition that an ambulance is in a particular location. A speaker could say that "the ambulance is at the corner of Park St. and State St.", or alternatively that "the ambulance is here." The former appears when the location cannot be otherwise recovered, while the latter appears when speaker and recipient share context to disambiguate the referent "here". Critically, the speaker relies on shared context in shaping her message. "Come here"[6] requires the speaker to believe that the recipient knows where "here" is.

Other measures of language behavior are consistent with the pursuit of alignment. D. Gibson (2010) shows that change in topic correlates with conversational latency, presumably to help listeners adjust to the new alignment. Cues for change, such as recruiting new speakers in a discussion may indicate a shift in focus. Dominance of one speaker does not eliminate the progress toward mutual alignment. Certainly, identifying

[6] Note that "come here" takes less effort to articulate and, given a pre-established referent for "here", less effort to comprehend, consistent with Wilson & Sperber's (2004) notion of Relevance Theory.

who among potential speakers performs a particular conversational act informs the illocutionary force (intention) of the preceding statements (D. Gibson, 2008). Despite a leader influencing linguistic variables more strongly, interactions still result in convergence from both parties (Scholand, Tausczik, & Pennebaker, 2010). That is, non-dominant conversational partners still influence the leader, simply to a lesser extent.

Alignment provides a collective narrative emanating from the affected area. I suggest that this narrative relies on a perceived match to the environment among the participants, revealed in their collective departures from normative language behavior. In sum, the model that guides my analysis of social media is triadic, distinguishes language from thought as an intention-laden artifact, and multi-agent, driven by the articulation of breach to establish alignment of grounded situation models.

Empirical Literature

Some of the relevant literature is best understood as providing empirical constraints on the processes that intervene between the world and language. In this section I organize this additional literature in terms of a) perception and action, b) production and comprehension, and c) mediation. These correspond to the processes identified in Figure 2.

Perception and action. Viewing the *function* of humans as sensors within the broader system (Sheth, 2009) risks misunderstanding the *processes* of human experience. Many different lines of research both within and outside of psychology counter the temptation to view human perception as sensor-like. Hong and Page (2008) distinguish between generated and interpreted signals. Generated signals map sensor values to world conditions. For example, an old-fashioned thermometer works because we understand the precise response of a set volume of mercury to an increase in temperature, graduating the side of the vial accordingly. In contrast, interpreted signals require a knowledge base. Crucially, this knowledge differs between agents, causing a complementary divergence in perspective. These processes, reflecting multiple perspectives, provide the well-established benefit to accuracy from wisdom of the crowds (Parunak, Brueckner, Hong, Page, & Rohwer, 2013).

Attempts to understand human perception as sensors trace to the earliest days of psychology as a science. Fechner (1860/1912) sought to define the relationship between the magnitude of physical stimuli and the intensity of perception. Derived from a combination of psychophysical methods (the method of limits, method of adjustment, and the method of constant stimuli), Fechner's Law states that the perception of a stimulus is proportional to the natural log of that physical intensity ($S = K \ln I$, where K is a modality-specific constant). While psychologists have argued the specifics of this relationship (e.g. Stevens, 1961), the crux of the argument stands without serious challenge. That is, although physical phenomena and human experience have a quantifiable relationship, it is non-linear.

J. J. Gibson (1977) argued that perception is inseparable from human action in an environment. Thus, observers naturally perceive how their environment affords or constrains their action, to pass through, reach, sit on, etc. More generally, people engage with a physical world, concerned more with qualitative states and processes than the quantitative measures that captivate the psychophysicist. For example, people are more likely concerned with whether a heated container will explode than its objectively measurable temperature, or where a system in motion will come to a final state rather than the coefficient of friction (Forbus, 1984).

Neisser (1976) adds a role for cognition in this perception action coupling. Accordingly, schemas direct exploration, which samples from the available information in the environment, in turn modifying the existing schemas. “When a subject reports verbally about an image, he is really reporting quite literally what he—or at least his visual system—*is prepared* to see. The referents of language about images are possible perceivable objects in the environment” (Neisser, 1978, p. 100). Neisser notes the ease with which we can track the motion of a running animal, which must involve some anticipation, some schema of animal movement, that directs our eyes toward an expected position. In the same way, schemas motivate language behavior (Roy, 2005).

Observers, then, will not likely report on the objective, physics inspired features of their environment. Rather, a sizable portion of their messages should reflect an egocentric and action-oriented perspective. Specifically, they may indicate the constraint on action (e.g., “Hurricane might stop me from going”, taken from a social media post in Hurricane Sandy), useful for both the user, recipient, and potentially, emergency personnel attempting to anticipate behavior.

In agreement with Sheth and Thirunarayan (2012) as well as Hong and Page (2008), I argue that conceptualizing humans as sensors does not properly acknowledge their knowledge-based interpretive processes. Consistent with Norman (1988), I target human interpretation specifically with respect to wide-ranging affordances of and constraints on behavior. This is decidedly advantageous, in that the specifics of disaster (e.g., earthquake or flood) may be largely overwhelmed by the transcendent consequences to pervasive human need (e.g., health, food/water supply, housing, transportation, communication). Because disasters disrupt normal functionality (Peacock & Ragsdale, 1997; Perry, 2007; Smith & Belgrave, 1995), and in particular social functionality (e.g., Erikson, 1995), I seek linguistic metrics that indicate compromised daily life through a social medium.

Language production and comprehension. Psycholinguists have distinct interests from linguists. Linguists partition inquiry into three topics: syntax, semantics, and pragmatics. *Syntax* concerns how words fit together to form sentences (e.g., Chomsky, 1957). For example, conjunctions and disjunctions allow for the concatenation of sentences and place agreement constraints between words—verbs require objects, prepositions require nouns, etc. *Semantics* focuses on the meaning of words as well as propositions (e.g., Cruse, 1986). Classically, semantics is the relationship between propositions and the set of world conditions in which that proposition is true. *Pragmatics* investigates the influence of context on meaning, encompassing speech act theory and conversational analysis (e.g., Mey, 1993).

Chomsky is arguably the most influential contemporary linguist. Chomsky's generative grammar (1965) enormously influenced the above partition by demonstrating syntax as separable from meaning, illustrated by the oddly comprehensible sentence "Colorless green ideas sleep furiously." Generative grammar illustrated how a finite number of rules can support the production and comprehension of an infinite set of utterances. While partitioning syntax from semantics, Chomsky also maintains Saussure's theoretical separability of both from the environment. This is a natural consequence of his objection to Skinnerian accounts of the role of feedback in language acquisition (Chomsky, 1959). The contribution of Chomsky takes the form of a competence model, i.e., the linguistic knowledge that native speakers possess about how words fit together ideally.

In contrast, psycholinguists are concerned with performance models, i.e., the specific processes that produce and comprehend language. Nevertheless, psycholinguists generally respect the Chomskian partition of syntax and semantics. Inspired by generative grammar, psycholinguists focus on questions such as: *How does the mind produce and comprehend novel language?,* and *What processes and mechanisms underlie this ability (e.g., syntax and semantics)?* However, the bulk of psycholinguistic research focuses on the investigation of comprehension, which affords greater control over the stimuli and therefore stronger inferences about causality. Psycholinguists debate whether we resolve potential phrase structure ambiguity (and similarly polysemy) at the end of sentences or pursue one pathway in real time, for example, garden path sentences such as the classic "The horse raced past the barn fell" (Dynel, 2009). Processing delays result from applying one phrase structure while reading only to have it contradicted by later syntactic cues, forcing a reevaluation, and thereby settled the psycholinguistic debate.

Controlled studies in production prove more difficult. These researchers commonly record utterances or collect fMRI readings as participants attempt to provide a particular word or write a sentence. The physical environment in production research context typically consists of no more than simple, closed-world, controlled stimuli such as a set of ambiguous shapes (Clark & Wilkes-Gibbs, 1986) or pre-defined problems (Straus & McGrath, 1994) that constrain intentionality and focus of attention. Further, researchers typically impose motivation on the participants artificially, e.g., by assigning a task with some reward for completion. The production tradition has pursued issues such as the role of priming on syntactic choice (e.g. Bock, 1986), or deciphering strategies used to recall words while learning a second language (Chamot & Kupper, 1989). This research tradition also seeks to identify the mechanisms of production, with Dell (1986) and Levelt (1993) providing two of the most influential models for the parallel and serial views of individual components, respectively.

Crucial for my argument is the separability of message content, syntax, and lexical choice characteristic of psycholinguistic theory and portrayed in Figure 4 (Pickering & Garrod's work being notably influenced by that of Dell and Levelt). Here I note that both syntax and lexical choice contribute to message articulation. This allows for different syntactic structures and lexical choices to correspond to highly similar content, e.g., "I wish the rain hadn't started" versus "I wish the rain would stop." These differences seem to reflect intentionality in message design related to how an audience will process the form and content (Horton & Gerrig, 2005). This consideration of recipient comprehension mechanisms suggests that the analysis of decontextualized production mechanisms, regardless of the level of detail, cannot predict linguistic behavior well enough on its own to reverse engineer language output and infer the conditions to which it corresponds, i.e., the grounding. That is, a thorough examination of the relationship between language and the world must account for the broader conversational framework, including intentionality, in which the language arises.

Research with fMRI suggests that production and comprehension share a neurological network (e.g., Opitz, Müller, & Friederici, 2003; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012), supporting a unified approach to studying these processes, and some researchers now consider them collectively (e.g., Tooley & Bock, 2014).

Considering production and comprehension as linked processes allows the study of one to inform hypotheses in the other. Clark (1969) explored the influence of lexical alternatives on comprehension, known as lexical marking, as a psychological phenomenon, defining foundational adjectives as those stored in memory in a more simple and accessible form than their “marked” antonyms. Response time methods revealed that marked adjectives in antonym pairs required longer processing time. Subsequently, Gilpin (1973) studied the use of bipolar rating scales such as “good” – “bad”. He compared ratings with bipolar scales to ratings with lexically unmarked unipolar scales (“good” – “not good”) and lexically marked unipolar scales (“bad” – “not bad”). He found that unmarked unipolar adjectives more closely resembled bipolar ratings than did marked unipolar adjectives. Increasing the difficulty of the task by imposing time constraints produced similar, but attenuated, results. This seems to indicate that comprehension differences arise from a structural asymmetry of the scales. Negation then can be used in conjunction with marked adjectives (“not bad”) to convey meaning distinct from the unmarked alternative (“good”), while negation used in conjunction with the unmarked (“not good”) would more closely resemble the marked (“bad”). How this asymmetry with respect to negation manifests in stressful events as conveyed by social media remains an open question.

Consistent with the Gricean maxim of manner (1975) speakers should prefer the least obscure expression, i.e., the unmarked option. The cognitive simplicity of an unmarked antonym suggests that it will be more frequent than its marked counterpart, and in many cases this is true when evaluated against word use baselines. Other evidence suggests developmental asymmetry in acquisition (e.g., E. V. Clark, 1971; Donaldson & Wales, 1970). This implies that the selection of the marked adjective and its attending processing consequences is intentional, for example to direct attention and convey emphasis.

Nevertheless, word choice should also respect informativeness (Grice, 1975). Baseline English proportions reflect a prevalence of the adjective “big” over its opposite “small” (as measured by the GloWbe database [Davies & Fuchs, 2015]), presumably because big things are informative. A particularly dangerous, transient storm should promote words pertaining to above average size and temporarily magnify the disparity with respect to baseline usage. Similarly, “bad” experiences should promote the word “bad” over “good”. Word choice in social media therefore cannot be a purely cognitive phenomenon, as it must to some extent reflect the environment of the message producer.

Mediation. Mediation of language by digital interface such as computers and mobile phones deserves independent consideration. Certainly, speech differs from written communication. Speech itself benefits from face to face contact, which allows monitoring for comprehension (Clark & Krych, 2004; Clark & Schaefer, 1989). Even speech over the telephone permits interruption. Primarily text-dependent social media therefore likely influence message properties, particularly Twitter with its character limit.

Social media also particularly support the participation of multiple contributors. Kiesler, Siegel, and McGuire (1984) compared face-to-face, anonymous mediated (via computer chat windows), and non-anonymous mediated communication (via the same chat interface) in choice–dilemma problems. They found that whether an identity is attached or not, a single party is less likely to dominate a conversation in mediated exchanges, with all parties more likely to contribute something. The researchers also noted reduced efficiency in reaching consensus and reduced inhibition (more swearing and general hostility) with mediation. Straus and McGrath (1994) compared mediated communication to face-to-face in group performance on a range of tasks. They found that as levels of interdependence needed to complete a goal increased (from additive idea generation, to intellective reasoning, and finally to a collective judgment task) the *quality of solutions* stayed relatively constant over mediation. However, mediated communication proved slower, with more misunderstanding. Participants preferred face-to-face in all but the least interdependent task. Mediated communication groups performed best when participants attempted to generate small pieces of independent content, in fact consistent with Twitter's short messaging.

Social media readily accommodate the findings of both written and spoken language research. Cleland and Pickering (2006) argue that the same mechanism is responsible for syntactic encoding in both spoken and written productions. Also, Hartsuiker and Westenberg (2000) demonstrate that word order remains largely unchanged between modalities. Social media rely (primarily) on written products, inviting consideration of deliberation and planning in the shaping of sentences (Bereiter & Scardamalia, 2013), while also exhibiting characteristics of conversation more typical of spoken language research paradigms. There exists evidence of conventional conversational markers in 37% of tweets (Ritter, Cherry, & Dolan, 2010). Purohit et al. (2013) argue for the pragmatic conversational characteristics of Twitter, expanding the definition to accommodate the implied conversation supported with the *Reply* and *Mention* features. Likewise, Danescu-Niculescu-Mizil, Gamon, and Dumais (2011) found significant linguistic style matching in Twitter exchanges, like what one would expect in face-to-face conversations, even controlling for homophily (where people conversing tend to be similar already and thus use similar linguistic conventions) and topic accommodation.

While Twitter's idiosyncrasies (particularly character limits on message length) no doubt affect message content, the consequences are likely fortuitous. Holtgraves and Paul (2013) studied text messages versus telephone conversations unobtrusively and found that people tended to speak more simply via text, using more words associated with the personal and affective than they did in recorded phone conversations. Simple, direct, and personally centered messages likely allow for easier automated processing as well as greater likelihood of relevance with respect to the sender's immediate situation.

In summary, the empirical, process oriented literature in psycholinguistics informs the analysis of social media, including an emphasis on disaster-independent, function oriented content, appreciation for alignment at multiple levels of analysis between interlocutors, the separability of message contents from syntactic and lexical choice, and concern for the influence of media on the message.

Computer Science Approaches to Mining Social Media

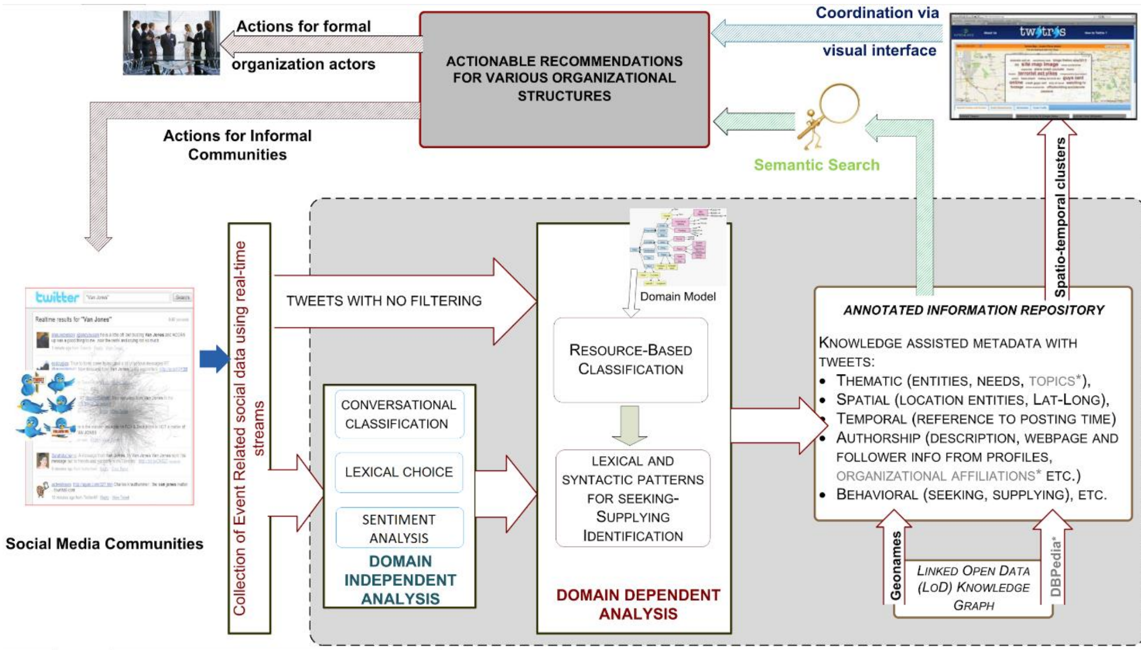

Figure 5. A process for filtering and formatting social media for emergency response. Event related social media data pass through domain independent analysis prior to domain dependent analysis that supports automated annotation for a searchable database. Adapted from "Identifying Seekers and Suppliers in Social Media Communities to Support Crisis Coordination," by Purohit et al., 2014, *Computer Supported Cooperative Work*, *23*(4-6), 513-545. Copyright 2014 by Springer. Adapted with permission.

Social media analysis constitutes a relatively new area of investigation, with early efforts focusing on defining and classifying behaviors, networks, and trends (e.g., Boyd, Golder, & Lotan, 2010; Huberman, Romero, & Wu, 2008; Java, Song, Finin, & Tseng, 2007). Sheth (2009) proposed a theory of citizen engagement wherein positioned observers report on their surroundings, somewhat analogous to the purpose of a physical sensor on the environment. Ideally such analyses identify and distill content to inform the formal response community regarding areas of need. The problem of course lies in the computational distillation of actionable content amid a firehose of mostly irrelevant posts. Figure 5, based on one presented in Purohit, Hampton, Bhatt, Shalin, Sheth, and Flach (2014), provides a basis for identifying many of the emergent issues. From a collection of social media data, Purohit et al. first posit domain independent analyses based on the presence of conversational features, to reduce the corpus to a more manageable size. From there a conventional domain dependent analysis may find actionable information. Subsequently, an annotated information repository and visualization software organize findings for the formal response community. I begin my overview of the relevant computer science literature with domain-dependent analyses, generally concerning the exploitation of keywords and human-mediated annotation. I follow these subsections with examples of domain-independent analysis.

Keyword Tallying. Tallying the prevalence of relevant and actionable keywords has proven a popular approach to information processing, through various stages of filtering, classifying, aggregating, etc. (e.g., Ashktorab, Brown, Nandi, & Culotta, 2014; Imran, Castillo, Diaz, & Vieweg, 2015; Imran, Elbassuoni, Castillo, Diaz, & Meier, 2013a; Imran, Elbassuoni, Castillo, Diaz, & Meier, 2013b; Jennex, 2012). Herein, researchers typically identify a target domain, then use a set of keywords to crawl target corpora to identify related, lexical entities from the bottom up. Lexical tallying and subsequent analysis purport to provide a summarized, computationally based representation of the situation.

The assumption is that a high absolute value of statements or mentions corresponds to a more pronounced presence of the subject of those statements in the designated area. However, the approach does not consider the base rate of the tallied variables. Also, the approach raises concern for veracity, particularly when the recognition of breach depends on infrequent, or even singular reports.

Moreover, the approach may reward exaggeration with attention. These message features are subject to intentional exploitation. There also exist examples of social media users extrapolating wildly from limited or dubious information, as in the aftermath of the Boston Marathon bombings when so-called “Reddit detectives” falsely identified completely innocent bombing suspects whose safety was then threatened (Lee, 2013). In disaster response, mischievous or uninformed rumors propagated via “retweet” (forwarded microblogs), could skew results to indicate false need for assistance. Such potential has motivated effort into characterizing the sender’s trustworthiness (e.g., Tapia, Bajpai, Jansen, Yen, 2011; Tapia, Moore, Johnson, 2013; Westerman, Spence, & Van Der Heide, 2014; Vedula, Parthasarathy, & Shalin, 2016)

Annotation. A further limitation of keyword tallying is lexical dependence. Without a supporting knowledge model, separate references to medical care such as “the hospital”, “the doctors”, or “urgent care” may not be recognized as high-frequency features in aggregate. This illustrates but one instance of a general problem in distilling relevance across the range of choice in language production. Recognizing the computational challenge this poses, some researchers advocate a role for human pre-processing to aggregated conceptually related content.

Prominent approaches to annotation of social media data include those of Palen, Starbird, and Vieweg (2010), who helped to define the interaction of official and publicly-generated information in disaster contexts, with users typically seeking out official reports and often appending it with their own eye-witness accounts or other more locally pragmatic details. Subsequent work addressed the ability to leverage digital volunteers to impose a *post hoc* structure on messages to afford processing and comprehension by the formal response community (Starbird & Stamberger, 2010; Starbird, 2011). However, the approach relies on the questionable assumption that citizen observers have the domain knowledge to classify citizen observations properly, e.g., the implications of flooding for the integrity of the water supply, or the potential public health crisis when the absence of electricity means that well water can no longer contribute to household sanitation.

Sheth (2009) advocates a computational approach to annotation with respect to three dimensions: spatial, temporal, and thematic, by straightforward inference from computationally accessible resources in the medium. Some users enable geographic tagging that provides important context for interpretation. Social media also contain temporal metadata by default, thus readily affording analysis on that dimension. Thematic analysis, from hashtags and other keywords benefits from knowledge based analysis. Synthesizing many of these ideas, Purohit et al. (2014) constructed a system to leverage emergent resources (suppliers) and match them to those in need of aid (seekers). This contributes to domain modeling and message analysis beyond the lexically bound approaches to tallying.

Other approaches to the location of key information rely less on specific content, which Purohit et al. identify as domain-independent analysis. These include sentiment analysis and conversation analysis.

Sentiment analysis. Sentiment analysis (e.g., Pang & Lee, 2008; Wang, Chen, Thirunarayan, & Sheth, 2012) exemplifies domain independence. Rather than tallying absolute frequencies of specific lexical items, sentiment analysis typically calculates the proportion of positive linguistic features to negative ones in proximity to a target entity (usually expressed as a keyword or hashtag). Investigators can then compare the proportion to some baseline or competing entity (for example, sentiment relative to stock prices over a matched period of time, or one politician against another). Higher proportions indicate more positive views expressed in the Twitter conversation overall. This approach has helped predict movie box office performance (Liu, Huang, An, & Yu, 2007; Mishne & Glance, 2006), political variables (Tumasjan, Sprenger, Sander, & Welpe, 2010; Wang, Can, Kazemzadeh, & Bar, 2012), and stock prices (Yu, Duan, Cao, 2013).

While sentiment analysis has resulted in the successful prediction of behavior, the relationship of sentiment to environmental conditions is unclear. Caragea (2014) maps sentiment metrics to the spatial and temporal properties of disaster events, generally supporting the claim that proximity is associated with negative sentiment. However, this pattern ignores widespread positive and prosocial behaviors of those most directly impacted (e.g., Rodríguez, Trainor, & Quarantelli, 2006). It also contrasts with disaster response theory, which tends to emphasize a positive, prosocial response in the affected population (Quarantelli, 1986). As Quarantelli makes clear, "organizations do worst when they assume the worst about human beings in a disaster" (1986, p. 3).

I suggest that the environment (e.g., a disaster scenario) influences the patterns of lexical choice in a more nuanced manner than sentiment readily captures. This is reflected both in the long term (through word baselines based on daily life), and in the short term (via disaster-mediated departures in lexical choice). Moreover, the multiple functions of speech (e.g., assertion versus promise [Searle, 1976]) muddies reliance on overall corpus sentiment as the sole (or even predominant) metric of human experience. Therefore, other measures of lexical choice beyond sentiment may reflect strong environmental influence on a community.

Conversation analysis. Purohit et al. (2013) emphasized the conversational properties of social media. They explored the conversational properties of a social media platform such as the presence of deictic pronouns as a screening heuristic for subsequent content analysis. They demonstrated that tweets containing conversational features contained more sensory experience, social interaction, and communication content than tweets without conversational features. Domain independent conversational features may serve a useful pre-screening function in the processing of social media.

Implications for Mining Social Media in Disasters

The psychological processes that underlie language production suggest complementary approaches to the conventional computational focus on tallying and annotating specific message content. This dissertation extends domain-independent, conversation-based screening presented in Purohit et al. (2014) to include lexical choice, to supplement disaster-specific methods. This work complements the characterization of disruption based on the simple tally of signal counts.

Twitter allows for a massive, computationally accessible corpus specific to surrounding events, locations, or topics designated by the researcher. Stable platform conventions combined with search functions afford manipulation of corpus subsets, such as eliminating retweets. In addition, studying participant production directly precludes any possible loss or bias in transcription. As such, the short message service platform reduces noise that may affect the interpretation of intentionality.

While the psychological research summarized above has a "micro" approach to language, the properties of social media afford analysis of *patterns* of language behavior. My approach relies on style patterns across a corpus to identify a variety of human interpretive sensors that respond to different breaches of normal functioning in the disaster context. The focus on style allows me to employ a baseline standard for evaluating observed patterns and supports comparison across varying population bases and event types. These measures will complement other sentiment and conversational analysis techniques by identifying sentinels of breach of normative conditions. The macro approach, combined with an elimination of retweets, displaces the reliance on trust and accuracy common to social media analysis by averaging over the population.

I suggest a heuristic-based method by which computationally inexpensive and rapid processing of linguistic cues within messages may identify significant breach of canonicity. The message author's perceptual and cognitive processes shape how he detects and interprets a blocked road or a downed power line, determining message production. These messages are intended to align the community's understanding of the situation with his. I shift the metaphor for human processes from sensor to narrative, to acknowledge the influence of knowledge and intentionality on message production. Borrowing from both Pickering and Garrod's (2004) conception of alignment in situation models, narrative processes tied to the environment (specifically the notion of breach) constitute both a mechanism and the results of that alignment. Individual posts contribute to an overall narrative structure by reporting on environmental affordances and constraints. Mutual understanding is therefore a mutually constructed narrative to explain circumstances. The current investigation essentially continues the spirit of psychophysics, mapping different physical stimuli to linguistic measures. Rather than steady light sources and audio tones, I aim to investigate detection of the problematic in a physical, socio-cultural environment. The long-term goal is to reason backward from how people are formulating messages to the identification of conditions in the world.

Here I describe the empirical implications of this view. Instead of asking participants for ratings on a constructed scale, I infer potential psychophysical metrics based on linguistic output with reference to a base rate. My purpose is not to substitute social media for individual calls for specific assistance, but rather to filter and mine the pattern of commentary to ascertain the degree of breach and in turn focus response. Below I suggest the properties of an informed analysis. My approach focuses on selection of events, consideration of base rates, geographic proximity, and metrics based on aggregated lexical content. To this end, event type and proximity to those events constitute the independent variables that should correspond with patterns of lexical choice, as I will discuss in greater detail below.

Research Question 1: Does the nature of an event (including disaster context, control over outcome, and sentiment) influence linguistic behavior?

Research Question 1a: What event differences influence linguistic behavior?

Event influences. The correlation of language behavior with ground truth requires variability in ground truth, particularly with respect to the presence of breach, and the resulting uncertainty and stress. While the construct of breach should produce similar patterns in any disaster scenario, there may exist correlations between some type of disaster and a particular word choice that skews results. To avoid this potential confound and ensure the measurement of a broadly defined breach of canonicity rather than, say, some idiosyncrasy in the response to inclement weather in Oklahoma, I include a variety of events. These include both natural and manmade (terrorist) events in different regions of the United States. The disaster scenarios I chose were Hurricane Sandy along the Northeast coast, tornadoes in central Oklahoma in 2012, and the Boston Marathon bombing, also in 2012. Hurricane Sandy and the Oklahoma tornado represent natural hazards. The Boston Marathon bombing represents an intentional, man-made conflict (Quarantelli, 2005) that serves to test the generality of my findings.

Control over outcome. As practical control conditions, I suggest identifiably and distinctly non-disaster contexts, invoking a stress response that may influence language choice, but without the type of breach common to disasters. For my control corpora, I collected two datasets both including and excluding geographic boundaries on a similar scale. In one, I randomly selected 50,000 tweets from a location-independent corpus assembled from search terms related to fantasy football. Crucially, participants in fantasy football can make choices to impact their outcome and reduce their stress. Fantasy football is, legally, a game of skill (Humphrey v. Viacom, Inc., 2007), indicating that the player's choices of trades, sits, and starts have directional non-random changes on the outcome. For the second, I crawled Twitter for the 2016 Major League Baseball World Series between the Chicago Cubs and the Cleveland Indians. As in disasters, participants are subject to precipitating circumstances outside their control. Unlike disasters, no participant action controls the World Series outcome—the classical conditions for stress.

Sentiment. To unconfound sentiment with event type, I divide the World Series datasets into winning and losing streaks (based on the combination of games won and the winning home city, switching which city's tweets fall into which category at each streak's end). This constitutes a single event (avoiding confound variability with topic, context, etc.) encapsulating both positive and negative sentiment—eustress and distress—that is easily distinguishable based on geographical indicators and likely communal attachment to a given city's home team. Sorting tweets by winning and losing decouples the notion of breach from sentiment valence. Further, the conversational patterns of the losing team's city may present a kind of intangible disaster for the most ardent of fans, potentially helping to define a sliding scale of community response to adverse events generally. I acknowledge the likelihood of patterns adhering to a communal narrative structure similar to that of disasters. Nevertheless, deviations elicited by true disaster contexts should prove distinguishable in both manner and magnitude. I identify the specific linguistic measure employed in Research Question 2.

Research Question 1b: How does event influence linguistic behavior: proportions of paired antonyms in situations of breach compared to an Internet-specific base rate—is the departure general or specific to antonym pairs?

Base rates. One of the limitations of the tallying approach to mining is the absence of base rates to scale observed responses. A single report is therefore assumed to be meaningful (provided trustworthiness is established). In contrast, psychological science does not typically invest single data points with meaning, but rather depends on statistical analysis to detect discrepancies from chance occurrence. Word baseline databases provide base rates for word frequencies, including databases specific to Internet language, e.g., the GloWbe database (Davies & Fuchs, 2015). These aggregate language behavior across all web-based circumstances over extended periods of time and provide a basis for estimated chance occurrence. Still, absolute frequencies during an event are not meaningful when compared to aggregated frequencies. To overcome magnitude issues, I examine proportions, in particular proportions of lexical choice between antonyms.

Lexical choice. As justified above, lexical choice corresponds to a combination of the environment and of schemas that capture the environment's relationship to normal functioning. I operationalize breach in relation to baseline conditions by exploiting the phenomenon of lexical choice. The lexicon offers alternatives for creating emphasis. When discussing size, standard language patterns in English show a baseline preference for the adjective "big" over its opposite "small". An increased relative prevalence of "big" likely indicates the increased prevalence of something big in the world. I treat negation as orthogonal to this choice, equally affecting the prevalence of "big" (as in "not big") and "small" ("not small"). A separate analysis of negation will address this assumption. Rather than rely on raw frequency counts as a metric of need, I compare the observed proportions of paired lexical alternatives in a disaster social media corpus to baseline proportions that provide a base rate. This Bayesian-inspired approach shifts the focus from individual reports to the altered pattern of such reports that reflect the combined influences on lexical choice, including lexical marking (Gilpin, 1973), Gricean maxims (Grice, 1975) and conversational alignment (Levelt, 1993). I suggest that message filtering and the general assessment of need is informed by comparing observed lexical choice between paired antonyms to word baselines, beyond that gleaned from expressions of sentiment.

Disaster response situations should provide the requisite breach of canonicity to serve as a magnifier for linguistic variables sensitive to stress in an individual, including lexical choice. Stress situated in a breach event provides motivation (if only to remove the stressor). The larger an event and the less ambiguous the goal (e.g., clear in a storm, conflicting in an election), the more observers within the affected area should converge on shared lexical choices across the corpus. Regardless of the relative strength of contextual influence or marked language emphasis, decisive changes in patterns of language behavior will arise in meaningful and detectable ways. I therefore decline to make a directional hypothesis, but instead propose a conservative metric to identify relatively certain changes attributable to salient environmental factors and the corresponding increase in stress of message producers.

Research Question 2: What impact does spatial proximity to a disaster have on language behavior (e.g., authenticity, sentiment)?

Just as the nature of the event may influence how observers express breach, an observer's proximity to that event should do the same. Certainly, the articulated concerns of someone with a flooded ground floor would diverge from the ubiquitous "thoughts and prayers" pouring in from afar. Authentic, situated concern should diverge in many ways from those simply referring to the event, including specificity and discussion of functionality. To identify the disparities in message content, I separate the disaster corpora above into rough boundaries of those in close proximity to the hardest hit area, those in the broader community, and those situated spatially beyond the general geographic area. These are rough approximations, not meant to set definitive boundaries but merely to identify a coarse gradient of exposure and corresponding reaction. While this relationship likely resists a "border" between those directly versus distally impacted, a gross aggregation of these broad categories may point toward the critical linguistic patterns that serve to differentiate. Resolution is a function of corpus size.

Lexically-aggregated metrics. RQ2 employs lexically aggregated metrics in lieu of antonym proportions for several reasons. When a corpus is partitioned for the analysis of geographic proximity, frequency counts plunge. Aggregated LIWC metrics support a study of behavior with a finer grained spatial analysis, when the base rates of individual lexical items become too low to detect trends. Linguistic Inquiry Word Count software (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015), provides a kind of knowledge-based tallying regarding categories for psychological processes (to indicate message authors' consideration of mental activities), spatial and temporal relativity (suggesting concern for dynamic properties), biological processes (suggesting concern for physiological rather than cognitive topics), and authenticity, all of which would likely vary in use between those immediately versus distally impacted by an event. LIWC classifies words into one or more categories, counts the number that occur for each, and presents that number as a percentage of total linguistic output. LIWC has proven useful in social media analytics for its rapid and naïve approach (e.g., Chen & Sakamoto, 2014; Kryvasheyeu et al., 2016; Landwehr & Carley, 2014; Purohit, Hampton, Shalin, Sheth, Flach, & Bhatt, 2013) that does not require extensive domain modeling or advanced computation.

Moreover, specific LIWC dictionaries address two issues raised earlier: 1) the role of sentiment in the analysis of social media and 2) the relationship between negation and lexical choice. LIWC can perform sentiment analysis (e.g., Pang & Lee, 2008; Wang, Chen, Thirunarayan, & Sheth, 2012) through measures for positive emotion, negative emotion, and aggregated tone[7]. This allows me to examine differences between breach and sentiment valence, revealing value added by consideration for antonym pairs beyond that of more conventional sentiment analysis approaches. LIWC also allows me to

[7] According to the LIWC User Manual (Pennebaker et al., 2015a), "tone" measures a continuum from positive and upbeat to anxious and sad.

investigate the relationship between antonym selection and the manner and density of content.

Research Question 3: Do high frequency members of the antonym pair correlate with content (e.g., sentiment, negation, motion processes)?

While investigating the processes of message generation carries import in its own right, the applications motivation for this work was that high frequency members of the antonym pair provide a useful heuristic for finding content useful to response agencies.

Content quality and density. Other LIWC measures relating to physical movement and processes may indicate that messages containing sentinels of breach also contain higher density of content that is relevant to those monitoring the disaster. Categories such as function, motion, perception, biological, and cognitive processing may shed light on this relationship. However, there may be a role for more specific analyses related to compromised functionality, such as transportation terminology. An examination of individual tweets addresses this issue.

Analysis Approach

My approach to distinguishing disaster from non-disaster corpora requires a preliminary evaluation to ensure that disaster and non-disaster corpora vary sensibly in relation to one another on predictable, theoretically sound variables (RQ1). Having established that, I will compare how event types influence language relative to other events (RQ1a). RQ1b employs paired antonyms, for which I use an objective external reference for each word in every pair's base frequency of Internet use to investigate departure in different event settings.

Complementing these analyses, the more aggregated nature of LIWC metrics allows for investigation broken down by geographical proximity to critical events without a compromising reduction of power (RQ2). LIWC content analysis allows me to address authenticity, negation, and sentiment. Finally, I will examine the words that reliably increase in frequency during disaster contexts (relative to their paired antonyms) and determine if the messages in which people elect to use them increase consistently on content measures like sentiment, negation, or motion processes (RQ3).

II. METHOD

Datasets

I collected several million unique tweets (i.e., omitting those with markers indicating they were retweets) from three disasters in different regions and of distinct types across the United States. I eliminated “retweets” or forwarded messages because they are likely heavily influenced by organizational reports, which contaminate my interest in personal narrative (Starbird & Palen, 2010). I used a social media analysis tool, Twitris (Purohit & Sheth, 2013), to identify the tweets within the target time frames (see Table 1 for inclusion criteria). As is typical of social media data, location specific corpuses constitute a small subset of the full data stream. I further segmented the resulting corpora according to location tags specified in Tables 2 and 3, isolating proximal and intermediate corpora, with the remaining location tagged tweets constituting the distal corpus. In choosing these bounding boxes, I balanced the size of the corpus, the frequency of pairs, and region size.

Table 1

Inclusion Criteria for Tweet Datasets

Event	Start	End	Crawling Word Set
Hurricane Sandy n ~ 4.6M	2012 Oct 27	2012 Nov 07	Hurricane Sandy, Frankenstorm, #Sandy
Boston Marathon Bombing n ~ 4.5M	2013 Apr 15	2013 Apr 25	Boston explosion, Boston explosions, Boston blast, Boston blasts, Boston tragedies, Boston tragedy, PrayForBoston, Boston attack, Boston attacks, Boston terrorist, Boston terrorists, Boston tragic, Boston Marathon, Boston Marathon, Boston explosive, Boston bomb, Boston bombing
Oklahoma Tornado n ~ 2.8M	2013 May 20	2013 May 30	Oklahoma tornado, Oklahoma storm, Oklahoma relief, Oklahoma volunteer, Oklahoma disaster, #Moore, Moore relief, Moore storm, Moore tornado, Moore flood, Moore disaster, Moore volunteer, #OKC relief, #OKC disaster, #OKC tornado, #OKC flood, #OKC volunteer, #OKC storm, #OKhaves, #OKwx, Shawnee, Norman, Pottawatomie, Mary Fallin, #OKC, #OKneeds, #OKhaves, #OK, #OKhaves, #OK tornado, #OK relief, #OK flood, #OK disaster, #OK volunteer, #OK storm
Fantasy Football n ~ 1.0M	2016 Sept 12	2016 Oct 11	NFL NFLFantasy DawgPound RiseUp SieTheDay GoPackGo Skol WhoDey FlyEaglesFly KeepPounding Jaguars Patriots Broncos Chargers Chiefs RaiderNation ForTheShoe GoNiners WeAreTexans OnePride GiantsPride GoBills TitanUp JetUp MobSquad RavensFlock WeAre12 BeRedSeeRed FinsUp HereWeGo FantasyFootball Seahawks Bengals Falcons CowBoys Texans 49ers Titans Redskins Vikings Buccaneers MiamiDolphins Eagles Steelers Cardinals
World Series n ~ 4.5M	2016 Oct 20	2016 Nov 7	World Series, WorldSeries, ALSC, NLCS, baseball world series, Baseball, MajorLeagueBaseball, Cubs, #FlyTheW, chicago cubs, #GoCubsGo, #MLB, #ChicagoCubs, Wrigley, Wrigleyville, #1908, #CubsWin, Cubbies, NLChampions, #NLChamps, National League Champs, National League Championship Series, #NationalLeagueChampionshipSeries, #NationalLeagueChamps, #fowlershowlers, #postseason, #CubsNation, #Game7, #HelloCleveland, #WorldSeries, #game7herewecome, #cubsvsindians, #iaintafraidofnogoat, #thecurseisover, #goodbyesomeday, Indians, #ClevelandRocks, AL Champs, #ALChamps, AmericanLeague, American League, #RallyTogether, Rally Together, AL Chamption, #ALChamption, Progressive Field, ProgressiveField

Table 2

Coordinates of the Bounding Boxes Used for Direct Comparison between Events

Event	Southwest Corner		Northeast Corner		*N*
	Latitude	Longitude	Latitude	Longitude	
Hurricane Sandy	39.270	-74.612	41.327	-71.816	146,764
Boston Marathon Bombing	42.022	-71.802	42.865	-70.572	54,348
Oklahoma Tornado	34.551	-98.465	36.008	-96.597	45,788
World Series – Chicago	41.469	-88.241	42.160	-87.209	88,701
World Series – Cleveland	41.112	-82.396	41.788	-81.003	33,071

Note. As Hurricane Sandy directly impacted a much larger area than the other disasters, I use only the smaller bounding box, where the Boston Marathon Bombing and the Oklahoma Tornado use the larger bounding box.

Table 3

Coordinates of the Bounding Boxes Used for the "Doughnut" Analysis of Events

Event		Southwest Corner		Northeast Corner		*N*
		Latitude	Longitude	Latitude	Longitude	
Hurricane Sandy	Small	39.270	-74.612	41.327	-71.816	146,764
	Doughnut	24.857	-87.931	47.469	-66.952	212,115
Boston Marathon Bombing	Small	42.332	-71.111	42.370	-71.053	33,977
	Doughnut	42.022	-71.802	42.865	-70.572	20,371
Oklahoma Tornado	Small	35.250	-97.653	35.400	-97.319	4,383
	Doughnut	34.551	-98.465	36.008	-96.597	41,405
World Series—Chicago	Metro Area	41.469	-88.241	42.160	-87.209	88,701
World Series—Cleveland	Metro Area	41.112	-82.396	41.788	-81.003	33,071

Note. Tweets in my dataset with geotags outside of these boxes include 366,604 from Hurricane Sandy, 974,314 from the Boston Marathon bombing, 477,336 from the Oklahoma tornado, and 190,374 from the World Series.

Using only tweets with geolocation information for the World Series analysis, I first eliminated all tweets arising from areas outside of either Chicago or Cleveland based on a spatially generous, but fairly arbitrary bounding box (see Figure 6). Then I divided the two corpora (Chicago and Cleveland) into winning and losing streaks (Tables 4 and 5). Winning streaks began after each team had clinched their respective leagues (i.e., secured a spot in the championship) and continued until that team's first loss (game one for Chicago, game two for Cleveland), with time of loss operationally defined as two hours after the game’s scheduled start time. To obtain a combined corpus of wins and losses, the city of collection for each category switched with each snapped winning streak. For example, after a Cleveland win, all tweets emanating from that city appear in the “wins” corpus and all from Chicago appear in the “losses” corpus until Chicago won a game, at which point the classification switched. The final “streak” continued from the end of game five, when Chicago fended off elimination, through their ultimate victory, and on until the United States’ election day began on November 8th, when factors not related to baseball were likely to contaminate any findings.

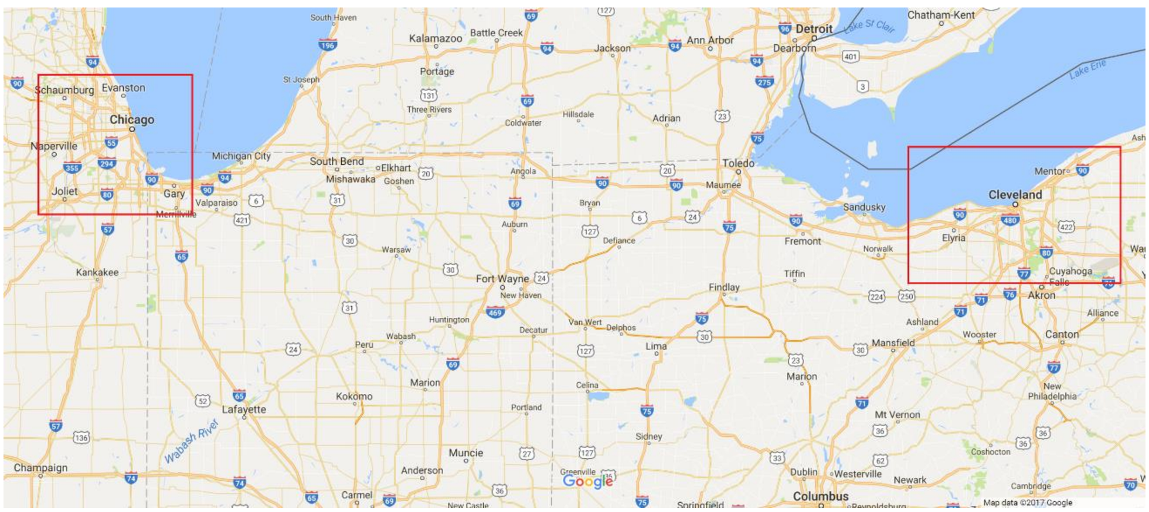

Figure 6. Bounding boxes for the World Series, including Chicago- and Cleveland-centered regions.

Table 4

Compilation of corpora samplings that constitute the ‘World Series Wins’ corpus

Source City	Start	Stop	n
Chicago	Oct. 20 16:44	Oct. 26 04:10	27,749
Cleveland	Oct. 20 16:44	Oct. 27 03:08	12,589
Chicago	Oct. 27 03:08	Oct. 29 04:08	13,296
Cleveland	Oct. 29 04:08	Oct. 31 04:17	7,819
Chicago	Oct. 31 04:17	Nov. 08 06:00	28,353
Total	Oct. 20 16:44	Nov. 08 06:00	89,806

Note. All times are in Greenwich Mean Time. The first time starts for both teams when both pennants had been secured.

Table 5

Compilation of corpora samplings that constitute the 'World Series Losses' corpus

Source City	Start	Stop	n
Chicago	Oct. 26 04:10	Oct. 27 03:08	5,440
Cleveland	Oct. 27 03:08	Oct. 29 04:08	4,146
Chicago	Oct. 29 04:08	Oct. 31 04:17	13,863
Cleveland	Oct. 31 04:17	Nov. 08 05:00	8,517
Total	Oct. 26 04:10	Nov. 08 05:00	31,966

Note. Different end times between the wins and losses corpora reflect the same time point in different time zones: midnight the morning of Election Day in the United States

Antonym Pairs

Previous research into marked language (e.g., Gilpin, 1973) provided a starting point for compiling the set of antonym pairs. Additionally, I consulted a list of common positive adjectives from the Oxford English Dictionary online and paired each with its most common antonym using www.thesaurus.com. Finally, I read through 100 randomly selected tweets for each disaster corpus and manually identified the adjectives used, later adding the corresponding antonym as above. I eliminated redundancies to complete the list of 36 pairs. The selected pairs appear in Tables 7 and 8 in the results section.

Word frequency baselines. The GloWbe database of Internet language (Davies & Fuchs, 2015) estimated the baseline frequency of words within the United States. I calculated a baseline proportion of use between each word and its matched alternate such that the less common word in the control corpus was represented as a fraction of the total use. For example, "all" appeared 1,306,886 times in the GloWbe corpus, while its marked alternate "some" appeared 724,227 times. Thus, the proportion of "some" to the pair total equals 0.36. This approach standardizes comparison across antonym pairs with different absolute frequencies.

Sentiment baselines. I used SentiStrength (Thelwall et al., 2010) to obtain sentiment scale values for each member of the antonym pair. I determined any difference in sentiment value greater than zero to be an affectively asymmetric pair and confirmed this determination against the valence values in the 14,000 item Warriner, Kuperman, and Brysbaert (2013) database when possible.

Tabulation

For the purposes of tabulating antonyms, I first used an open-source tokenization tool from Apache OpenNLP (2010) to separate each word from any adjacent punctuation to facilitate automatic identification without capturing word inflections that may carry different denotation or connotation (e.g., I wanted to capture the "real" in "for *real*!" but not in "*real*ly"). After tokenizing the text files, I ran each through a custom Python-based program that counted each instance of the target words, including ones that occurred at the beginning or end of a line (i.e., without leading or following spaces), and output the count into a table complete with the filename and target word.[8] This contrasts with the method reported in Hampton & Shalin (2016), which employed a simple search function that required empty space preceding and following the target word in order to be counted. The custom program employed here breaks words apart from punctuation for more thorough counting. This yielded two scores (one for each member of the antonym pair) against which to compare the corresponding two scores in the baseline database.

I calculated several other percentages using the LIWC tool (Pennebaker et al., 2015b) for further content analyses. LIWC tallies content ranging from linguistic

[8] A preliminary analysis with computationally unsophisticated searches resulted in slightly different word counts. These results were closely related in relative magnitude, but a few borderline word pairs only reached significance with the latter approach.

variables (personal pronouns, punctuation) to abstract concepts (social or environmental interaction) by counting the number of words in a target corpus that appear in a given LIWC category (e.g., every "me" counts toward personal pronouns) and calculating a percentage of that tally relative to the total number of words. All LIWC categories are organized into an overarching "dictionary". LIWC tallies also informed my pilot work.

Statistical Analysis

I conducted two sets of statistical analyses. The first set examined the pattern of lexical choice across events (RQ1, RQ1a, RQ1b). These analyses largely concern correlations and effect size departures from baselines. The second set concerned the spatial (RQ2) and quantitative (RQ3) distribution of content by lexical choice. RQ3 analyses employ an ANOVA framework.

Pattern of lexical choice. The goal of these analyses was to establish trends in lexical choice across events expressed as proportions, and identify the specific word pairs responsible for any observed trends. I calculated proportions of use between antonym pairs by taking the count of one alternative (in either baseline or event-specific corpora) and dividing that number by the count of that alternative (the same number) plus the count of its opposite. This way, the two proportions always add up to 1.0 (see Figure 7).

$$\frac{(\text{Base Count "some"})}{(\text{Base Count "some"}) + (\text{Base Count "all"})} = \text{Baseline Proportion}$$

$$\frac{724{,}227}{724227 + 1{,}306{,}886} = 0.36$$

Figure 7. Proportion formula for the less common "some" relative to "all" in baseline circumstance. Observed proportions are calculated the same way, with observed values.

To compare the proportions for an antonym pair with the word baseline corpus, I employed effect size metrics. The Cox Logit method (Lipsey & Wilson, 2001) compares proportions using *d* values. For example, the pair all/some has a baseline proportion of use of 0.36 (with “all” as the more common alternative; see Figure 7) and an observed proportion in the Hurricane Sandy corpus of 0.23 (shifting toward “all”). Consistent with the Cox Logit *d* value (Figure 8), I divided the observed proportion, 0.23, by its inverse, 0.77, and divided that by the baseline proportion, 0.36, over its inverse of 0.64. *D* increases with asymmetry in the contributing binary proportions, reinforcing the standard caution regarding over-emphasis on effect size (Cohen, 1988).

Moreover, although the source corpora are quite large, the number of instances of an antonym pair can be small, particularly as the geographic span of the corpus shrinks. As a result, a 95% confidence interval including zero may surround otherwise impressive *d* values. Determining a confidence interval (Figure 9) follows a similar procedure but takes into account the magnitude of the observed and baseline counts. I employed an effect size calculator (Wilson, 2000) to spot check both *d* values and the surrounding 95% confidence intervals.

$$\left(\mathrm{Ln}\frac{\left(\frac{\text{Observed Proportion A}}{1-\text{Observed Proportion A}}\right)}{\left(\frac{\text{Base Proportion A}}{1-\text{Base Proportion A}}\right)}\right)/1.65 \qquad \left(Ln\,\frac{\left(\frac{0.23}{1-0.23}\right)}{\left(\frac{0.36}{1-0.36}\right)}\right)/\ 1.65$$

Figure 8. Cox Logit *d* formula, instantiated for all/some in the Hurricane Sandy corpus

$$\text{Cox Logit } d\ \pm 1.96\sqrt{\frac{\left(\frac{1}{\text{Base Count A}}\right)+\left(\frac{1}{\text{Base Count B}}\right)+\left(\frac{1}{\text{Obs. Count A}}\right)+\left(\frac{1}{\text{Obs. Count B}}\right)}{1.65^2}}$$

$$\text{Cox Logit } d\ \pm 1.96\sqrt{\frac{\left(\frac{1}{724{,}227}\right)+\left(\frac{1}{1{,}306{,}886}\right)+\left(\frac{1}{2{,}880}\right)+\left(\frac{1}{9{,}669}\right)}{1.65^2}}$$

Figure 9. Cox Logit *d* 95% confidence interval formula, where “Base Count A” is the

count of instances of word A (“some”) in the baseline corpus; “Base Count B” is the same for word B (“all”); “Obs. Count A” is the count of word A observed in the target corpus; and “Obs. Count B” is the same for word B.

Several statistical concerns challenge the comparison of observed proportions between the corpora, such as assumptions of linear relationships, the underlying distribution of proportion values, and the fundamental correlation of the observed proportions with a baseline. I describe the relationship between two sets of observed proportions using a Spearman’s rank correlation and recover partial correlations between disaster corpora controlling for the common baseline value (National Institute of Standards and Technology, 2012).

Spatial proximity. I will examine changes in language behavior with respect to spatial event proximity, to demonstrate that linguistic behavior is a valid measure of speaker experience, changing in expected fashion. *D* values should attenuate with distance from the disaster epicenter. However, spatial proximity analysis requires segmenting the corpora. I am unable to conduct this analysis with the above pairs owing to the sparsity of geo-tagged data. The resulting small subsets generally result in prohibitively low frequency values for the observed pairs, with a preponderance of missing data and non-significant *d* values. Aggregated metrics, resulting from LIWC tallying, increase the base rates.

To examine trends with respect to spatial proximity I will change metrics from individual words to LIWC categories, each of which aggregates over dozens of words, and returns a value for the observed frequency in a percentage of total language. The target LIWC categories include language about (roughly) interaction with the environment (function, motion, relative, focus on the present), personal factors (cognitive processes, perception, personal pronouns, biological, ingest), manner of utterances (authenticity, clout, certainty, analytic, tone), content density (word count, six letter words), social interaction (affiliation, positive emotion), and a consideration for negation.

I will examine three subsets of data for each disaster event as indicated in Table 3: a small bounding box, a doughnut consisting of the remainder of content in the large bounding box less the content in the small bounding box, and a distal subset consisting of content known to originate outside the large bounding box. LIWC categories are not orthogonal, and I do not report all of them. Many of the trends across proximity are consistent for all three events. Most are consistent for Sandy and Oklahoma.

Distribution of content by lexical choice. Having established correlations in word choice between events, and attributing these to specific antonym pairs, I conducted quantitative content analysis, sentiment analysis, and negation analysis comparing the content in those tweets containing the high frequency member of the pair with the content of tweets with the low frequency member of the pair. All these analyses employed a three-way ANOVA, with *Pair*, *Event*, and *Hi/Lo Frequency* as independent variables. Pair refers to a specific subset of antonym pairs (nine of which appeared informative according to the above analyses), and Event is one of the six events in my study. Hi/Lo Frequency separates tweets according to whether they contained the member of those nine antonym pairs that increased or decreased significantly and consistently in use in the presence of a disaster.

To obtain a dependent measure for the corpora in direct comparison in the ANOVAs (small bounding box for Hurricane Sandy, large bounding boxes for Oklahoma tornadoes and the Boston Marathon bombing, wins and losses from the World Series cities, and the full fantasy football corpus) I also separated out the tweets for each event category that individually contained target variables (i.e., 6 events by 9 pairs of 2 individual words for 108 total files[9]). I analyzed each file with the LIWC tool. I then examined which of the 67 LIWC categories (see Pennebaker et al., 2015b, for a more detailed explanation of these variables) had the most consistent patterns between tweets containing high versus low frequency antonyms, looking especially for pairs that moved in one direction for disaster corpora and the other for non-disaster. Adding to this list those LIWC variables for which I had *a priori* hypotheses (e.g., negation, positive and negative emotion) yielded 20 total ANOVAs.

I organized the scores on these LIWC variable for each of the 108 files into the three-way ANOVA table described above. For example, with the LIWC category negative emotion, I ran separate 108 separate LIWC tallies—one each for tweets arising from a unique combination of event and word in the antonym pair list. Each LIWC analysis gave a single score for the percentage of negative emotion words appearing in that subset of tweets. I organized these scores into the pair by event by high versus low frequency member decomposition described above. The resulting ANOVAs determine if particular pairs, particular events, or the high frequency word in an antonym pair co-occur with increased measures of content density. There are possible confounds in this analysis as the word used to segment the corpus (words in the antonym pairs) could appear in the LIWC dictionary. For example, in the case of the LIWC category

[9] "Sane" was never used in the location specific World Series losses corpus, nor was "severe" used in the fantasy football corpus, resulting in null results for those two files.

“certainty”, I eliminated all/some; “all” confounds results across pairs because it also serves a seed word for the “certainty” category. In these cases, I removed the pair in question from the analysis.

III. Results and Discussion

First, I review pilot data conducted with relatively small sample sizes and analyzed by the LIWC software to demonstrate the computationally accessible differences in language between disaster and non-disaster scenarios (RQ1). For the complete data set, I then correlate the observed proportions of lexical choice between pairs of disasters to examine generality across disaster corpora (RQ1a). Next I examine specific antonym pairs to determine which ones depart from baseline proportions consistently across corpora (RQ1b), the directionality of this departure, and the (in)dependence of these findings on sentiment. I then split the corpora by proximity to disaster epicenter and employ LIWC (RQ2) to investigate the impact of proximity to content manner (RQ2a). This includes demonstrating that my manipulation taps into social media users who are genuinely attached to the event, by the “Authenticity” metric. “Posemo” demonstrates that sentiment analysis proves more complex than a naïve interpretation might suggest. Other metrics for biological and spatiotemporal concerns indicate personal reflections concerning food, space, time, and motion that support my narrative argument that people in the event remain more concerned with practicalities than those surrounding the event. Finally, a comparison only of tweets containing individual target words (e.g., a corpus of tweets that all contain the word “wonderful” originating from the area around Hurricane Sandy) tests the relationship between lexical choice and sentiment, the hypothesized orthogonality with negation, and the notion of relative information density concurrent with shifts toward more stressed linguistic choices (RQ3). A sample of tweets determined by my lexical choice heuristic illustrates the notion of breach underlying the observed pattern of results.

Research Question 1: Does the nature of an event (including disaster context, control over outcome, and sentiment) influence linguistic behavior?

LIWC pilot. Pilot analyses established the potential differences in language behavior between disaster events and normal circumstances and provided a grounded theory approach (Corbin & Strauss, 1990) to the development of hypotheses. Consistent with the method of Danescu-Niculescu-Mizil, Gamon, and Dumais (2011), I used the LIWC software (Pennebaker et al., 2001) on small (1,000) geographically dispersed Twitter datasets collected under nominal conditions with a sample from tornado laden Oklahoma to reveal instances of collective deviation from baseline language behavior, indicative of breach of canonicity. Three datasets contain selections from tweets in one of three major cities: Los Angeles, Miami, and Manhattan. Each represents a different week with nothing of particular interest happening, certainly with respect to disaster management. The other dataset contains a selection of tweets from a section of Oklahoma during a period when it was subjected to numerous tornadoes. These datasets all contain 1,000 tweets. Qualitative analysis suggests differences in language between disaster and non-disaster settings (see Table 6).

Table 6

Preliminary LIWC metrics for disaster versus non-disasters in percentage of total tweet volume

	Region	Positive Emotion	Negative Emotion	"I"	Prepositions	Function
Non-Disaster	Los Angeles	4.71	2.49	5.66	08.37	39.25
	Miami	4.41	1.94	3.36	07.27	30.71
	Manhattan	4.79	2.13	4.71	08.35	38.19
Disaster	Central Oklahoma	2.67	3.17	1.76	13.29	34.33

Positive emotion and negative emotion both trend in the expected direction, considering the large areas in question[10]. The next column presents the personal pronoun "I" and shows a marked decrease in the disaster scenario, indicating a shift away from the egocentric perspective among the population. This counterintuitive result may stem from the established prosocial tendencies of people in disasters (Rodríguez, Trainor, & Quarantelli, 2006; Quarantelli, 1986). Again, this averages over the entire city, rather than a relatively small area that was directly impacted, which may show different patterns from more spatially restricted analyses. Prepositions appear to increase dramatically, possibly indicating a focus on spatial language, introducing new subjects with reference to established ones (e.g. "North *of* town"). Finally, function words do not differ from the general population of tweets, likely because this category collapses over several others which may cancel out (e.g. articles, conjunctions, adverbs), indicating a need to closely examine how variables are aggregated. These results support the notion that analysis of linguistic behavior can detect both reactions to disaster situations as well as assertions about them.

[10] I will later address the nuances inherent in sentiment analysis of disaster scenarios.

RQ1a: What event differences influence linguistic behavior?

Lexical choice by event. I used Spearman partial correlations to demonstrate positive relationships between the observed language behavior, generally across all events. I report the observed partial Spearman *r* values for the relationship between two event corpora, controlling for the underlying relationship of both to baseline values. The resulting partial *r* values confirm positive relationships between the observed proportions despite the influence of a common baseline (see Table 7). These tables include minor discrepancies from previously published work on the same data sets (Hampton & Shalin, 2016) based on changes in measurement method.

Table 7

Partial Spearman Rank Correlation Values for Observed Proportion Values Controlling for Baseline Influence

	Sandy	Oklahoma	Boston	FF	WSL	WSW
Oklahoma	.78	–	–	–	–	–
Boston	.37	.56	–	–	–	–
FF	.41	.49	.11	–	–	–
WSL	.54	.64	.31	.64	–	–
WSW	.39	.48	.50	.47	.78	–

Note. n = 36 for all comparisons. Approximate critical r = 0.33 for a = .05 (Noether, 1976, p. 203). "*FF*" indicates fantasy football, "*WSL*" for World Series losses, and "*WSW*" for World Series wins.

Almost all the correlations are significant. The strongest correlations do exist between theoretically related events: Hurricane Sandy and the Oklahoma tornadoes and the World Series wins (WSW) and losses (WSL). In both cases, this suggests some influence of experience on lexical choice. The high correlation between World Series wins and losses further suggests that my lexical choice metric captures something beyond sentiment analysis, as these should diverge strongly on that dimension.

With 15 pairings I hesitate to suggest substantive distinctions. The relatively low relationship between the Boston Marathon bombing and Hurricane Sandy (r = .37) could reflect substantive event differences (natural versus intentional disasters). Other relatively high correlations could reflect spurious relationships owing to the relevance of specific word pairs (e.g., “high” and “low” could both refer to weather characteristics in Hurricane Sandy and Oklahoma tornadoes).

These analyses do not consider the statistical significance of the individual metric (antonym proportion) relative to the baseline, and the potential relationship between specific pairs and events. This concern is addressed further below.

Research Question 1b: How does event influence linguistic behavior: proportion of paired antonyms in situations of breach compared to an Internet-specific base rate—is the departure general or specific to antonym pairs?

Examination of specific antonym pairs. Spearman values are a function of the entire dataset; interpreting significant correlations is not straightforward. Below I examine the specific pairs that are responsible for the observed correlations, focusing particularly on those pairs that seem to be consistent across disasters. I also examine the potential confounding of these patterns with sentiment, and illustrate some of the content associated with high frequency antonyms.

Pair effects. Using an effect size metric to examine the discrepancy of observed proportion from the baseline, I present my results regarding specific antonym pairs in two tables. The first (Table 8) focuses on those d values that exceed an (arbitrary) absolute value of 0.37 with 95% confidence intervals that do not include zero for at least two disasters. The direction of the discrepancy is roughly evenly split between increases and decreases in the prevalence of the rarer term. Nine pairs are consistent across all three disaster events. This presents an increase of two pairs (soft/hard and stupid/smart) from findings reported in a previous publication on the same data (Hampton & Shalin, 2016) based on the change in word count methods. Significant variations may occur if word selection is correlated with punctuation, captured by the more computationally advanced tally employed here.

Only two word pairs diverge from word baselines in opposing directions between disaster events (warm/cool and east/west). Of the nine pairs that show a consistent pattern across disaster events, none are consistent in the same direction in *all* non-disaster events. Three are significant in the same direction in more than one non-disaster event (wonderful/horrible, crazy/sane, and together/alone) consistent with the above noted correlations. Stop/start, minor/severe, and hard/soft shift in the opposite direction when comparing all three disasters to one of the non-disaster datasets, and fall either below my threshold or below significance thresholds in the other two non-disaster datasets.

Table 8

Moderate to Large Effect Size Departures from Baselines by Disaster

Word Pair	Hurricane Sandy *d*	Oklahoma Tornado *d*	Boston Marathon Bombing *d*	Fantasy Football *d*	World Series Wins *d*	World Series Losses *d*
Horrible/ Wonderful*	1.10	0.67	1.38	1.70	0.42	*NS*
Stop/ Start	0.79	0.43	0.80	**-0.75**	*BT*	*BT*
Severe/ Minor	0.48	1.91	1.09	*ZU*	*NS*	**-0.89**
Warm/ Cool	**0.42**	-0.94	-1.12	-0.79	*NS*	-0.46
Some/ All	-0.38	-0.39	-0.60	*BT*	*BT*	*BT*
Sane/ Crazy*†	-0.71	-1.24	-1.45	-1.21	*ZU*	-1.37
Soft/ Hard†	-0.89	-1.06	-0.47	**0.75**	*NS*	*NS*
Smart/ Stupid*	-0.95	-0.53	-0.63	-0.44	*NS*	*NS*
Alone/ Together	-1.21	-1.12	-0.45	*BT*	-3.35	-2.94
Tiny/ Massive	-1.26	-1.43	-1.44	*NS*	*NS*	-0.58
Worse/ Better*	0.70	0.45	*NS*	*NS*	**-0.47**	**-0.56**
Under/ Over	-0.59	-0.96	*BT*	-0.62	-0.92	-0.85
East/ West	**1.03**	*NS*	-1.03	*NS*	-0.67	-0.88
Fake/ Real*	0.88	*NS*	1.10	*NS*	*NS*	*BT*
Terrible/ Great*	0.54	*BT*	0.60	0.60	*NS*	*BT*
Large/ Small	-0.73	*NS*	1.06	*BT*	*NS*	*NS*
Unsafe/ Safe*	-1.52	*ZU*	-1.03	**0.60**	*NS*	-1.16
Global/ Local	*BT*	-0.83	-1.33	*NS*	*NS*	-1.50

Note. Less frequent words according to GloWbe appear first in the pair description, so that positive *d* indicates an observed increase in the less frequent word and negative *d* indicates an observed increase in the more frequent word. Bolded pairings indicate a reversal of direction across disasters. “NS” indicates that a *d* 95% confidence interval contained 0. “BT” indicates that the *d* value fell below my threshold of an absolute value of 0.37. “ZU” indicates zero instances of the uncommon alternative. Less frequent words according to GloWbe appear first in the pair description, so that positive *d* indicates an observed increase in the less frequent word and negative *d* indicates an observed increase

in the more frequent word. Horizontal lines separate pairs that were significant in all three disaster events, then two out of three without the Boston Marathon bombing, the Oklahoma tornadoes, and Hurricane Sandy, respectively. Pairings with results unique to Hurricane Sandy are not presented.
* indicates a sentiment asymmetry per SentiStrength
† indicates disagreement between Warriner sentiment classification and SentiStrength

Table 9

Hurricane Sandy d *Analysis for Pairs Not Included in Table 8*

Significant *d* above threshold	*d*	Significant *d* below threshold	*d*	Non-significant *d*
Out/In	0.51	Bad/Good*	0.37	Every/Any
Last/First	0.47	Slow/Fast	0.34	Imperfect/Perfect*
Black/White	0.38	Left/Right	0.29	Shorter/Longer
Large/Small	-0.72	Big/Little	0.17	Stale/Fresh*
Dull/Amusing*	-1.26	Down/Up	-0.09	
		Few/Much	-0.13	
		Dead/Live*	-0.23	
		Low/High* †	-0.25	
		Boring/Fun*	-0.27	

Note. Less frequent words according to GloWbe appear first in the pair description, so that positive *d* indicates an observed increase in the less frequent word and negative *d* indicates an observed increase in the more frequent word.
* indicates a sentiment asymmetry per SentiStrength
† indicates disagreement between Warriner sentiment classification and SentiStrength

The second table completes the list of 36 pairs, illustrating singleton effects for Hurricane Sandy, where the large corpus provides narrow confidence intervals. For Hurricane Sandy, most of the word pairs (24/36) meet my discrepancy criteria. Annotations on the entries in Tables 8 and 9 indicate affective (sentiment) asymmetry in word pairings. Many pairs do not reflect affective asymmetry. Some pairs such as minor/severe, massive/tiny and in/out have clear context relevance. But neither affective asymmetry nor context relevance explain pairs such as some/all, stop/start and alone/together. The pair some/all is particularly interesting because it includes two high frequency “stop words” that are generally ignored in text mining.

Research Question 2: What impact does spatial proximity to a disaster have on language behavior (e.g., authenticity, sentiment, and negation)?

I explored a range of LIWC variables to find those that vary consistently with increased distance from an event epicenter. Figures 10–12 illustrate four of these metrics for the three events, with gradated distinctions for distance (central, doughnut, and distal). To aid comparison, I converted observed words-per-hundred values (standard LIWC output) to z-scores based on my three samples. High *authentic* scores indicate a more honest, personal, and disclosing text (note that this does not incorporate personal pronouns); lower scores suggest a more guarded, distanced text. *Ingest* refers to kinds of food and ingestion terminology such as “taste” and “dine”. *Posemo* refers to positively valenced emotional words and word combinations. *Relativ* combines spatial, temporal, and motion references.

Across all three disaster events, authenticity increases with proximity to the epicenter. This pattern supports the claim that my corpora are tapping personal comments and not organizational reports, suggesting that messages do in fact come from those individuals most directly impacted by the event. It appears that those closer to the event will be more authentic by virtual of direct rather than reflected exposure, while also shifting away from an egocentric perspective (as indicated the LIWC pilot).

The positive emotion analysis further supports the claim that my findings provide value added beyond conventional sentiment analysis. The trend is for *positive* emotion to *increase* with proximity to the event, contrary to expectation. The seemingly contradictory findings in the LIWC pilot arise from the aggregated nature of that corpus, including those directly impacted along with the wider geographic area. It seems positive emotion decreases overall, but to a lesser degree for those directly impacted. This reinforces concern for reliance on sentiment as a metric of need, and also the orthogonality of my analysis to those of sentiment analysis. Finally, measures for relative words and ingestion words increase with proximity to the disaster epicenter. The former indicates that people are in motion or monitoring dynamic events in specific terms. The latter suggests a concern for the basic necessity of food and water. Both provide convergent validity for my identification of breach as the primary driver of changes in lexical choice, with no suggestion that the use of negation systematically differs in its presence.

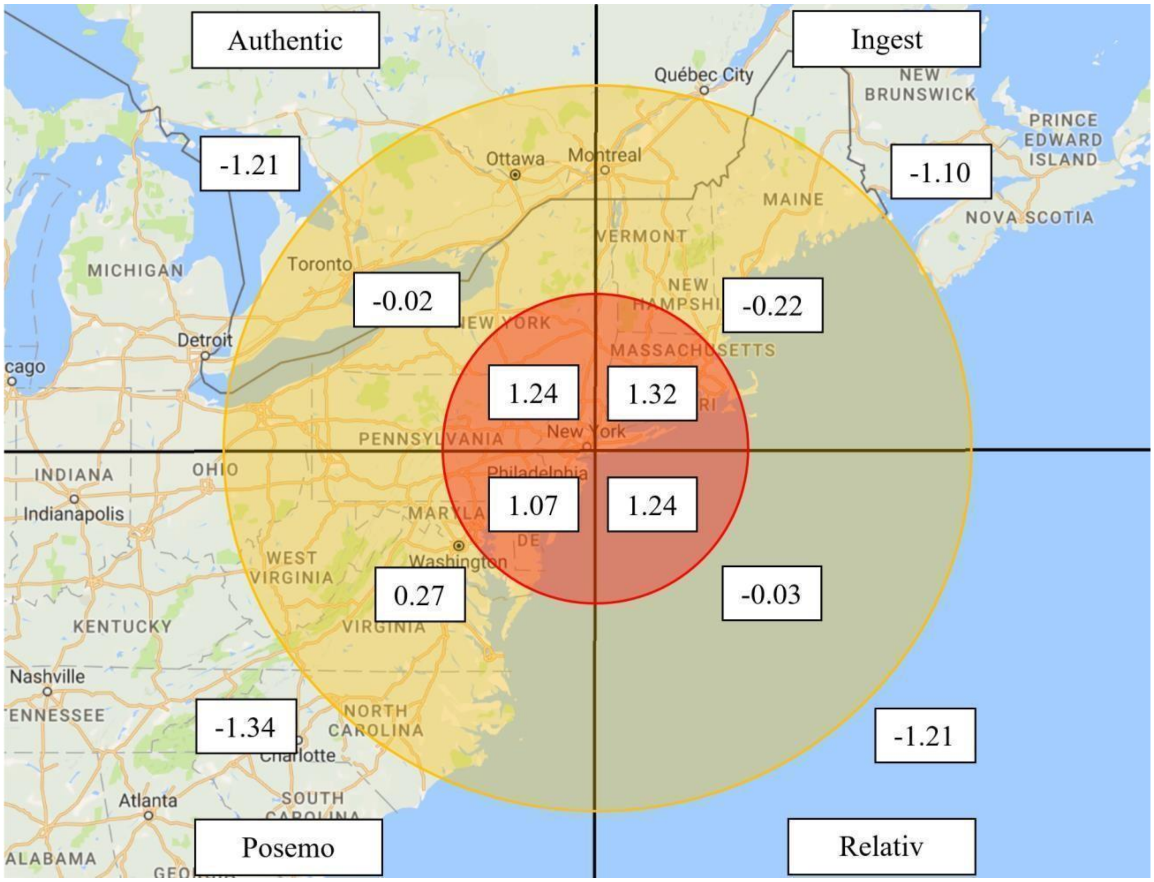

Figure 10. Spatial representation of LIWC categories in z-scores relative to the hardest hit areas of Hurricane Sandy. Map image credit: Google Inc.

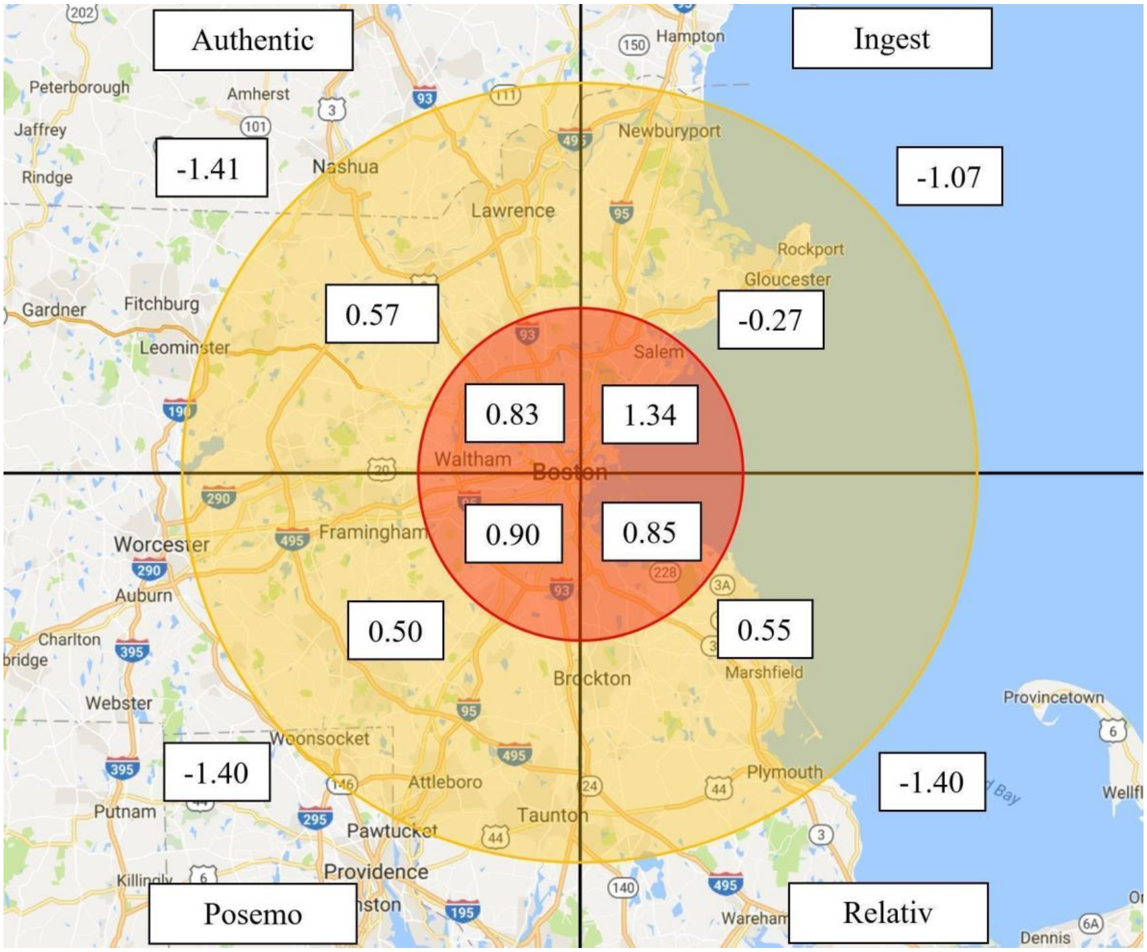

Figure 11. Spatial representation of LIWC categories in z-scores relative to the Boston Marathon finish line. Map image credit: Google Inc.

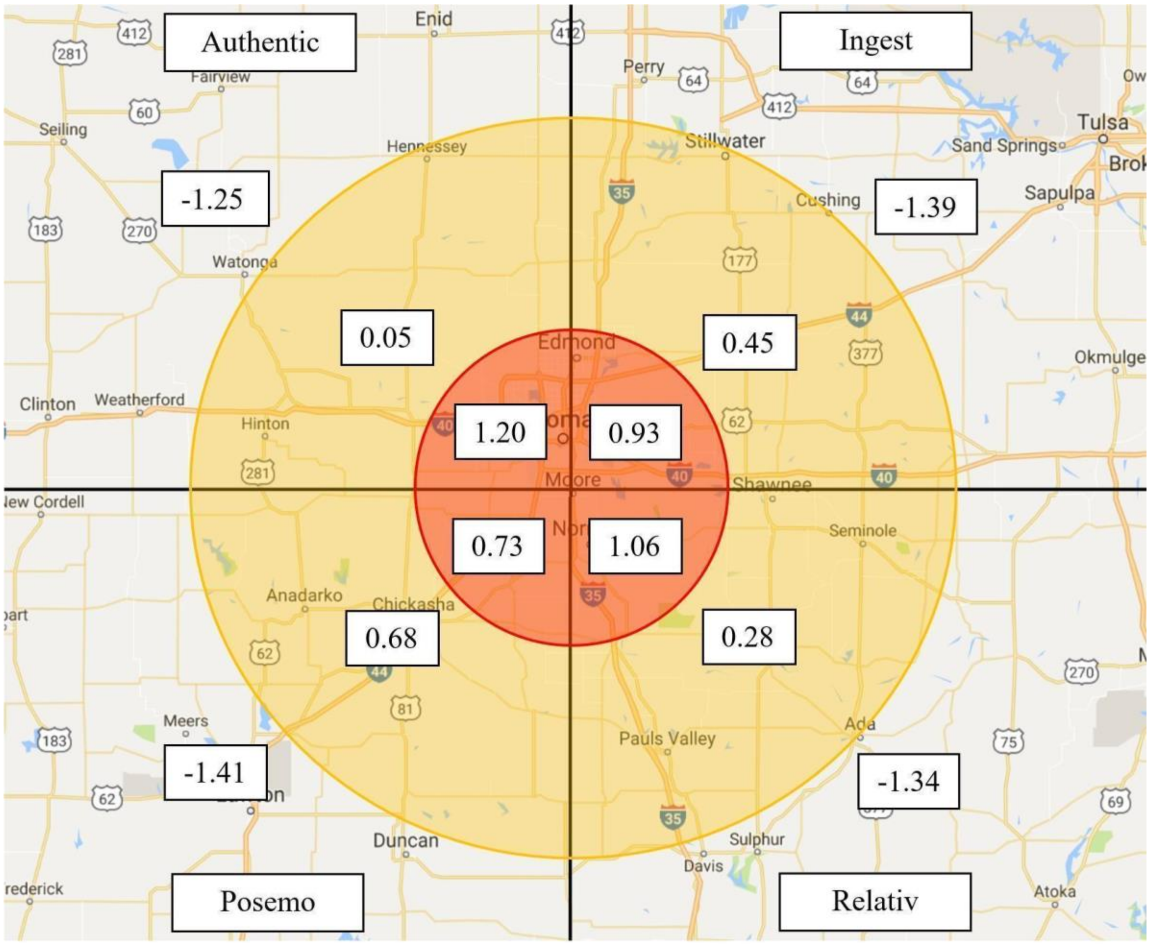

Figure 12. Spatial representation of LIWC categories in z-scores relative to hardest hit areas of the storms in central Oklahoma. Map image credit: Google Inc.

Research Question 3: Do high frequency members of the antonym pair correlate with content (e.g., sentiment, negation, motion processes)?

This question allows me to investigate three things. First, I can determine whether the use of words that increase in use in disaster contexts can be exploited to mine actionable content. Second, it examines the effect of negation with respect to my antonym pairs and if these two aspects of lexical choice are related. Finally, I can address the relationship between antonym selection and sentiment, the dominant method of large-scale social media analysis.

I used ANOVA to determine the presence or absence of a quantitative relationships in these areas, as measured by LIWC constructs. I conducted three-way ANOVAs, with *Pair* (nine), *Event* (six) and *Hi/Lo Frequency* (two) as independent variables. I considered *Pair* as a random effect, and *Event* and *Frequency* as fixed effects. Proper tests of *Event* and *Frequency* were therefore against *Pair**Fixed Effect* error terms. Due to the small number of pairs, these were conservative, low-power tests. A less conservative test uses the pooled residual from all *Pair* interactions. I also decomposed the five *Event* degrees of freedom into separate orthogonal contrasts.

I am less interested in differences among the nine pairs, as these would be difficult to disentangle and of questionable utility or generalizability to other circumstances. Instead, I focus on differences between the high and low frequency alternatives with respect to disaster datasets (i.e., those that increase or decrease in relative use in the presence of disasters). Differences between disaster and non-disaster datasets, or among the various constituent datasets could potentially provide useful insight, particularly if these interacted with high and low frequency antonym selection.

Tweet content. Other LIWC measures could support the claim that high frequency members of the antonym pair are associated with a greater content density. If antonym pairs are to function as screening tools, it may follow that messages containing words that consistently increase in relative use during disasters would also contain more disaster-relevant information. To test this, I ran an analysis of variance (ANOVA) on 20 LIWC categories. An example analysis in Table 10 shows that this is largely not the case.

Table 10

ANOVA for Prevalence of Motion Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	20.53	8	2.57	2.14	0.04
Event	26.87	5	5.37	4.79	0.00
Non-Disaster vs Disaster	2.90	1	2.90	2.58	0.12
Football vs. Baseball	22.05	1	22.05	19.65	0.00
Manmade vs. Natural Disaster	0.38	1	0.38	0.34	0.56
Baseball Win vs. Loss	1.31	1	1.31	1.16	0.29
Oklahoma vs. Sandy	0.24	1	0.24	0.21	0.65
High/Low	0.74	1	0.74	0.73	0.42
Event*High/Low	7.57	5	1.51	1.16	0.35
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.53	1	0.53	0.41	0.53
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	0.39	1	0.39	0.29	0.59
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	1.73	1	1.73	1.32	0.26
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	4.78	1	4.78	3.65	0.06
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.15	1	0.15	0.11	0.74
Residual	105.36	88	1.20		
Event*Pair	44.89	40	1.12		
High/Low*Pair	8.07	8	1.01		
Event*High/Low*Pair	52.40	40	1.31		
Total	161.08	107			

Several categories overlapped with one or more of the target words and had to be re-run with that pair removed. (For full ANOVA result tables, see Appendix entries 1-19.) Despite the large corpora, segmentation on both word pair and semantic categories resulted in several categories with no representative tweets for analysis. To combat the unequal *n* that resulted, I either deleted the subject (word pair) if more than one score was missing, or replaced the score using Kirk's formula for individual null results (1982, p. 268-270).

Despite an exhaustive, multi-faceted deconstruction of the events, pairs, and contrasts, my conservative error term prevented discovery of any consistent departure toward increased content density in the presence of words that increase in frequency in the presence of disasters. Using a more liberal test with pooled degrees of freedom, I did find that LIWC measures for word count demonstrated significant departures. This indicates that when people elect the stressed alternative, their messages tend to be longer. The messages also use more relative words, concerning spatiotemporal relationships.

The relatively weak signal, combined with the risk of family-wise error, serves as a cautionary warning. Such an approach to the identification of content requires substantially more pairs, which is the limiting factor on power in the error term.

However, manually-selected examples of tweet content illustrate breach as a more promising heuristic for content identification. All the examples in Table 11 illustrate the disruption in normal activity—notable but not uniformly negative. The mix of sentiment reflects the range of communicative function, including commissives, directives, and beliefs, along with factual assertion.

The Boston police in Example 5 were surely aware of their presence at the train stations; the tweet does, however, indicate the public response. And while Example 7 does not require an organizational response, it does inform the response organizations of community activity, which can be highly influential in distributing resources. I note the wide-ranging idiosyncratic content and language apart from my antonym-pair heuristic, indicating power outages, downed trees, and disrupted traffic. My stylistic heuristic indicators support the identification of numerous specific compromises, phrased in virtually unlimited fashion.

Table 11

Tweet Examples

	Word pairs	Tweet example ***[SentiStrength Rating]***
1	Severe/ Minor	At work wanting to go home! ***[2]*** I've busted my butt *all* week and worked every day of the hurricane. ***[1]*** I'm tired and in *severe* pain! ***[-5]***
2	Stop/ Start	If you are driving through Moore on I-35, *stop* pausing to look at the wreckage. ***[-2]*** It's making traffic a problem ***[-2]***
3	Horrible*/ Wonderful	@Drew_Hampton *horrible* I still don't have any power from hurricane sandy & I'm freezing :(***[-4]***
4	Some/ All	Yes, and *all* flights to Boston are totally full due to the bombing ***[-2]***
5	Sane†/ Crazy	Wow there's a lot more security and police at the train stations now. ***[3]*** This is *crazy*?? #BostonBombing ***[-2]***
6	Tiny/ Massive	Storm knocked down one of the *massive* trees in front of my #house ***[-2]*** #rip #sandy #hurricane #HurricaneSandy ***[NA]***
7	Alone/ Together	Getting a team *together* to go up near Moore to cut tree limbs. ***[1]*** Call me if you're interested #Oklahoma #Tornado #Relief ***[2]***
8	Soft/ Hard†	$2B in Okla. tornado damage means *hard* recovery: ***[1]*** *All* that is left of my friend's three-bedroom home is this http://t.co/dAgDFiA ***[2]***
9	Smart/ Stupid*	During the moment of silence there was a bomb threat called into a Boston school. ***[-3]*** Probably a false alarm from a *stupid* kid! ***[-4]***

Note. The SentiStrength scale ranges from -5 to 5. I have superficially altered tweet examples in compliance with Twitter's privacy policy. These examples illustrate a mix of assertions (3), directives (2, 7) and commissives (7) that influence overall sentiment ratings.
* indicates affectively asymmetric pairs according to SentiStrength
† indicates a disagreement on affective asymmetry between Warriner et al. and SentiStrength

While the presence of a significant effect for positive emotion words could indicate an overlap with sentiment analysis, the spatial analysis and the ANOVAs discount such a straightforward conclusion.

Negation. Antonym analysis does not account for the use of negation, e.g., "not tiny", "not wonderful". Negation provides a possible mechanism for articulating thought comparable to the prevalent antonym but drawing attention by using the less prevalent alternative. This would manifest in an increased use of negation in the low-frequency alternative, possibly depending on an interaction with event that warranted increased emphasis. Alternatively, observers could attempt to de-escalate anxieties in a disaster by negating the emotionally valenced alternatives— “not massive” or “not horrible”. This possibility would manifest in a pair by event interaction. Combining both possibilities suggests that the function of negation may be largely orthogonal to the above outlined influences of lexical choice. The ANOVA on the prevalence of negation showed little in the way of interpretable significant variation with respect to proposed interactions (see Table 12). While differences among pairs were significant ($F[8, 88] = 2.18$, $p = 0.04$) as a main effect (i.e., some pairs co-occur with negation more than others), that finding that is largely irrelevant to my argument in the absence of high versus low frequency or event type interactions. Likewise, disaster versus non-disaster datasets differed ($F[1, 40] = 5.87$, $p = 0.02$, $\mu_{\text{Disaster}} = 0.71$, $\mu_{\text{Non-disaster}} = 0.53$) but without relation to the pairs or their baseline frequency.

Table 12

ANOVA for Prevalence of Negation Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	3.95	8	0.49	2.18	0.04
Event	1.96	5	0.39	2.42	0.04
Non-Disaster vs Disaster	0.50	1	0.95	5.87	0.02
Football vs. Baseball	0.02	1	0.31	1.91	0.17
Manmade vs. Natural Disaster	5.67	1	0.55	3.39	0.07
Baseball Win vs. Loss	0.09	1	0.04	0.26	0.61
Oklahoma vs. Sandy	4.51	1	0.11	0.66	0.42
High/Low (Conservative)	0.07	1	0.07	0.10	0.76
High/Low (Liberal)	0.07	1	0.07	0.29	0.59
Event*High/Low	0.78	5	0.16	0.82	0.54
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.04	1	0.04	0.23	0.63
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	0.24	1	0.24	1.29	0.26
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.09	1	0.09	0.49	0.49
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.05	1	0.05	0.25	0.62
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.35	1	0.35	1.84	0.18
Residual	19.94	82	0.24		
Event*Pair	6.48	40	0.16		
High/Low*Pair	5.88	8	0.74		
Event*High/Low*Pair	7.57	40	0.19		
Total	14.06	107			

Sentiment. The pairs that appear to be diagnostic could simply reflect trends in sentiment. If so, the high frequency member of the antonym should elicit sentiment content, either positive emotion, negative emotion, or changes in tone. One complication of this analysis is that the antonym pairs themselves contribute to sentiment. As I have already dismissed the correlation of high-frequency members of the pair with sentiment, in the following analyses I remove the confounding pair (horrible/wonderful for both positive and negative emotion, and together/alone for the inclusion of “alone” in the negative emotion category). This allows me to assess whether additional tweet sentiment is related to the high frequency choice. Unfortunately, this further reduces power.

Table 13

ANOVA for Prevalence of Positive Emotion Words with Confound Removed

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	39.20	7	5.60	2.36	0.03
Event	10.52	5	2.10	1.28	0.30
Non-Disaster vs Disaster	0.94	1	0.94	0.57	0.46
Football vs. Baseball	0.04	1	0.04	0.02	0.88
Manmade vs. Natural Disaster	5.17	1	5.17	3.13	0.09
Baseball Win vs. Loss	0.06	1	0.06	0.04	0.85
Oklahoma vs. Sandy	4.31	1	4.31	2.62	0.11
High/Low (Conservative)	20.02	1	20.02	2.00	0.20
High/Low (Liberal)	20.02	1	20.02	8.43	0.00
Event*High/Low	3.95	5	0.79	0.50	0.77
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.35	1	0.35	0.22	0.64
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.28	1	1.28	0.81	0.37
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.02	1	0.02	0.01	0.91
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	1.12	1	1.12	0.71	0.41
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	1.18	1	1.18	0.75	0.39
Residual	182.92	77	2.38		
Event*Pair	57.73	35	1.65		
High/Low*Pair	70.05	7	10.01		
Event*High/Low*Pair	55.14	35	1.58		
Total	256.61	95			

Table 14

ANOVA for Prevalence of Negative Emotion Words with Confounds Removed

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	194.96	6	32.49	6.97	0.00
Event	44.97	5	8.99	8.68	0.00
Non-Disaster vs Disaster	29.23	1	29.23	28.20	0.00
Football vs. Baseball	0.09	1	0.09	0.08	0.77
Manmade vs. Natural Disaster	12.32	1	12.32	11.89	0.00
Baseball Win vs. Loss	3.20	1	3.20	3.08	0.08
Oklahoma vs. Sandy	0.14	1	0.14	0.14	0.71
High/Low (Conservative)	115.90	1	115.90	3.91	0.05
High/Low (Liberal)	115.90	1	115.90	24.85	0.00
Event*High/Low	13.54	5	2.71	0.82	0.54
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	2.26	1	2.26	0.81	0.38
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	3.72	1	3.72	1.33	0.26
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.03	1	0.03	0.01	0.92
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.58	1	0.58	0.21	0.65
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.11	1	0.11	0.04	0.85
Residual	307.84	66	4.66		
Event*Pair	31.10	30	1.04		
High/Low*Pair	177.96	6	29.66		
Event*High/Low*Pair	98.78	30	3.29		
Total	677.22	83	8.16		

Table 15

ANOVA for Tone

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	13714.21	8	1714.28	1.74	0.10
Event	6737.98	5	1347.60	4.32	0.00
Non-Disaster vs Disaster	1923.39	1	1923.39	6.17	0.02
Football vs. Baseball	987.12	1	987.12	3.17	0.08
Manmade vs. Natural Disaster	2945.44	1	2945.44	9.44	0.00
Baseball Win vs. Loss	62.07	1	62.07	0.20	0.66
Oklahoma vs. Sandy	819.96	1	819.96	2.63	0.11
High/Low (Conservative)	20079.45	1	20079.45	2.71	0.14
High/Low (Liberal)	20079.45	1	20079.45	20.37	0.00
Event*High/Low	1265.81	5	253.16	0.68	0.64
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	283.01	1	283.01	0.76	0.39
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	46.90	1	46.90	0.13	0.72
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.17	1	0.17	0.00	0.98
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	432.85	1	432.85	1.16	0.29
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	502.88	1	502.88	1.35	0.25
Residual	86739.84	88	985.68		
Event*Pair	12474.86	40	311.87		
High/Low*Pair	59322.13	8	7415.27		
Event*High/Low*Pair	14942.84	40	373.57		
Total	128537.29	107			

While positive emotion (Table 13) shows almost no significant effects, with the exception of loading on different antonym pairs and high versus low use pairs when using a more liberal criterion[11]. Negative emotion (Table 14) and tone (Table 15) prove more complex. In Table 14, the influence of event demonstrates a clear impact on the expression of negative emotion (F[5,30] = 8.68, p < 0.00). And while the decomposition

[11] More on this in a moment.

of that effect indicates that a difference arises, as one might expect, between disaster and non-disaster contexts ($F[1, 30] = 28.20$, $p = 0.00$, $\mu_{\text{Disaster}} = 3.84$, $\mu_{\text{Non-disaster}} = 2.66$)[12], but also in differentiating among types of disaster. The significant difference between manmade and natural disasters ($F[1, 30] = 11.89$, $p = 0.00$, $\mu_{\text{Manmade}} = 4.61$, $\mu_{\text{Natural}} = 3.46$) suggests that there exists nuance in the expression of response to disasters that may call into question the context-independent utility of sentiment analysis.

Further, the nearly significant effect of high versus low use members of the pairs ($F[1, 6] = 3.91$, $p > .05$ when using a conservative error term) suggests that there may exist a difference between expression of negative sentiment in tweets that contain antonym alternatives more commonly used in proximity to disasters. Closer inspection reveals that negative emotion increases when used in conjunction with the members of pairs that increase in relative popularity in disaster settings. Complementing this finding, positive emotion shows a relative depression when used in messages with the same words ($F[1, 77] = 8.43$, $p < .00$, when using the liberal error term). While this at first seems compelling, the confused results of the decomposition in negative emotion and the lack of significant event findings in positive emotion suggest that matched antonyms may provide a correlated, but more flexible tool than either positive or negative sentiment.

Table 15 further complicates the detectability of sentiment during community response to disaster. Here the effect of events ($F[5, 40] = 4.32$, $p < .00$) suggests that differences certainly exist in tone, where “a high number is associated with a more positive, upbeat style; a small number reveals greater anxiety, sadness, or hostility” (Pennebaker, Booth, Boyd, & Francis, 2015, p. 22). While this definition seems to adhere

[12] Where I find theoretically interesting significant differences, I include cell means.

closely to the tenets of sentiment analysis, the decomposition of event influences only serves to confuse an intuitive reading. While disaster versus non-disaster events shows a strong effect ($F[1, 40] = 6.17$, $p = .02$), the strongest effect appears for man-made versus natural disasters ($F[1, 40] = 9.44$, $p < .00$). The reversed order of these (relative to the expected direction of sentiment in disasters generally) may allow speculation that minor difference between findings could be due to sampling or idiosyncratic anomalies, but that same rationale would necessarily question why fantasy football versus baseball ($F[1, 40] = 3.17$, $p = .08$) and Hurricane Sandy versus Oklahoma tornadoes ($F[1, 40] = 2.63$, $p = .11$) trailed so closely behind, without an apparent difference in sentiment content. In fact, the only events that demonstrated no detectable difference in tone were wins versus losses in the World Series ($F[1, 40] = 0.20$, $p = .66$), which I included explicitly to differentiate sentiment from event influences.

Consideration of tone used in conjunction with words that appear more commonly in disaster scenarios versus those that decrease in relative frequency (high vs. low) shows a significant difference when using the more liberal pooled error term ($F[1, 88] = 20.37$, $p < .00$). As one may expect, the pattern follows that expressed in positive and negative emotion, with lower tone scores (more anxious, sad, and hostile) coinciding with words that increase in disaster settings. But again, given the mixed results of the event decomposition and combined with the nuance illustrated by the epicenter distance analysis (Figures 10–12), I would find it difficult to detect stable patterns of sentiment across disasters using conventional methodology.

IV. General Discussion

By examining lexical choice across variety of situations, I identify trends suggestive of disruption to normalcy, which I call breach. To reveal disruption, I changed the way linguistic data are typically analyzed, from a frequency-based approach to a baseline referenced approach. Accordingly, my measure is the departure of observed antonym proportions in disasters relative to the same proportion across all instances of language behavior on the Internet. I review the implication of my findings relative to the literature that inspired my research questions. As one might expect at the beginning of a novel approach to observation, I note a number of limitations prior to identifying the major contributions and future avenues of this research.

To be sure, any consistency across disaster types in the face of geographic and cultural diversity is encouraging for the text mining effort. Two sets of results support my view that psychological functions influence the description of environmental variability. Certainly, in most cases, environmental variability provides the simplest explanation of word choice, e.g., “massive” in the case of disaster. However, an intriguing set of antonym pairs more generally related to human observation and function appears relevant to the problem of interpreting linguistic response to disaster. Second, a supporting analysis of sentiment with respect to disaster proximity suggests that human response provides an important foundation for the interpretation of social media, consistent with the enriched correspondence model that I suggested at the outset. In the following general discussion, I place my results in the context of the literature that I reviewed at the outset, addressing the influence of events, the selection of antonyms, spatial proximity and content density within antonym selection. I address limitations specific to my research questions where they arise but postpone the discussion of broader issues and future work to a closing section on contributions.

Event Influence

I found that events did influence language behavior relative to baseline standards. The strongest evidence for this is the high correlations in language choice between the Hurricane Sandy corpus and the Oklahoma tornado corpus, and between the win and loss responses to the World Series corpora. A match between language and environment is a fundamental requirement for reliance on social media as a source of information, and an oversight in much of the scholarly literature on verbal exchange between interlocutors. For example, Pickering & Garrod (2004) refer to a situation model that is not grounded in an environment, overlooking the very problem I seek to address. Clearly, the nature of the event influences language behavior (RQ1), as the triadic semiotic perspective demands (e.g. Flach & Voorhorst, 2017; Peirce, 1894; Roy, 2005).

I intentionally examined events with different qualities of stress using a common set of terminology. I included natural and man-made disasters and sports events with varying sentiment and degree of control. While sports fans surely engage in hyperbole, few view the proximal outcome as truly life or death. In addition, the separation of winning versus losing corpora theoretically follows a fault line in sentiment analysis, serving to discriminate the empirical contributions of proportional lexical choice from more traditional methods[13]. Finally, examining both city-specific (World Series) and diffuse (fantasy football) sports corpora allows for differential analyses related to the potential influence of geographic homogeneity.

Correlations in language choice relative to baseline standards confirm the importance of language analysis with respect to environmental influences. However, the

[13] However, the differing sizes of the cities and relative lengths of winning and losing streaks do introduce class imbalance for which I was unable to compensate in the current investigation.

correlations in word choice between *all* events are relatively high. Moreover, the large Hurricane Sandy corpus yields significant findings for nearly ⅔ of my measures, even using conservative effect size criteria. I cannot dismiss the possibility that social media are used to document general breach, with no sensitivity to the type or degree of breach, or the role of control over outcome. This is of course consistent with classical Gricean maxims, which suppress the articulation of the insignificant.

Nevertheless, I only examined six events permitting only fifteen event correlations. Reliable identification of what constitutes high or low correlation would require at least dozens of events. The key disaster and non-disaster events depicted here are meant to cover a range of event types and geographies, but different combinations of these variables (e.g., a hurricane in the Gulf of Mexico instead of the East coast) as well as new areas and event types (e.g., earthquakes, wildfires) should supplement them. Although all the events I examined provide an opportunity for breach, they confound several influential factors (particularly natural versus man-made disaster, and degree of control).

An additional concern is the conceptualization of event control conditions. Retroactive, location-specific data collection (e.g., central Oklahoma before the tornado hit) proves cost prohibitive (Twitter, 2017c). Also, interpretation would be suspect primarily because there likely exists no noteworthy event or location (i.e., crawlable corpus) in which nothing stressing happens. An alternative is to exploit the less-eventful—crawling for benign topics like breakfast or prospective (as yet unexperienced) events like trip planning. Creating a sufficiently diffuse and flexible sample to meet these conditions would constitute an ambitious, though clearly valuable, multi-year project. Commonly entire research grants go toward establishing how often certain populations use certain words in a particular medium under “normal” circumstances (e.g., Davies & Fuchs, 2015).

Even with a comprehensive sampling of event type, location, and relative stress, I would still be unable to demonstrate definitively that the breach characteristics I observed do not constitute a fundamental property of Twitter exchanges. All these limitations merely underscore a primary motivation for my work: the evaluation of any metric regarding human behavior requires comparison to something. Computationally inspired analyses of social media traffic do not typically employ this common feature of psychological research. Difficulty does not justify the absence of an attempt.

Antonym Selection

I examined word choice between antonyms and significant differences in their use relative to Internet-specific baselines (RQ1b). From a theoretical perspective, the idea is that word choice reflects intentionality in addition to environmental influences. The evidence here is mixed, and a primary contribution is methodological.

I employed a metric that has a baseline standard. Using proportions allowed me to adjust for differences in frequency of usage to allow for comparison across different word pairs. The psychological phenomenon of lexical marking (Gilpin, 1973) led me to study word choice in antonym pairs, based on the notion that choice between lexical alternatives reflects not just the environment but the intentional manipulation of emphasis. Among the several advantages of this approach, it aggregates over multiple contributors and thereby lessens dependence on the report of a specific individual and his or her trustworthiness.

While the evidence does suggest preference, this does not map to marking phenomena. There are several cases of increase in the more prevalent word, which is inconsistent with the idea that the less prevalent word is employed to create emphasis. No evidence suggests that negation is systematically employed to preserve a preferred word. Perhaps social media restrictions in message length made the addition of negation less attractive.

I did dismiss sentiment as the guiding principle in the indication of breach by demonstrating the absence of systematic sentiment differences in the pattern of lexical choice results and by the absence of systematic content differences that correspond to lexical choice. My most intriguing finding concerns word choice preferences that constitute generic reflections of human functionality: such as some/all, stop/start, and alone/together. These words escape the criticism of disaster specificity and have the potential to detect unknown, emergent disturbances. I argue that stop/start and alone/together reflect the consequence of breach with respect to human experience, which necessitate the intervening cognitive processes that I added in Figure 2. The findings with respect to all/some take on practical significance in text mining, as these words are often eliminated in computational analysis due to their high frequency. Instead, I employ relative metrics, with respect to both corpus prevalence and base rates. This constitutes a departure from conventional, frequency based text mining practice.

The selection of word pairs, and pairings within words, remains subjective. A good example of the resulting problem is big/little and large/small—they diverge from baselines in opposite directions. My approach does not yet escape the empirical tradition of text mining, driven both by *a priori* review of the corpora and highlighting those pairs that happen to work for the corpora at hand. The specificity of my word pairs to the disaster setting also requires consideration. Many of the diagnostic words have clear relevance to event specifics (e.g., “severe” is often linked to weather disasters, terrorism naturally engenders the use of “crazy”, both have objectively “horrible” aspects). Though useful in the detection of public perception, they require a priori knowledge of the event in order to tally. However, an important motivation for my approach is domain-independence that does not require a model of the disaster to flag relevant information. Potentially, any two words could serve this purpose. Replication and theory should guide future selection of diagnostic pairs and their direction in order to restrict the candidate set to a manageable expansion and limit exposure to both misses and false alarms. I do not require all pairs of antonyms to be useful, but rather that some subset proves consistently, and ideally *a priori*, diagnostic.

Spatial Proximity

Distance from a critical event constitutes a natural manipulation of context, leveraging the location tagging capacity of Twitter to measure distance and allow comparison of user-generated language most likely to be stressed against others within the same event that are less stressed (RQ2). To examine the effect of spatial proximity, low base rates required that I change metrics from specific words to the general conceptual categories that LIWC provides. In so doing, I adopt a knowledge-based approach to text-mining (Purohit et al., 2013; Sheth & Thirunarayan, 2012), which aggregates over multiple lexical items that refer to the same conceptual category. I infer gross differences by defining and comparing relative *zones*. These boundaries are represented in a “bulls-eye” graphic scaled with z-scores to reveal trends across conceptual categories with different base rates.

Proximity to the epicenter of disaster scenarios does appear to impact language behavior for at least some LIWC metrics (RQ2). However, the direction of the effect departs from the expectations of sentiment analysis. A counterintuitive direction of change for positive emotion language within a given event highlights the limitations of sentiment analysis. Increased positivity close to the disaster epicenter paired with the decrease in personal pronouns demonstrated in the LIWC pilot converges with Quarantelli's admonition not to assume the worst of those in disaster scenarios (1986). More relevant to the role of human language as a *response* the environment as opposed to absolute environmental conditions, positive sentiment here cannot mean the absence of breach. Language behavior tells us about human interpretation and response at least as much as language behavior reflects environmental reality. While sentiment analysis has certainly proven valuable in determining public perception of products and politicians (e.g., Liu, Huang, An, & Yu, 2007; Mishne & Glance, 2006; Tumasjan, Sprenger, Sander, & Welpe, 2010; Wang, Can, Kazemzadeh, & Bar, 2012), the direct mapping to the environment is not straightforward.

The empirically ideal event would present a defined epicenter and steadily decreasing severity scaled directly to distance, such as an earthquake. Earthquake, though fortunately absent within the United States recently, provides a well-defined epicenter. However, interpretation is still challenging, because the consequence to human experience reflects more than the epicenter, for example due to location of population centers and more susceptible infrastructure and demographics. These correspond with breach of canonicity, dependent upon cultural variables that define the collective narrative. Such scaling lies outside the scope of the current inquiry.

Content Density

A truly useful measure of public response would help to separate out messages with content (RQ3) from the distracting "thoughts and prayers"-type that hampers the pragmatic utility of social media. I investigated this by separating messages with high frequency choices from those with low frequency choices, to separately characterize the content in each. Using previously defined aggregated LIWC metrics, I found little quantitative evidence for any content differences between high frequency and low frequency antonyms. Positive and negative sentiment, already dismissed above as not directly interpretable, do not vary with high frequency terms, or even winning and losing World Series corpora.

However, the quantitative content analyses were low power, limited by the number of antonym pairs submitted to analysis and a treatment of these pairs as a random effect. Power can be addressed with more pairs, but the inclusion of additional antonym pairs would necessitate a multiplicative increase in manually created data sets. It may also be the case that content density differs between diagnostic and nondiagnostic pairs, as the latter were not included in my quantitative analyses. An intriguing possibility is that the power of my analysis for the identification of content lies not in the mining of content within the high frequency member of the pair, but rather informative dimensions of human experience e.g., size and social engagement. Relative departure from baselines could still provide the sentinel, but content mining would exploit both members of an antonym pair.

Manual examination of tweet content is somewhat more encouraging for the identification of relevant content. My quantitative analysis of tweet content exploited generic pre-established LIWC categories. However, the hand-selected tweet examples (Table 11), address issues of transportation and environmental consequence. For example, the creation of dictionaries pertaining to infrastructure damage, means of transportation, or the formal emergency response community would likely be more sensitive to changes in ground truth during a disaster scenario. A formal knowledge-driven model tied to a disaster specific dictionary will likely prove more helpful to the identification of actionable content.

General Limitations

Generic concerns related to the platform and scope of my inquiry require consideration. Mediated communication differs from conventionally-studied interactive research paradigms, and the Twitter platform has its own eccentricities. Location-tagging limitations as well as concerns for how to appropriately aggregate and interpret results elude definitive solutions. Finally, the restricted scope of my analyses calls into question the generalizability of my findings to other cultures.

Mediated communication concerns. Several eccentricities of the Twitter platform influence the relevance of my results to general issues of human communication processes: the availability of message production data but not message comprehension data; the broader function of Twitter; and limitations in both the availability of data and its potential contamination.

Text-based communication interferes with monitoring recipient comprehension (Clark & Schaefer, 1989; Clark & Krych, 2004) and likely alters communication behavior. The absence of a baseline corpus restricted to Twitter usage raises the possibility that the observed patterns reflect more general discrepancies between Twitter mediated communication and other functions of Internet language. I partially addressed this concern by demonstrating somewhat attenuated effects between disaster and non-disaster corpora (e.g., Boston with Fantasy Football and WSL, and Sandy with WSW). However, as noted above in event selection, this pattern was not at all definitive. Finally, the disasters covered all took place years before the analysis and I cannot rule out the emergence of pertinent new language trends on the social media platform. Notably, Twitter expanded their maximum message size from 140 to 280 characters in 2017 (Perez, 2017).

Identifying location. The standard Twitter API provides a limited sample of the full content stream (Twitter, 2017c). An even smaller portion of the stream employs voluntarily message location tagging. The net result is reduced sample size and low power in the *d* metric, reducing the number of diagnostic pairs and prohibiting the examination of my proportion metric over segmented space and time. This problem will yield to a larger corpus of location tagged messages, necessary not only for research purposes but also to deploy a real time metric. Several techniques exist for estimating location with varying degrees of success, accuracy, and data loss. Many rely on user profiles or analysis of network connections to determine a home location. Newer approaches to the problem include the analysis of the text itself to identify location (Jurgens, Finethy, McCorriston, Xu, & Ruths, 2015; Al-Olimat, Thirunarayan, Shalin, & Sheth, 2017).

Institutional reports (e.g., from the Red Cross or local hospitals) need to be screened out if a lexical choice metric is meant to identify public experience. Eliminating retweets, as I did in my analysis, and the LIWC authenticity metric somewhat mitigates this concern, but a complete scrubbing would require a comprehensive list of relevant agencies' and news outlets' Twitter handles.

Modeling issues. The number of variables that I studied, coupled with the overwhelming statistical power inherent in social media, suggests that I will find numerous spurious significant departures and correlations which complicate inference about the environment. For example, I can obtain high correlations of some unimportant terminology with air pressure, simply because social media usage is correlated with time of day, which is correlated with air pressure.

Working with patterns of behavior as opposed to individual assertions requires aggregation across both spatial and temporal dimensions. These relatively unprincipled aggregations could very well affect my conclusions regarding sentiment, which did not partition the corpora by time. In addition, the potential for lag complicates temporal mapping, with no standard time interval between onset of an event and reporting on social media. Furthermore, style matching (Niederhoffer & Pennebaker, 2002) and alignment (Pickering & Garrod, 2013) suggest a cascading influence of initial posts on subsequent patterns of language usage. My primarily cognitive approach to lexical choice overlooks social influences that potentially blur interpretation. Because a recipient is more likely to reproduce the term or adjective that someone else just used, future attempts should assess this effect on word choice in social media. Modeling in the spirit of multiple regression is likely required to tease out the multiple additive, if not interactive, influences on lexical choice.

Sociocultural scope. This dissertation addresses only incidents occurring in the United States. Aside from the obvious difficulties in language, syntax, and spelling differences (even with other countries that primarily use English), cultural differences threaten broad generalizability. For example, personal control, a collective mindset, or even the tolerance of breach could well be a cultural phenomenon (Nisbett, Peng, Choi, & Norenzayan, 2001) Also, more subtle differences in conventions of interpersonal communication could influence the preponderance of factual assertion versus an account of belief, and contaminate variation in properties like indirect speech acts (Clark, 1979; Searle, 1975). The Canadians, for example, are widely known for their extreme politeness and may require considerably more dire circumstances to abandon those principles, even incrementally. Rüsenberg (personal communication, February 2018) is pursuing extensions of my work in German, using events related to the recent election that will expand the understanding of these influences.

V. General Contributions & Future Work

The purpose of this dissertation is to examine the relationship between language use in social media and the environment it reflects. The application to the analysis of social media provides several methodological advantages. Social media produce voluminous data for every type of situation while simultaneously providing valuable metadata on time and location. These data exist naturally in the world, not as an artifact of an experiment. They are computationally accessible and allow for correlation with ground truth. My analysis reinforces an integrated view of language production processes and raises the opportunity to consider how a psychological perspective influences computational approaches to text mining.

An Integrated View of Language Production Processes

Within the field of psychology, various research traditions have separately investigated production, comprehension, dialogue, speech act theory, environmental grounding, and mediation, often with limited mutual acknowledgement. The work of (among others) fMRI researchers (e.g., Opitz, Müller, & Friederici, 2003; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012) has served to help bridge the gap between the first two, and Pickering and Garrod (2004) make a compelling case to consider those plus dialogue collectively. Psychophysics has certainly addressed the relationship between an environmental stimulus and human experience, albeit typically using highly restricted metrics and experimental paradigms that are bereft of motivating context. At the same time, the larger psychophysical issue has had virtually no influence on models of speech production, making those dyadic efforts of limited value in exploring the relationship between the environment and its description. Moreover, experimental researchers such as Clark and Wilkes-Gibbs (1986) and Strauss and McGrath (1994) impose a limited and artificial motivation on language production. In this respect, my research converges with a triadic view consistent with Roy (2005) and Flach and Voorhorst (2017) in which the environment is critical. The linking of environmental context with mediated communication (itself becoming part of the environment) constitutes, I believe, a novel consolidation while motivating a principled unification of the all-too-distributed research histories involved.

To guide my own research, I provided a schematic model to unify the various research traditions that address language production. The significance of this model is not its testability; the model is arguably both too complex and too underspecified to test. But the model does change the measures, the methods, and the interpretation of results. I focused on unconstrained lexical choice between options rather than absolute lexical frequency. To provide the aggregated measures that psychological methods employ, I examined patterns in lexical choice across a population. To scale these patterns, I employed word baselines. To provide comparison I examined lexical choice in response to different environments. To interpret the results, I suggested the narrative construct of breach as a motivation for message production. This constitutes an ambitious extension to a longstanding research tradition of psychophysics. Though still dependent on representations of experience, the representations are not the artificial product of an experiment, but the result of an inherent need to represent the purposes of coordinated activity.

In this case, interpretation involves the construction of a collective narrative. While others have noted the potential of social media for capturing the individual narrative (Anderson et al., 2016) my analysis points to the entire disaster corpus as exhibiting narrative properties in its own right, in line with what Mejova, Weber, and Macy (2015) call the "digital socioscope". Further, a collective narrative construction, sensitive to the social and conversational conventions of the Twitter medium, could potentially follow a "…yes and…" structure similar to improvisational theater, where the goal is to augment and adjust create a narrative as more information appears (represented by the "Schemas about the World" considerations in Figure 2). This constitutes a form of alignment (Garrod & Anderson, 1987).

A Psychological Perspective on Text Mining

I have demonstrated interpretable patterns of language behavior in social media during disasters using a novel, psychologically-inspired metric of lexical choice relative to a baseline standard. Comparison to an external standard constitutes an alternative approach to reliance on internal trend detection within a corpus (Bifet, Holmes, Pfahringer, & Gavaldà, 2011) or concern with the veracity of any individual report contributing to a tally. Methodological considerations such as concern for base rates and control corpora showcase the benefits of an experimentally oriented psychological approach to the analysis of big data.

In some respects, this complements widely employed frequency based approaches to mining social media text, illustrated by sentiment analysis. Sentiment analysis represents a substantial research base, both academically and commercially. One strength of sentiment analysis lies in its simplicity and domain independence. More positive words than negative words suggests a positive public opinion, and those words readily present themselves in neatly partitioned groups—one could easily suggest an unimpeachable list of a dozen words in each category. Sentiment analysis most closely resembles public polling in terms of information gathering. It exploits naturally occurring public conversations on virtually any topic to gauge the prevailing public perspective (at least among the Twitter-user demographic). This can help predict election results (e.g., Tumasjan, Sprenger, Sander, & Welpe, 2010; Wang, Can, Kazemzadeh, & Bar, 2012) or a movie's box office performance (e.g., Liu, Huang, An, & Yu, 2007; Mishne & Glance, 2006), in relation to how the public as a whole feels about the entities involved. I share some commonalities with the method involved—for example comparing the prevalence of positive terms relative to negative terms.

One way in which my approach complements sentiment analysis concerns interpretation of an unexpected result such as positive sentiment associated with the event epicenter. I have suggested that the positive sentiment is associated with pro-social behavior consistent with Quarantelli, 2008. Alternatively, my LIWC results may be revealing the amplified dispersal of tragic content or even misinformation by those just outside the impacted area (Lin & Margolin, 2014; Starbird et al., 2014; Thelwall, Buckley, & Paltoglou, 2011). Assuming my result replicates, interpretation demands intervening psychological processes between the events in the environment and the human response that generates social media messages. The naive model presented in the introduction (see Figure 1) cannot be sufficient.

A second way in which my approach complements sentiment analysis concerns the selection of relevant dimensions to monitor. While some researchers have expanded the dimensions with finer-grained emotion, I have pursued dimensions that reflect affordances for action (e.g., stop/start, all/some) and the properties of that action (together/alone). How people feel is only one dimension of human behavior. I forego relatively explicit statements of preference and approval for more implicit indications of breach associated with stress. I leverage relative use of common words and their matched opposites to infer stress using some of the same intuition as sentiment analysis. However, these are not mere substitutes for sentiment, which is generally uncorrelated with the word pairs that I examined. Generic sentinels of breach enable analysis across different disasters and do not require an *a priori* set of content terms. Sentinels naturally accommodate the diversity of perspective that supports wisdom of the crowds by potentially revealing a variety of unanticipated, specific “problems of living”.

Future Research

I have already identified a number of methodological limitations specific to my study and how they might be addressed in future replication and extension. Here I turn a critical eye to the larger research agenda, including aggregation over space and time, a Bayesian concern and the function of social media in disaster. These challenges pervade any large-scale social media analysis.

Aggregation over space and time. I made relatively arbitrary decisions concerning the levels of proximity to an event epicenter. I was able to segment the World Series corpus temporally according to known win-loss events. Yet, both modern physics and modern psychology (Boroditsky, 2000) acknowledge the common conceptual foundation of space and time. These must be considered together to provide a principled approach to partitioning a corpus for aggregation. With this in mind, an expansion of the current spatial analysis to the temporal dimension would likely yield valuable insights. I aim to quantify reaction to shifting environmental factors, mapping changes in a community's affordances for action to changes in language. This shift is subtle but has significant implications for the evaluation of ground truth. Methodologically, this requires a substantial expansion of target events, as obtaining sufficiently large corpora will require events with extended recovery periods that afford segmentation into meaningfully spaced windows. Spatial proximity requires consideration, so this would constitute a *further* segmentation and accompanying decrease in power. Moreover, spatial segmentation changes over time, for example as electrical service is restored in some areas but not others.

The ability to identify relative levels of breach, as the essential elements of this research may someday facilitate, will allow disaster management professionals to deploy aid with more precise information in both spatial and temporal frames of reference. In addition to initial conditions, social media provide continuing status reports on the affected areas. Often with disasters that extend beyond days into weeks (or months, as with Hurricanes Sandy and Katrina), the challenges lie in understanding continuing unmet resource needs (Anderson et al., 2016; Purohit et al., 2014). Social media, then, can provide an information stream for gauging the effectiveness of extended interventions. However, aggregation is relevant—a measure that reflects the collective narrative of the digital community obscures the identification of more directly impacted individuals.

A Bayesian concern. I exploited base rates to identify language patterns given a known breach, providing a key step in the identification of true social media alarms, that is, breach given observed language patterns. In this sense, the analysis is Bayesian-inspired. Though suggestive, my analysis of word preference given breach does not address the Bayesian distinction of breach given preference. The usual Bayesian concerns apply in such a conversion, particularly concerning the base rates of breach, which is a potentially intractable problem. A full accounting would require damage estimates, fatality totals, electrical outages, flood levels, and other potential objective correlates to breach of canonicity, *all* at a population-density-scaled granularity and somehow aggregated into a comprehensive score. This breach of canonicity depends upon cultural variables defining the collective narrative.

Function of social media in disaster. I remain concerned that Twitter is unique in its function relative to the GloWbe database used for base rate comparison. GloWbe includes an extensive array of Internet language. If the microblogging forum afforded by Twitter inherently encourages communication of breach of canonicity, only a base rate accounting for exclusively Twitter language properly determines deviations from baseline language behavior. One approach to the assessment of Twitter is to examine other social media platforms, and other verbal exchanges such as blogs or forums. However, other platforms have their own idiosyncrasies that certainly influence message creation. The most prudent course would be individual analyses of each medium followed by a comparison to the others—far beyond the scope of the current work.

While the challenges of evaluating social media in disaster remain substantial, there is promise for substantive contributions to disaster response down the road. Twitter reached 100 million users in only five and a half years (Twitter, 2017b). Harnessing this medium for the exchange of information between the public and the authorities lacks the procedures we have for other media such as 9-1-1. Grounded in the psychology of language production and springing from an effort to identify computationally inexpensive and available features of conversation, let this effort better formulate the interpretation problem and provide a step in that direction.

References

Al-Olimat, H. S., Thirunarayan, K., Shalin, V. L., & Sheth, A. P. (2017). Location name extraction from targeted text streams using Gazetteer-based statistical language models. arXiv preprint arXiv:1708.03105.

Anderson, J., Kogan, M., Bica, M., Palen, L., Anderson, K., Stowe, K., Morss, R., Demuth, J., Lazrus, H., Wilhelmi, O., & Henderson, J. (2016). Far, far away in far Rockaway: Responses to risks and impacts during Hurricane Sandy through first-person social media narratives. In *Proceedings of the 13th Information Systems for Crisis Response and Management*, Rio de Janeiro, Brazil, 16 pp.

Apache Software Foundation. (2010). OpenNLP (Version 1.7.2.) [Software]. Available from https://opennlp.apache.org/download.html

Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014, May). Tweedr: Mining twitter to inform disaster response. In *ISCRAM*.

Austin, J. L. (1962). *How to do things with words*. London: Oxford University Press.

Bereiter, C., & Scardamalia, M. (2013). *The psychology of written composition*. Routledge.

Bifet, A., Holmes, G., Pfahringer, B. & Gavaldà, R. (2011). Detecting sentiment change in Twitter streaming data. *Journal of Machine Learning Workshop and Conference Proceedings, 17*, 5-11.

Bochkarev, V. V., Shevlyakova, A. V., & Solovyev, V. D. (2012). Average word length dynamics as indicator of cultural changes in society. *arXiv preprint* *arXiv:1208.6109*.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355-387.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*(1), 1-28.

Boyd, D., Golder, S., & Lotan, G. (2010, January). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.

Bruner, J. (2003). The narrative construction of reality. *Narrative Intelligence*, 41-62.

Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K., & Tapia, A. (2014). Mapping moods: geo-mapped sentiment analysis during hurricane Sandy. *Proceedings of the 11th Information Systems for Crisis Response and Management*, University Park, Pennsylvania, USA, 10 pp.

Card, S. K., Newell, A., & Moran, T. P. (1983). *The psychology of human-computer interaction.* Hillsdale, New Jersey; Lawrence Erlbaum Associates, Inc.

Chamot, A. U., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign language annals, 22*(1), 13-22.

Chen, R., & Sakamoto, Y. (2014, January). Feelings and perspective matter: Sharing of crisis information in social media. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on System Sciences* (pp. 1958-1967). IEEE.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language, 35*(1), 26-58.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1966). *Topics in the theory of generative grammar*. Walter de Gruyter.

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences, 3,* 1-61.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of verbal learning and verbal behavior, 10*(3), 266-275.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76*(4), 387-404.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive psychology*,*11*(4), 430-477.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*, 127-149.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 332-373.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive psychology, 3*(3), 472-517.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50,* 62-81.

Clark, H. H., & Marshall, C. A. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science,* *13*(2), 259-294.

Clark, H. H., & Schunk, D. H. (1980). Polite responses to polite requests. *Cognition, 8*(2), 111-143.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Cleland, A. A., & Pickering M. J. (2006, February). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language, 54*(2), 185-198.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2. Auflage)*. Hillsdale, NJ: Erlbaum.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3-21.

Cruse, D. A. (1986). *Lexical Semantics.* Cambridge, MA.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011, March). Mark my words!: Linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web* (pp. 745-754). ACM.

Davies, M, & Fuchs, R. (2015). Expanding horizons in the study of world Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide 36*(1), 1-28.

De Saussure, F. (2011). *Course in general linguistics.* Columbia University Press.

Dell, G. S. (1986). A Spreading activation theory of retrieval and sentence production. *Psychological Review, 93*, 283-321.

Donaldson, M., & Wales, R. J. (1970). *On the acquisition of some relational terms.*

Cognition and the development of language. New York: Wiley, 235-268.

Dynel, M. (2009). *Humorous garden-paths: A pragmatic-cognitive study*. Cambridge Scholars Publishing.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178-186.

Erikson, K. T. (1995). *A new species of trouble: The human experience of modern disasters*. New York: WW Norton & Company.

Flach, J. M. & Voorhorst, F. (2017). *What matters: Putting common sense to work.* Dayton, OH: Flach & Voorhorst.

Fechner, G. T. (1912). The measurement of sensation. (Herbert Sidney Langfeld, Trans.). In B. Rand (Ed.), *The Classic Psychologists* (562-572). Boston: Houghton Mifflin. (Original work published 1860)

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Forbus, K. D. (1984). Qualitative process theory. *Artificial intelligence*, *24*(1), 85-168.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Gibson, D. R. (2008). How the outside gets in: Modeling conversational permeation. *Annual Review of Sociology 34*:359-384.

Gibson, D. R. (2010). Marking the turn: Obligation, engagement, and alienation in group discussions. *Social Psychology Quarterly 73*(2):132-151.

Gibson, J. J. (1977). *The theory of affordances.* Hilldale, USA.

Gilpin, A. R. (1973). Lexical marking effects in the semantic differential. *Journal of*

Psychology, 85(2), 277. doi:10.1080/00223980.1973.9915658

Gleitman, L. & Papafragou, A. (2013). Relations between language and thought. *In* D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology*, pp. 504-523. Oxford University Press 10.1093/oxfordhb/9780195376746.013.0032

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics, Vol. 3.* Edited by P. Cole and J. Morgan, Academic Press.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, *75*(2), B27-B39.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* (Vol. 13). Oxford: Blackwell.

Holtgraves, T., & Paul, K. (2013). Texting versus talking: An exploration in telecommunication language. *Telematics and Informatics, 30*(4), 289-295.

Hong, L., & Page, S. E. (2008). Some microfoundations of collective wisdom. *Collective Wisdom*, 56-71.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127-142.

Hotz, R. L. (2011). Decoding our chatter. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052970204138204576598942105167646.html.

Huberman, B. A., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. *First Monday, 14*(1). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2317/2063

Humphrey, Jr., C. E. v. Viacom, Inc., et al. Case No. 06-2768 (DMC) (D.N.J., June 20, 2007).

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR), 47*(4), 67.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013a, May). Extracting information nuggets from disaster-Related messages in social media. In *ISCRAM*.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013b, May). Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 1021-1024). ACM.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (pp. 56-65). ACM.

Jennex, M. E. (2012). Implementing social media in crisis response using knowledge management. *Managing Crises and Disasters with Emerging Technologies: Advancements: Advancements, 216*.

Jurgens, D., Finethy, T., McCorriston, J., Xu, Y. T., & Ruths, D. (2015, May). Geolocation prediction in twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Kawamura, T., & Ohsuga, A. (2013). Extraction and Estimation of Human Activity from Twitter for Information Sharing in Disaster. *Journal of Convergence Information*

Technology, *8*(11), 707.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123-1134.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). John Wiley & Sons, Inc.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). *Rapid assessment of disaster damage using social media activity. Science advances*, *2*(3), e1500779.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Lee, D. (2013). Boston bombing: How internet detectives got it very wrong. *BBC News*. Retrieved from http://www.bbc.com/news/technology-22214511

Levelt, W. J. (1993). Lexical access in speech production. In *Knowledge and Language* (pp. 241-251). Springer Netherlands.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). Boston, MA: MIT press.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106-131.

Li, J., & Rao, H. R. (2010). Twitter as a rapid response news service: An exploration in the context of the 2008 China earthquake. *The Electronic Journal of Information Systems in Developing Countries, 42*.

Lin, Y. R., & Margolin, D. (2014). The Ripple of fear, sympathy and solidarity during the Boston bombings. *EPJ Data Science*, *3*(1), 1.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage Publications.

Liu, Y., Huang, X., An, A., & Yu, X. (2007, July). ARSA: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 607-614). ACM.

Lopez, S. (2017, November 7). Twitter officially expands its character count to 280 starting today. *TechCrunch.* Retrieved from https://techcrunch.com/2017/11/07/twitter-officially-expands-its-character-count-to-280-starting-today/

Mejova, Y., Weber, I., & Macy, M. W. (Eds.). (2015). *Twitter: A digital socioscope*. Cambridge University Press.

Mey, J. L. (1993) *Pragmatics: An introduction.* Oxford: Blackwell.

Mishne G., & Glance N. (2006) Predicting movie sales from Blogger sentiment. In *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*. pp. 155–158.

National Institute of Standards and Technology (2012). *Partial rank correlation*. Retrieved from http://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/partraco.htm

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. WH Freeman/Times Books/Henry Holt & Co.

Neisser, U. (1978). Perceiving, anticipating, and imagining. *Minnesota Studies in the Philosophy of Science*, *9*, 89-106.

Newell, A. (1980). Physical symbol systems*. *Cognitive science*, *4*(2), 135-183.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337-360.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291.

Noether, G.E. (1976). *Introduction to statistics: A nonparametric approach.* Houghton Mifflin, Boston.

Norman, D. (1988). *The psychology of everyday things.* Basic Books.

Noth, W. (2009). On the instrumentality and semiotic agency of signs, tools, and intelligent machines. *Cybernetics & Human Knowing*, *16*(3-4), 11-36.

Nunberg, G. (1979). The Non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*(2), 143-184.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the influence of thought and of the science of symbolism.* New York: Harcourt, Brace & World, Inc.

Opitz, B., Müller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*, *16*(2), 285-296.

Palen, L., & Liu, S. B. (2007, April). Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 727-736). ACM.

Palen, L., Starbird, K., & Vieweg, S. (2010). Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bulletin of the Association for Information Science and Technology, 36*(5), 13-17.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1-2), 1-135.

Parunak, H., Brueckner, S. A., Hong, L., Page, S. E., & Rohwer, R. (2013). Characterizing and aggregating agent estimates. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 1021-1028). International Foundation for Autonomous Agents and Multi-Agent Systems.

Peacock, W. G. & Ragsdale, A. K. (1997). Social systems, ecological network and disasters. In W. G. Peacock, B. H. Morrow, & H. Gladwin (Eds.), *Hurricane Andrew: Ethnicity, gender, and the sociology of disasters* (20-35). New York, NY: Routledge.

Peirce, C. S. (2015). *On a new list of categories*. Media Galaxy.

Peirce, C. S. (1894). What is a Sign?. *Philosophical Writings of Peirce*, 98-104.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015a). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015b). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*, 2001.

Perry, R. W. (2007). What is a disaster? *In* H. Rodriguez, E. L. Quarantelli & R. R. Dynes (Eds.) *Handbook of Disaster Research*, pp. 1-15. Springer Science and Business Media.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530-535.

Pinker, S. (1995). *The Language instinct: The New science of language and mind* (Vol. 7529). Penguin UK.

Purohit, H., Hampton, A., Bhatt, S., Shalin, V. L., Sheth, A. P., & Flach, J. M. (2014). Identifying seekers and suppliers in social media communities to support crisis coordination. *Computer Supported Cooperative Work (CSCW)*, *23*(4-6), 513-545.

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J. M., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, *29*(6), 2438-2447.

Purohit, H. & Sheth, A. P. (2013). Twitris v3: From citizen sensing to analysis, coordination and action. In *Proceedings of the Seventh International Conference on Weblogs and Social Media, 8-11 July 2013, Cambridge, MA.* Palo Alto, CA: The AAAI Press, pp. 746-747.

Quarantelli, E. L. (1986). Research findings on organizational behavior in disasters and their applicability in developing countries. *Disaster Research Center.* Preliminary paper 107. Boulder, CO: University of Delaware.

Quarantelli, E. (Ed.) (2005). *What is a disaster? A dozen perspectives on the question.* Routledge, London.

Quarantelli, E. (2008). Conventional beliefs and counterintuitive realities. *Social Research: An International Quarterly of the Social Sciences,* 75(3), 873-904.

Ritter, A., Cherry, C., & Dolan, B. (2010). Unsupervised modeling of twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL* (pp. 172–180). National Research Council Canada.

Rodríguez, H., Trainor, J., & Quarantelli, E. L. (2006). Rising to the challenges of a catastrophe: The emergent and prosocial behavior following Hurricane Katrina. *The Annals of the American Academy of Political and Social Science, 604*(1), 82-101.

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, *167*(1), 170-205.

Searle, J. R. (1975). Indirect speech acts. *Syntax and Semantics, 3*, 59-82.

Searle, J. R. (1976). *A taxonomy of illocutionary acts* (pp. 355-68). Linguistic Agency University of Trier.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex*, *22*(7), 1662-1670.

Selye, H. (1978). *The Stress of Life: Rev Ed*. Chicago: McGraw-Hill.

Schoenharl, T., Madey, G., Szabó, G., & Barabási, A. L. (2006, May). WIPER: A multi-agent system for emergency response. In *Proceedings of the 3rd International ISCRAM Conference* (Vol. 1, No. 7).

Scholand, A. J., Tausczik, Y. R., & Pennebaker, J. W. (2010, February). Social language network analysis. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 23-26). ACM.

Sheth, A. P. (2009). Citizen sensing, social signals, and enriching human experience. *IEEE Internet Computing 13*(4), 87.

Sheth, A. P. & Thirunarayan, K. (2012). Semantics empowered web 3.0: Managing enterprise, social, sensor, and cloud-based data and services for advanced applications. *Synthesis Lectures on Data Management*, *4*(6), 1-175.

Smith, K. J., & Belgrave, L. L. (1995). The reconstruction of everyday life: Experiencing Hurricane Andrew. *Journal of Contemporary Ethnography*, *24*(3), 244-269.

Starbird, K. (2011, May). Digital volunteerism during disaster: Crowdsourcing information processing. In *Conference on Human Factors in Computing Systems* (pp. 7-12).

Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R.M. (2014). Rumors, false flags and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon bombing. In M. Kindling & E. Greifeneder (Eds.) *iConference 2014 Proceedings*, pp. 654–662.

Starbird, K. & Palen, L. (2010). Pass It On?: Retweeting in mass emergencies. In S. French, B. Tomaszewski, and C. Zobel (eds.): *The 7th International Information Systems for Crisis Response and Management Conference*, Seattle, WA, USA, May 2010.

Starbird, K., & Stamberger, J. (2010). Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In *Proceedings of the 7th International ISCRAM Conference,* Seattle, WA.

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science, 133*, 80-86.

Straus, S. G., & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology*, *79*(1), 87.

Suedfeld, P. (1997). Reactions to societal trauma: Distress and/or eustress. *Political Psychology, 18*(4), 849-861.

Tapia, A. H., Bajpai, K., Jansen, B. J., Yen, J., & Giles, L. (2011, May). Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. In *Proceedings of the 8th International ISCRAM Conference* (pp. 1-10).

Tapia, A. H., Moore, K. A., & Johnson, N. J. (2013, May). Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. In *ISCRAM*.

Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*(2), 406-418.

Tooley, K. M., & Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition*, *132*(2), 101-136.

Tumasjan A., Sprenger T. O., Sander P. G., & Welpe I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 178–185.

Tversky, A., & Kahneman, D. (1981). The Framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Twitter. (2017a). *Company*. Retrieved from https://about.twitter.com/company

Twitter. (2017b). *Milestones*. Retrieved from https://about.twitter.com/milestones

Twitter. (2017c). *Twitter terms of service.* Retrieved from https://twitter.com/tos?lang=en

Vedula, N., Parthasarathy, S. & Shalin, V. L. (2016). Predicting trust relations among users in a social network. On the roles of influence, cohesion and valence. *Proceedings of ACM SKGKDD WISDOM.*

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, pp. 115-120.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012, September). Harnessing twitter" big data" for automatic emotion identification. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. IEEE, pp. 587-592.

Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191-1207.

Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social media as information source: Recency of updates and credibility of information. *Journal of Computer-Mediated Communication, 19*(2), 171-183.

Whorf, B. L. (1956). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. (Edited by John B. Carroll.). Oxford, England: Technology Press of MIT.

Wilson, D. & Sperber, D. (2004). Relevance theory. *In* Horn, L.R. & Ward, G. (eds.) *The Handbook of Pragmatics*, pp. 607-632. Oxford: Blackwell.

Wilson, D. B. (2000). Practical meta-analysis effect size calculator [Software]. Available from http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD10.php

Wittgenstein, L. (2009). *Philosophical investigations* (4th ed.) (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.). Hoboken, New Jersey: Wiley-Blackwell. (Original work published 1953.)

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems, 55*(4), 919-926.

APPENDIX

Table A1

ANOVA for Word Count

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	19135927456.83	8	2391990932.10	6.05	0.00
Event	5290610486.97	5	1058122097.39	2.44	0.05
Non-Disaster vs Disaster	200516400.75	1	200516400.75	0.46	0.50
Football vs. Baseball	554839200.75	1	554839200.75	1.28	0.26
Manmade vs. Natural Disaster	642783856.33	1	642783856.33	1.48	0.23
Baseball Win vs. Loss	421925834.03	1	421925834.03	0.97	0.33
Oklahoma vs. Sandy	3470545195.11	1	3470545195.11	8.00	0.01
High/Low	2297080394.08	1	2297080394.08	3.97	0.08
Event*High/Low	1702211134.31	5	340442226.86	1.05	0.40
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	1879944.45	1	1879944.45	0.01	0.94
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	477368421.12	1	477368421.12	1.48	0.23
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	17935705.04	1	17935705.04	0.06	0.81
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	231815850.25	1	231815850.25	0.72	0.40
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	973211213.44	1	973211213.44	3.01	0.09
Residual	34890065870.06	88	396478021.25		
Event*Pair	17342865895.94	40	433571647.40		
High/Low*Pair	4623260284.17	8	577907535.52		
Event*High/Low*Pair	12923939689.94	40	323098492.25		
Total	63315895342.25	107			

Table A2

ANOVA for Prevalence of Certainty Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	61.26	8	7.66	8.30	0.00
Event	5.17	5	1.03	5.19	0.00
Non-Disaster vs Disaster	2.04	1	2.04	10.27	0.00
Football vs. Baseball	1.93	1	1.93	9.71	0.00
Manmade vs. Natural Disaster	0.00	1	0.00	0.02	0.90
Baseball Win vs. Loss	1.02	1	1.02	5.10	0.03
Oklahoma vs. Sandy	0.17	1	0.17	0.86	0.36
High/Low	10.62	1	10.62	1.31	0.29
Event*High/Low	2.61	5	0.52	2.54	0.04
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	1.03	1	1.03	5.03	0.03
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.40	1	1.40	6.84	0.01
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.02	1	0.02	0.08	0.77
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.15	1	0.15	0.75	0.39
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.00	1	0.00	0.00	0.99
Residual	81.18	88	0.92		
Event*Pair	7.97	40	0.20		
High/Low*Pair	65.01	8	8.13		
Event*High/Low*Pair	8.21	40	0.21		
Total	160.83	107			

Table A3

ANOVA for Prevalence of Certainty Words with Confound Removed

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	5.11	7	0.73	2.92	0.01
Event	5.51	5	1.10	5.13	0.00
Non-Disaster vs Disaster	1.98	1	1.98	9.23	0.00
Football vs. Baseball	2.21	1	2.21	10.28	0.00
Manmade vs. Natural Disaster	0.00	1	0.00	0.00	0.95
Baseball Win vs. Loss	1.13	1	1.13	5.27	0.03
Oklahoma vs. Sandy	0.19	1	0.19	0.87	0.36
High/Low (Conservative)	0.21	1	0.21	0.40	0.55
High/Low (Liberal)	0.21	1	0.21	0.86	0.36
Event*High/Low	2.80	5	0.56	2.47	0.05
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	1.23	1	1.23	5.43	0.03
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.39	1	1.39	6.13	0.02
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.01	1	0.01	0.04	0.84
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.17	1	0.17	0.73	0.40
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.00	1	0.00	0.00	0.97
Residual	19.23	77	0.25		
Event*Pair	7.52	35	0.21		
High/Low*Pair	3.77	7	0.54		
Event*High/Low*Pair	7.94	35	0.23		
Total	32.87	95			

Table A4

ANOVA for Prevalence of Perception Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	74.03	8	9.25	3.18	0.00
Event	120.65	5	24.13	9.26	0.00
Non-Disaster vs Disaster	56.39	1	56.39	21.64	0.00
Football vs. Baseball	0.04	1	0.04	0.01	0.90
Manmade vs. Natural Disaster	9.19	1	9.19	3.53	0.07
Baseball Win vs. Loss	0.71	1	0.71	0.27	0.60
Oklahoma vs. Sandy	54.32	1	54.32	20.84	0.00
High/Low	0.79	1	0.79	0.16	0.70
Event*High/Low	18.49	5	3.70	1.51	0.21
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.09	1	0.09	0.04	0.85
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	4.39	1	4.39	1.79	0.19
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	12.80	1	12.80	5.21	0.03
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.93	1	0.93	0.38	0.54
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.27	1	0.27	0.11	0.74
Residual	241.45	83	2.91		
Event*Pair	104.24	40	2.61		
High/Low*Pair	39.01	8	4.88		
Event*High/Low*Pair	98.21	40	2.46		
Total	455.41	107			

Table A5

ANOVA for Prevalence of Biological Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	22.56	8	2.82	2.29	0.03
Event	11.65	5	2.33	1.25	0.30
Non-Disaster vs Disaster	2.36	1	2.36	1.27	0.27
Football vs. Baseball	0.01	1	0.01	0.00	0.95
Manmade vs. Natural Disaster	8.37	1	8.37	4.49	0.04
Baseball Win vs. Loss	0.85	1	0.85	0.45	0.50
Oklahoma vs. Sandy	0.06	1	0.06	0.03	0.86
High/Low	0.45	1	0.45	0.48	0.51
Event*High/Low	2.01	5	0.40	0.61	0.70
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	1.36	1	1.36	2.06	0.16
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	0.21	1	0.21	0.32	0.58
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.03	1	0.03	0.04	0.85
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.39	1	0.39	0.59	0.45
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.02	1	0.02	0.03	0.87
Residual	108.55	88	1.23		
Event*Pair	74.60	40	1.86		
High/Low*Pair	7.45	8	0.93		
Event*High/Low*Pair	26.50	40	0.66		
Total	145.22	107			

Table A6

ANOVA for Prevalence of Affiliation Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	185.74	8	23.22	8.15	0.00
Event	53.69	5	10.74	9.06	0.00
Non-Disaster vs Disaster	0.70	1	0.70	0.59	0.45
Football vs. Baseball	37.03	1	37.03	31.23	0.00
Manmade vs. Natural Disaster	1.80	1	1.80	1.52	0.23
Baseball Win vs. Loss	13.30	1	13.30	11.21	0.00
Oklahoma vs. Sandy	0.86	1	0.86	0.73	0.40
High/Low	11.66	1	11.66	0.68	0.43
Event*High/Low	1.40	5	0.28	0.17	0.97
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.18	1	0.18	0.11	0.74
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.11	1	1.11	0.67	0.42
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.07	1	0.07	0.04	0.84
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.03	1	0.03	0.02	0.89
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.00	1	0.00	0.00	0.96
Residual	250.76	88	2.85		
Event*Pair	47.43	40	1.19		
High/Low*Pair	136.66	8	17.08		
Event*High/Low*Pair	66.66	40	1.67		
Total	503.24	107			

Table A7

ANOVA for Prevalence of Affiliation Words with Confound Removed

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	19.62	7	2.80	1.86	0.09
Event	49.28	5	9.86	7.98	0.00
Non-Disaster vs Disaster	1.01	1	1.01	0.81	0.37
Football vs. Baseball	33.84	1	33.84	27.40	0.00
Manmade vs. Natural Disaster	2.24	1	2.24	1.82	0.19
Baseball Win vs. Loss	12.10	1	12.10	9.80	0.00
Oklahoma vs. Sandy	0.08	1	0.08	0.06	0.80
High/Low	0.22	1	0.22	0.11	0.75
Event*High/Low	4.17	5	0.83	0.50	0.78
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.63	1	0.63	0.38	0.54
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	3.20	1	3.20	1.90	0.18
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.19	1	0.19	0.11	0.74
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.16	1	0.16	0.09	0.76
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.00	1	0.00	0.00	0.97
Residual	116.21	77	1.51		
Event*Pair	43.24	35	1.24		
High/Low*Pair	14.17	7	2.02		
Event*High/Low*Pair	58.81	35	1.68		
Total	189.50	95			

Table A8

ANOVA for Prevalence of Analytic Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	3797.67	8	474.71	2.60	0.01
Event	426.84	5	85.37	0.53	0.75
Non-Disaster vs Disaster	86.65	1	86.65	0.54	0.47
Football vs. Baseball	0.00	1	0.00	0.00	1.00
Manmade vs. Natural Disaster	199.43	1	199.43	1.23	0.27
Baseball Win vs. Loss	109.69	1	109.69	0.68	0.41
Oklahoma vs. Sandy	31.06	1	31.06	0.19	0.67
High/Low	444.33	1	444.33	1.15	0.31
Event*High/Low	1386.90	5	277.38	1.71	0.15
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	12.70	1	12.70	0.08	0.78
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1076.80	1	1076.80	6.65	0.01
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.23	1	0.23	0.00	1.00
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	295.61	1	295.61	1.83	0.18
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	1.55	1	1.55	0.01	0.92
Residual	16049.75	88	182.38		
Event*Pair	6475.84	40	161.90		
High/Low*Pair	3100.23	8	387.53		
Event*High/Low*Pair	6473.69	40	161.84		
Total	22105.48	107			

Table A9

ANOVA for Prevalence of Words that Focus on the Present

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	185.68	8	23.21	5.36	0.00
Event	52.25	5	10.45	3.06	0.02
Non-Disaster vs Disaster	2.24	1	2.24	0.66	0.42
Football vs. Baseball	35.21	1	35.21	10.30	0.00
Manmade vs. Natural Disaster	0.21	1	0.21	0.06	0.80
Baseball Win vs. Loss	4.40	1	4.40	1.29	0.26
Oklahoma vs. Sandy	10.18	1	10.18	2.98	0.09
High/Low	3.21	1	3.21	0.29	0.61
Event*High/Low	7.61	5	1.52	0.39	0.85
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.09	1	0.09	0.02	0.88
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	0.00	1	0.00	0.00	0.97
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	1.77	1	1.77	0.46	0.50
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	5.22	1	5.22	1.34	0.25
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.52	1	0.52	0.13	0.72
Residual	381.35	88	4.33		
Event*Pair	136.73	40	3.42		
High/Low*Pair	88.78	8	11.10		
Event*High/Low*Pair	155.84	40	3.90		
Total	630.10	107			

Table A10

ANOVA for Prevalence of Clout Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	3628.37	8	453.55	2.71	0.01
Event	2374.13	5	474.83	3.59	0.01
Non-Disaster vs Disaster	420.44	1	420.44	3.17	0.08
Football vs. Baseball	492.12	1	492.12	3.72	0.06
Manmade vs. Natural Disaster	207.86	1	207.86	1.57	0.22
Baseball Win vs. Loss	312.35	1	312.35	2.36	0.13
Oklahoma vs. Sandy	941.36	1	941.36	7.11	0.01
High/Low	0.04	1	0.04	0.00	0.99
Event*High/Low	868.70	5	173.74	1.06	0.40
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	138.65	1	138.65	0.85	0.36
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	148.12	1	148.12	0.90	0.35
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	98.67	1	98.67	0.60	0.44
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	455.25	1	455.25	2.78	0.10
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	28.00	1	28.00	0.17	0.68
Residual	14745.59	88	167.56		
Event*Pair	5296.90	40	132.42		
High/Low*Pair	2889.10	8	361.14		
Event*High/Low*Pair	6559.59	40	163.99		
Total	21616.82	107			

Table A11

ANOVA for Prevalence of Authentic Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	10407.22	8	1300.90	10.04	0.00
Event	4784.94	5	956.99	9.36	0.00
Non-Disaster vs Disaster	971.58	1	971.58	9.51	0.00
Football vs. Baseball	3750.75	1	3750.75	36.70	0.00
Manmade vs. Natural Disaster	14.37	1	14.37	0.14	0.71
Baseball Win vs. Loss	0.00	1	0.00	0.00	1.00
Oklahoma vs. Sandy	48.23	1	48.23	0.47	0.50
High/Low	334.08	1	334.08	2.87	0.13
Event*High/Low	527.43	5	105.49	0.66	0.65
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	210.98	1	210.98	1.32	0.26
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	3.59	1	3.59	0.02	0.88
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	30.14	1	30.14	0.19	0.67
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	221.81	1	221.81	1.39	0.25
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	60.92	1	60.92	0.38	0.54
Residual	11397.46	88	129.52		
Event*Pair	4088.03	40	102.20		
High/Low*Pair	932.08	8	116.51		
Event*High/Low*Pair	6377.35	40	159.43		
Total	27451.13	107			

Table A12

ANOVA for Prevalence of Six Letter Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	573.23	8	71.65	2.69	0.01
Event	284.91	5	56.98	2.44	0.05
Non-Disaster vs Disaster	257.83	1	257.83	11.06	0.00
Football vs. Baseball	0.48	1	0.48	0.02	0.89
Manmade vs. Natural Disaster	17.35	1	17.35	0.74	0.39
Baseball Win vs. Loss	5.95	1	5.95	0.26	0.62
Oklahoma vs. Sandy	3.29	1	3.29	0.14	0.71
High/Low	5.29	1	5.29	0.10	0.76
Event*High/Low	82.96	5	16.59	0.68	0.64
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	9.36	1	9.36	0.38	0.54
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	46.20	1	46.20	1.89	0.18
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	21.86	1	21.86	0.89	0.35
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	1.23	1	1.23	0.05	0.82
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	4.31	1	4.31	0.18	0.68
Residual	2340.71	88	26.60		
Event*Pair	932.34	40	23.31		
High/Low*Pair	428.78	8	53.60		
Event*High/Low*Pair	979.59	40	24.49		
Total	3287.10	107			

Table A13

ANOVA for Prevalence of Function Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	984.28	8	123.04	1.93	0.07
Event	1453.21	5	290.64	3.65	0.01
Non-Disaster vs Disaster	991.17	1	991.17	12.44	0.00
Football vs. Baseball	444.08	1	444.08	5.57	0.02
Manmade vs. Natural Disaster	7.02	1	7.02	0.09	0.77
Baseball Win vs. Loss	1.38	1	1.38	0.02	0.90
Oklahoma vs. Sandy	9.55	1	9.55	0.12	0.73
High/Low	50.40	1	50.40	0.43	0.53
Event*High/Low	190.98	5	38.20	1.03	0.41
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	26.44	1	26.44	0.71	0.40
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	3.03	1	3.03	0.08	0.78
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	4.92	1	4.92	0.13	0.72
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	122.10	1	122.10	3.29	0.08
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	34.50	1	34.50	0.93	0.34
Residual	5609.31	88	63.74		
Event*Pair	3186.99	40	79.67		
High/Low*Pair	937.90	8	117.24		
Event*High/Low*Pair	1484.42	40	37.11		
Total	8288.20	107			

Table A14

ANOVA for Prevalence of Personal Pronouns

Source		*SS*	*df*	*MS*	*F*	*p*
Pairs		137.78	8	17.22	5.10	0.00
Event		72.52	5	14.50	4.63	0.00
	Non-Disaster vs Disaster	45.55	1	45.55	14.56	0.00
	Football vs. Baseball	7.92	1	7.92	2.53	0.12
	Manmade vs. Natural Disaster	15.05	1	15.05	4.81	0.03
	Baseball Win vs. Loss	3.99	1	3.99	1.27	0.27
	Oklahoma vs. Sandy	0.02	1	0.02	0.01	0.94
High/Low		0.01	1	0.01	0.00	0.97
Event*High/Low		7.70	5	1.54	0.61	0.69
	Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.67	1	0.67	0.27	0.61
	Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	5.41	1	5.41	2.16	0.15
	Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.32	1	0.32	0.13	0.72
	Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.36	1	0.36	0.14	0.71
	Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.94	1	0.94	0.37	0.54
Residual		297.42	88	3.38		
	Event*Pair	125.18	40	3.13		
	High/Low*Pair	72.01	8	9.00		
	Event*High/Low*Pair	100.23	40	2.51		
Total		515.44	107			

Table A15

ANOVA for Prevalence of Cognitive Processing Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	383.12	8	47.89	9.10	0.00
Event	17.63	5	3.53	0.71	0.62
Non-Disaster vs Disaster	17.50	1	17.50	3.54	0.07
Football vs. Baseball	0.01	1	0.01	0.00	0.96
Manmade vs. Natural Disaster	0.01	1	0.01	0.00	0.96
Baseball Win vs. Loss	0.02	1	0.02	0.00	0.95
Oklahoma vs. Sandy	0.08	1	0.08	0.02	0.90
High/Low	0.27	1	0.27	0.02	0.89
Event*High/Low	15.39	5	3.08	0.77	0.58
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	6.08	1	6.08	1.53	0.22
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	2.49	1	2.49	0.63	0.43
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	1.04	1	1.04	0.26	0.61
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	5.74	1	5.74	1.44	0.24
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.03	1	0.03	0.01	0.93
Residual	463.27	88	5.26		
Event*Pair	197.52	40	4.94		
High/Low*Pair	106.41	8	13.30		
Event*High/Low*Pair	159.33	40	3.98		
Total	879.68	107			

Table A16

ANOVA for Prevalence of Relative Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	561.23	8	70.15	9.63	0.00
Event	297.37	5	59.47	11.23	0.00
Non-Disaster vs Disaster	15.44	1	15.44	2.91	0.10
Football vs. Baseball	263.39	1	263.39	49.74	0.00
Manmade vs. Natural Disaster	1.15	1	1.15	0.22	0.64
Baseball Win vs. Loss	7.51	1	7.51	1.42	0.24
Oklahoma vs. Sandy	9.88	1	9.88	1.87	0.18
High/Low	29.55	1	29.55	1.62	0.24
Event*High/Low	16.11	5	3.22	0.45	0.81
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	3.89	1	3.89	0.55	0.46
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	10.05	1	10.05	1.42	0.24
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.46	1	0.46	0.06	0.80
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.96	1	0.96	0.14	0.71
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.76	1	0.76	0.11	0.74
Residual	641.03	88	7.28		
Event*Pair	211.82	40	5.30		
High/Low*Pair	145.53	8	18.19		
Event*High/Low*Pair	283.67	40	7.09		
Total	1545.28	107			

Table A17

ANOVA for Prevalence of Positive Emotion Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	153.50	8	19.19	5.87	0.00
Event	10.78	5	2.16	1.47	0.22
Non-Disaster vs Disaster	0.50	1	0.50	0.34	0.56
Football vs. Baseball	0.02	1	0.02	0.01	0.92
Manmade vs. Natural Disaster	5.67	1	5.67	3.86	0.06
Baseball Win vs. Loss	0.09	1	0.09	0.06	0.81
Oklahoma vs. Sandy	4.51	1	4.51	3.07	0.09
High/Low	69.31	1	69.31	3.21	0.11
Event*High/Low	4.13	5	0.83	0.59	0.71
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.31	1	0.31	0.22	0.64
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.32	1	1.32	0.94	0.34
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.01	1	0.01	0.01	0.94
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.77	1	0.77	0.54	0.47
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	1.73	1	1.73	1.23	0.27
Residual	287.61	88	3.27		
Event*Pair	58.68	40	1.47		
High/Low*Pair	172.55	8	21.57		
Event*High/Low*Pair	56.38	40	1.41		
Total	525.34	107			

Table A18

ANOVA for Prevalence of Positive Emotion Words with Confound Removed

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	39.20	7	5.60	2.36	0.03
Event	10.52	5	2.10	1.28	0.30
Non-Disaster vs Disaster	0.94	1	0.94	0.57	0.46
Football vs. Baseball	0.04	1	0.04	0.02	0.88
Manmade vs. Natural Disaster	5.17	1	5.17	3.13	0.09
Baseball Win vs. Loss	0.06	1	0.06	0.04	0.85
Oklahoma vs. Sandy	4.31	1	4.31	2.62	0.11
High/Low	20.02	1	20.02	2.00	0.20
Event*High/Low	3.95	5	0.79	0.50	0.77
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.35	1	0.35	0.22	0.64
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	1.28	1	1.28	0.81	0.37
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.02	1	0.02	0.01	0.91
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	1.12	1	1.12	0.71	0.41
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	1.18	1	1.18	0.75	0.39
Residual	182.92	77	2.38		
Event*Pair	57.73	35	1.65		
High/Low*Pair	70.05	7	10.01		
Event*High/Low*Pair	55.14	35	1.58		
Total	256.61	95			

Table A19

ANOVA for Prevalence of Ingestion Words

Source	*SS*	*df*	*MS*	*F*	*p*
Pairs	0.59	8	0.07	0.60	0.78
Event	1.64	5	0.33	2.95	0.02
Non-Disaster vs Disaster	0.58	1	0.58	5.23	0.03
Football vs. Baseball	0.37	1	0.37	3.29	0.08
Manmade vs. Natural Disaster	0.42	1	0.42	3.77	0.06
Baseball Win vs. Loss	0.07	1	0.07	0.65	0.43
Oklahoma vs. Sandy	0.20	1	0.20	1.81	0.19
High/Low	0.48	1	0.48	3.48	0.10
Event*High/Low	0.48	5	0.10	0.72	0.61
Sports Lows/Disaster Highs vs. Sports Highs/Disaster Lows	0.35	1	0.35	2.62	0.11
Football Highs/Baseball Lows vs. Baseball Wins/Baseball Losses Highs	0.00	1	0.00	0.00	0.99
Manmade Highs/Natural Lows vs. Manmade Lows/Natural Highs	0.00	1	0.00	0.00	0.96
Baseball Wins Highs/Losses Lows vs. Baseball Wins Lows/Losses Highs	0.10	1	0.10	0.76	0.39
Oklahoma Highs/Sandy Lows vs. Oklahoma Lows/Sandy Highs	0.03	1	0.03	0.21	0.65
Residual	10.88	88	0.12		
Event*Pair	4.45	40	0.11		
High/Low*Pair	1.11	8	0.14		
Event*High/Low*Pair	5.32	40	0.13		
Total	14.07	107			